Multivariate Normal Distribution

Shyh-Kang Jeng
Department of Electrical Engineering/
Graduate Institute of Communication/
Graduate Institute of Networking and Multimedia

1

Outline

- Introduction
- The Multivariate Normal Density and Its Properties
- Sampling from a Multivariate Normal Distribution and Maximum Likelihood Estimation
- The Sampling Distribution of $\bar{X}$ and *S*
- Large-Sample Behavior of $\bar{X}$ and *S*

Outline

- Assessing the Assumption of Normality
- Detecting Outliers and Cleaning Data
- Transformations to Near Normality

Outline

- Introduction
- The Multivariate Normal Density and Its Properties
- Sampling from a Multivariate Normal Distribution and Maximum Likelihood Estimation
- The Sampling Distribution of $\bar{X}$ and *S*
- Large-Sample Behavior of $\bar{X}$ and *S*

Questions

- What is the univariate normal distribution?
- What is the multivariate normal distribution?
- Why to study multivariate normal distribution?

Multivariate Normal Distribution

- Generalized from univariate normal density
- Base of many multivariate analysis techniques
- Useful approximation to "true" population distribution
- Central limit distribution of many multivariate statistics
- Mathematical tractable

6

Outline

- Introduction
- The Multivariate Normal Density and Its Properties
- Sampling from a Multivariate Normal Distribution and Maximum Likelihood Estimation
- The Sampling Distribution of $\overline{X}$ and *S*
- Large-Sample Behavior of $\overline{X}$ and *S*

Questions

- What is the formula for the probability density function of a univariate normal distribution?
- What are the probability meaning of parameters μ and σ?
- How much probability are in the intervals $(\mu-\sigma, \mu+\sigma)$ and $(\mu-2\sigma, \mu+2\sigma)$?
- How to look up the accumulated univariate normal probability in Table 1, Appendix?

Questions

- What is the Mahalanobis distance for univariate normal distribution?
- What is the Mahalanobis distance for multivariate normal distribution?
- What are the symbol for and the formula of the probability density of a p-dimensional multivariate normal distribution?

Questions

- What are the possible shapes in a surface diagram of a bivariate normal density?
- What is the constant probability density contour for a p-dimensional multivariate normal distribution?
- What are the eigenvalues and eigenvectors of the inverse of Σ? (Result 4.1)

Questions

- What is the region that the total probability inside equals 1-α?
- What is the probability distribution for a linear combination of p random variables with the same multivariate-normal distribution? (Result 4.2)
- How to find the marginal distribution of a multivariate-normal distribution by Result 4.2?

Questions

- What is the probability distribution for a random vector obtained by multiplying a matrix to a random vector of p random variables with the same multivariate-normal distribution? (Result 4.3)
- What is the probability distribution of a random vector of multivariate normal distribution plus a constant vector? (Result 4.3)

Questions

- Given the mean and covariance matrix of a multivariate random vector, and the random vector is partitioned, how to find the mean and covariance matrix of the two parts of the partitioned random vector? (Result 4.4)

Questions

- What are the if-and-only-if conditions for two multivariate normal vectors X_1 and X_2 to be independent? (Result 4.5)
- If two multivariate normal vectors X_1 and X_2 are independent, what will be the probability distribution of the random vector partitioned into X_1 and X_2? (Result 4.5)

Questions

- A random vector X is partitioned into X_1 and X_2, then what is the conditional probability distribution od X_1 given $X_2 = x_2$? (Result 4.6)
- What is the probability distribution for the square of the Mahalanobis distance for a multivariate normal vector? (Result 4.7)

Questions

- How to find the value of the Mahalanobis distance for a multivariate normal vector when the probability inside the corresponding ellipsoid is specified? (Result 4.7)

Questions

- What is the shape of a chi-square distribution curve?
- How to look up the accumulated chi-square probability from Table 3, Appendix?
- What is the joint distribution of two random vectors which are two linear combinations of *n* different multivariate random vectors? (Result 4.8)

Univariate Normal Distribution

$N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-[(x-\mu)/\sigma]^2/2} \qquad -\infty < x < \infty$$

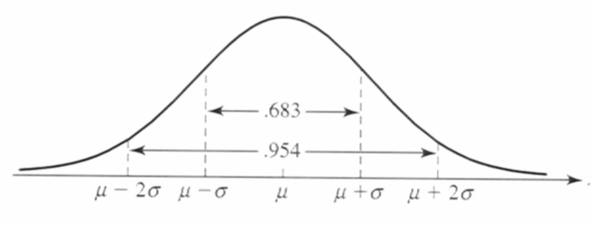
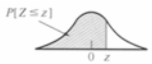


Table 1, Appendix



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389

Square of Distance (Mahalanobis distance)

$$\left(\frac{x-\mu}{\sigma}\right)^2 = (x-\mu)(\sigma^2)^{-1}(x-\mu)$$

$$\Downarrow$$

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

p-dimensional Normal Density

$N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

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

$-\infty < x_i < \infty, \quad i = 1, 2, \cdots, p$

$\mathbf{x}$ is a sample from random vector

$\mathbf{X}' = [X_1, X_2, \cdots, X_p]$

21

Example 4.1 Bivariate Normal

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

$\sigma_{11} = \mathrm{Var}(X_1), \sigma_{22} = \mathrm{Var}(X_2)$

$\rho_{12} = \sigma_{12} / \left(\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}\right) = \mathrm{Corr}(X_1, X_2)$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}, \boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}$$

$\sigma_{11}\sigma_{22} - \sigma_{12}^2 = \sigma_{11}\sigma_{22}(1 - \rho_{12}^2)$

22

Example 4.1 Squared Distance

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

$$= [x_1 - \mu_1, x_2 - \mu_2] \frac{1}{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)} \begin{bmatrix} \sigma_{22} & -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} \\ -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} & \sigma_{11} \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}$$

$$= \frac{1}{1-\rho_{12}^2} \left[\left(\frac{x_1 - \mu_1}{\sqrt{\sigma_{11}}} \right)^2 + \left(\frac{x_2 - \mu_2}{\sqrt{\sigma_{22}}} \right)^2 - 2\rho_{12} \left(\frac{x_1 - \mu_1}{\sqrt{\sigma_{11}}} \right) \left(\frac{x_2 - \mu_2}{\sqrt{\sigma_{22}}} \right) \right]$$

23

Example 4.1 Density Function

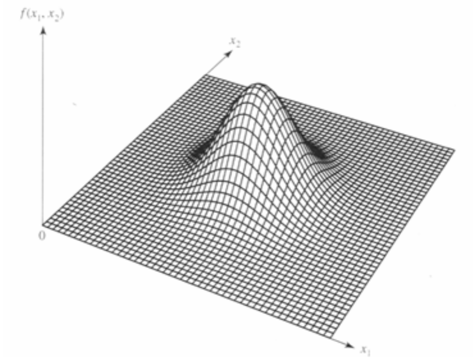
$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}}$$

$$\exp\{-\frac{1}{2(1-\rho_{12}^2)}[\left(\frac{x_1 - \mu_1}{\sqrt{\sigma_{11}}} \right)^2 + \left(\frac{x_2 - \mu_2}{\sqrt{\sigma_{22}}} \right)^2$$

$$-2\rho_{12} \left(\frac{x_1 - \mu_1}{\sqrt{\sigma_{11}}} \right) \left(\frac{x_2 - \mu_2}{\sqrt{\sigma_{22}}} \right)]\}$$

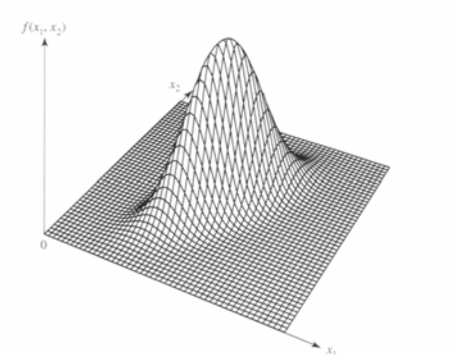
24

Example 4.1 Bivariate Distribution



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

Example 4.1 Bivariate Distribution



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

Contours

Constant probability density contour

$$= \left\{ \text{all } \mathbf{x} \text{ such that } (\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = c^2 \right\}$$

$$= \text{surface of an ellipsoid centered at } \boldsymbol{\mu}$$

$$\text{axes}: \pm c\sqrt{\lambda_i}\,\mathbf{e}_i$$

$$\boldsymbol{\Sigma}\mathbf{e}_i = \lambda_i \mathbf{e}_i, \quad i = 1, 2, \cdots, p$$

Result 4.1

$\boldsymbol{\Sigma}$: positive definite

$$\boldsymbol{\Sigma}\mathbf{e} = \lambda\mathbf{e} \Rightarrow \boldsymbol{\Sigma}^{-1}\mathbf{e} = \frac{1}{\lambda}\mathbf{e}$$

$$(\lambda, \mathbf{e}) \text{ for } \boldsymbol{\Sigma} \Rightarrow (1/\lambda, \mathbf{e}) \text{ for } \boldsymbol{\Sigma}^{-1}$$

$\boldsymbol{\Sigma}^{-1}$ positive definite

Example 4.2 Bivariate Contour

Bivariate normal, $\sigma_{11} = \sigma_{22}$

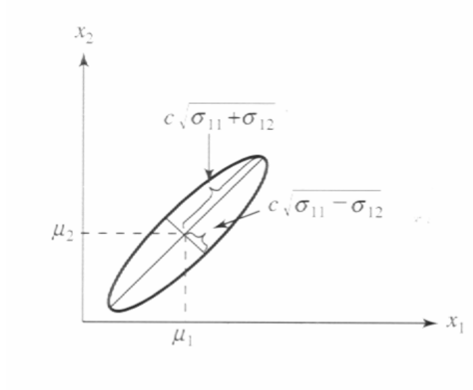
eigenvalues and eigenvectors

$$\lambda_1 = \sigma_{11} + \sigma_{12}, \quad \mathbf{e}_1' = [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]$$

$$\lambda_2 = \sigma_{11} - \sigma_{12}, \quad \mathbf{e}_2' = [\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}]$$

29

Example 4.2 Positive Correlation



30

Probability Related to Squared Distance

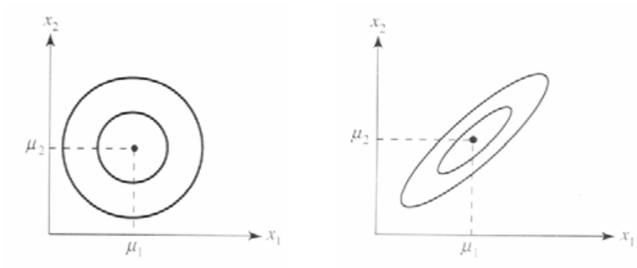
Solid ellipsoid of $\mathbf{x}$ values satisfying

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

has probability $1-\alpha$

31

Probability Related to Squared Distance



32

Result 4.2

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

$\mathbf{a}'\mathbf{X} = a_1X_1 + a_2X_2 + \cdots + a_pX_p :$

$N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$

$\mathbf{a}'\mathbf{X} : N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$ for every $\mathbf{a} \Rightarrow$

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

33

Example 4.3 Marginal Distribution

$\mathbf{X} = [X_1, X_2, \cdots, X_p]' : N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$\mathbf{a}' = [1, 0, \cdots, 0], \quad \mathbf{a}'\mathbf{X} = X_1$

$\mathbf{a}'\boldsymbol{\mu} = \mu_1, \quad \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} = \sigma_{11}$

$\mathbf{a}'\mathbf{X} : N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}) = N(\mu_1, \sigma_{11})$

Marginal distribution of X_i in $\mathbf{X}$:

$N(\mu_i, \sigma_{ii})$

34

Result 4.3

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

$$\mathbf{AX} = \begin{bmatrix} a_{11}X_1 + \cdots + a_{1p}X_p \\ a_{21}X_1 + \cdots + a_{2p}X_p \\ \vdots \\ a_{q1}X_1 + \cdots + a_{qp}X_p \end{bmatrix} : N_q(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$$

$\mathbf{X} + \mathbf{d} : N_p(\boldsymbol{\mu} + \mathbf{d}, \boldsymbol{\Sigma})$

35

Proof of Result 4.3: Part 1

Any linear combination $\mathbf{b}'(\mathbf{AX}) = \mathbf{a}'\mathbf{X}$,

$\mathbf{a} = \mathbf{A}'\mathbf{b} \Rightarrow$

$(\mathbf{b}'\mathbf{A})\mathbf{X} : N((\mathbf{b}'\mathbf{A})\boldsymbol{\mu}, (\mathbf{b}'\mathbf{A})\boldsymbol{\Sigma}(\mathbf{A}'\mathbf{b}))$

$\Rightarrow$

$\mathbf{b}'(\mathbf{AX}) : N(\mathbf{b}'(\mathbf{A}\boldsymbol{\mu}), \mathbf{b}'(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')\mathbf{b})$

valid for every $\mathbf{b} \Rightarrow \mathbf{AX} : N_q(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$

36

Proof of Result 4.3: Part 2

$$\mathbf{a}'(\mathbf{X}+\mathbf{d}) = \mathbf{a}'\mathbf{X}+\mathbf{a}'\mathbf{d}$$

$$\mathbf{a}'\mathbf{X} : N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$$

$$\mathbf{a}'\mathbf{X}+\mathbf{a}'\mathbf{d} : N(\mathbf{a}'\boldsymbol{\mu}+\mathbf{a}'\mathbf{d}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$$

$\mathbf{a}$ is arbitrary $\Rightarrow$

$$\mathbf{X}+\mathbf{d} : N_p(\boldsymbol{\mu}+\mathbf{d}, \boldsymbol{\Sigma})$$

37

Example 4.4 Linear Combinations

$\mathbf{X} : N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$\begin{bmatrix} X_1 - X_2 \\ X_2 - X_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \mathbf{AX}$$

$$\mathbf{A}\boldsymbol{\mu} = \begin{bmatrix} \mu_1 - \mu_2 \\ \mu_2 - \mu_3 \end{bmatrix}$$

$$\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}' = \begin{bmatrix} \sigma_{11} - 2\sigma_{12} + \sigma_{22} & \sigma_{12} + \sigma_{23} - \sigma_{22} - \sigma_{13} \\ \sigma_{12} + \sigma_{23} - \sigma_{22} - \sigma_{13} & \sigma_{22} - 2\sigma_{23} + \sigma_{33} \end{bmatrix}$$

$\mathbf{AX} : N_2(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$

can be verified with $Y_1 = X_1 - X_2$, $Y_2 = X_2 - X_3$

38

Result 4.4

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

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

$\Rightarrow \mathbf{X}_1 : N_q(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$

Proof : Set $\underset{(q\times p)}{\mathbf{A}} = \begin{bmatrix} \underset{(q\times q)}{\mathbf{I}} & | & \underset{(q\times(p-q))}{\mathbf{0}} \end{bmatrix}$ in Result 4.3

39

Example 4.5 Subset Distribution

$\mathbf{X} : N_5(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$\mathbf{X}_1 = \begin{bmatrix} X_2 \\ X_4 \end{bmatrix}, \boldsymbol{\mu}_1 = \begin{bmatrix} \mu_2 \\ \mu_4 \end{bmatrix}, \boldsymbol{\Sigma}_{11} = \begin{bmatrix} \sigma_{22} & \sigma_{24} \\ \sigma_{24} & \sigma_{44} \end{bmatrix}$$

$$\mathbf{X}_1 : N_2\left(\begin{bmatrix} \mu_2 \\ \mu_4 \end{bmatrix}, \begin{bmatrix} \sigma_{22} & \sigma_{24} \\ \sigma_{24} & \sigma_{44} \end{bmatrix} \right)$$

40

Result 4.5

(a) $\underset{(q_1\times 1)}{\mathbf{X}_1}, \underset{(q_2\times 1)}{\mathbf{X}_2}$: independent, $\text{Cov}(\mathbf{X}_1, \mathbf{X}_2) = \underset{(q_1\times q_2)}{\mathbf{0}}$

(b) $\begin{bmatrix} \mathbf{X}_1 \\ --- \\ \mathbf{X}_2 \end{bmatrix} : N_{q_1+q_2}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ --- \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & | & \boldsymbol{\Sigma}_{12} \\ --- & + & --- \\ \boldsymbol{\Sigma}_{21} & | & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right)$

$\Rightarrow \mathbf{X}_1, \mathbf{X}_2$: independent if and only if $\boldsymbol{\Sigma}_{12} = 0$

(c) $\mathbf{X}_1 : N_{q1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}), \mathbf{X}_2 : N_{q2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$ independent

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

41

Example 4.6 Independence

$\mathbf{X} : N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$\boldsymbol{\Sigma} = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

X_1, X_2 : not independent

$\mathbf{X}_1 = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ and X_3 are independent

(X_3 is independent of X_1 and also X_2)

42

Result 4.6

$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ --- \\ \mathbf{X}_2 \end{bmatrix} : N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ --- \\ \boldsymbol{\mu}_2 \end{bmatrix},$

$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & | & \boldsymbol{\Sigma}_{12} \\ --- & + & --- \\ \boldsymbol{\Sigma}_{21} & | & \boldsymbol{\Sigma}_{22} \end{bmatrix}, \quad |\boldsymbol{\Sigma}_{22}| > 0 \Rightarrow$

conditional distribution of $\mathbf{X}_1$ given $\mathbf{X}_2 = \mathbf{x}_2$ is normal with mean $= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$ and covariance $= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$

43

Proof of Result 4.6

$\mathbf{A} = \begin{bmatrix} \mathbf{I} & | & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ --- & + & ---- \\ 0 & | & \mathbf{I} \end{bmatrix},$

$\mathbf{A}(\mathbf{X} - \boldsymbol{\mu}) = \begin{bmatrix} \mathbf{X}_1 - \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2) \\ -------------- \\ \mathbf{X}_2 - \boldsymbol{\mu}_2 \end{bmatrix}:$

joint normal with covariance

$\mathbf{A}\boldsymbol{\Sigma}\,\mathbf{A}' = \begin{bmatrix} \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & | & \mathbf{0}' \\ -------- & + & -- \\ \mathbf{0} & | & \boldsymbol{\Sigma}_{22} \end{bmatrix}$

44

Proof of Result 4.6

$$\mathbf{X}_1-\boldsymbol{\mu}_1-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2-\boldsymbol{\mu}_2) \text{ and } \mathbf{X}_2-\boldsymbol{\mu}_2 \text{ are independent}$$

$$A,B \text{ independent} \Rightarrow P(A\,|\,B)=P(A,B)/P(B)=P(A)$$

$$f(\mathbf{X}_1-\boldsymbol{\mu}_1-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2-\boldsymbol{\mu}_2)=\mathbf{x}_1-\boldsymbol{\mu}_1-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2-\boldsymbol{\mu}_2)\,|$$
$$\mathbf{X}_2=\mathbf{x}_2)=$$

$$f(\mathbf{X}_1-\boldsymbol{\mu}_1-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2-\boldsymbol{\mu}_2)=\mathbf{x}_1-\boldsymbol{\mu}_1-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2-\boldsymbol{\mu}_2))$$

$$\mathbf{X}_1-\boldsymbol{\mu}_1-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2-\boldsymbol{\mu}_2):N_q(0,\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})$$

$$\mathbf{X}_1 \text{ given } \mathbf{X}_2=\mathbf{x}_2:$$

$$N_q(\boldsymbol{\mu}_1+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2-\boldsymbol{\mu}_2),\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})$$

Example 4.7 Conditional Bivariate

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}:N_2(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix},\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix})$$

show that

$$f(x_1\,|\,x_2)=N(\mu_1+\frac{\sigma_{12}}{\sigma_{22}}(x_2-\mu_2),\sigma_{11}-\frac{\sigma_{12}^2}{\sigma_{22}})$$

Example 4.1 Density Function

$$f(x_1,x_2)=\frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}}$$

$$\exp\{-\frac{1}{2(1-\rho_{12}^2)}[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2+\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2$$

$$-2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)]\}$$

Example 4.7

$$\frac{1}{2(1-\rho_{12}^2)}[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2+\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2-2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)]$$

$$=\frac{1}{2\sigma_{11}(1-\rho_{12}^2)}\left(x_1-\mu_1-\frac{\sigma_{12}}{\sigma_{22}}(x_2-\mu_2)\right)^2+\frac{1}{2}\frac{(x_2-\mu_2)^2}{\sigma_{22}}$$

$$2\pi\sqrt{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}=\sqrt{2\pi}\sqrt{\sigma_{11}(1-\rho_{12}^2)}\sqrt{2\pi\sigma_{22}}$$

$$f(x_1\,|\,x_2)=f(x_1,x_2)/f(x_2)$$

$$=\frac{1}{\sqrt{2\pi}\sqrt{\sigma_{11}(1-\rho_{12}^2)}}e^{-\left(x_1-\mu_1-(\sigma_{12}/\sigma_{22})(x_2-\mu_2)\right)^2/2\sigma_{11}(1-\rho_{12}^2)}$$

Result 4.7

$\mathbf{X} : N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad |\boldsymbol{\Sigma}| > 0$

(a) $(\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}) : \chi^2_{\mathrm{p}}$

(b) The probability inside the solid ellipsoid $\{\mathbf{x} : (\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}) \le \chi^2_{\mathrm{p}}(\alpha)\}$ is $1-\alpha$, where $\chi^2_{\mathrm{p}}(\alpha)$ denotes the upper (100α)th percentile of the χ^2_{p} distribution

χ^2 Distribution

$X_1 : N(\mu_1, \sigma_1^2), \quad X_2 : N(\mu_2, \sigma_2^2), \quad \cdots,$

$X_\nu : N(\mu_\nu, \sigma_\nu^2); \quad Z_i = \dfrac{X_i - \mu_i}{\sigma_i} : N(0,1)$

$$\chi^2 = \sum_{i=1}^{\nu}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2, \quad \nu : \text{degrees of freedom (d.f.)}$$

$$f_\nu(\chi^2) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)}\left(\chi^2\right)^{\nu/2-1} e^{-\chi^2/2}, & \chi^2 > 0 \\ 0, & \chi^2 \le 0 \end{cases}$$

(Gamma distribution with $\alpha = \nu/2$)

χ^2 Distribution Curves

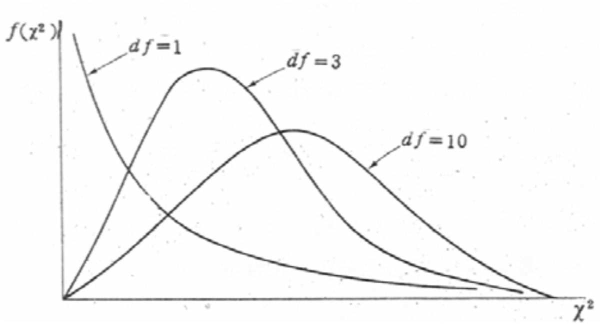
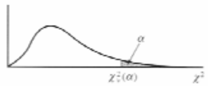


Table 3, Appendix



d.f. ν	α .990	.950	.900	.500	.100	.050	.025	.010	.005
1	.0002	.004	.02	.45	2.71	3.84	5.02	6.63	7.88
2	.02	.10	.21	1.39	4.61	5.99	7.38	9.21	10.60
3	.11	.35	.58	2.37	6.25	7.81	9.35	11.34	12.84
4	.30	.71	1.06	3.36	7.78	9.49	11.14	13.28	14.86
5	.55	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	.87	1.64	2.20	5.35	10.64	12.59	14.45	16.81	18.55
7	1.24	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.65	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.95
9	2.09	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.56	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	3.05	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.57	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	4.11	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.66	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32

Proof of Result 4.7 (a)

$$(\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})=\sum_{i=1}^{p}\frac{1}{\lambda_i}(\mathbf{X}-\boldsymbol{\mu})'\mathbf{e}_i\mathbf{e}_i'(\mathbf{X}-\boldsymbol{\mu})$$

$$=\sum_{i=1}^{p}\left[\frac{1}{\sqrt{\lambda_i}}\mathbf{e}_i'(\mathbf{X}-\boldsymbol{\mu})\right]^2=\sum_{i=1}^{p}Z_i^2,\quad \mathbf{Z}=\mathbf{A}(\mathbf{X}-\boldsymbol{\mu}): N_p(0,\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$$

$$\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'=\begin{bmatrix}\mathbf{e}_1'/\sqrt{\lambda_1}\\ \mathbf{e}_2'/\sqrt{\lambda_2}\\ \vdots\\ \mathbf{e}_p'/\sqrt{\lambda_p}\end{bmatrix}\left[\sum_{i=1}^{p}\lambda_i\mathbf{e}_i\mathbf{e}_i'\right]\begin{bmatrix}\frac{\mathbf{e}_1}{\sqrt{\lambda_1}} & \frac{\mathbf{e}_2}{\sqrt{\lambda_2}} & \cdots & \frac{\mathbf{e}_p}{\sqrt{\lambda_p}}\end{bmatrix}=\mathbf{I}$$

$$Z_i: N(0,1), (\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})=\sum_{i=1}^{p}Z_i^2:\chi_p^2$$

53

Proof of Result 4.7 (b)

$P\left[(\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})\le c^2\right]$ is the probability assigned to the ellipsoid by $\mathbf{X}: N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$

$(\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})$ new random variable distributed by χ_p^2

$$P\left[(\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})\le \chi_p^2(\alpha)\right]=1-\alpha$$

54

Result 4.8

$\mathbf{X}_1,\mathbf{X}_2,\cdots,\mathbf{X}_n$: mutually independent

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

$$\mathbf{V}_1=c_1\mathbf{X}_1+c_2\mathbf{X}_2+\cdots+c_n\mathbf{X}_n : N_p\left(\sum_{j=1}^{n}c_j\boldsymbol{\mu}_j,(\sum_{j=1}^{n}c_j^2)\boldsymbol{\Sigma}\right)$$

$\mathbf{V}_2=b_1\mathbf{X}_1+b_2\mathbf{X}_2+\cdots+b_n\mathbf{X}_n$ and $\mathbf{V}_1$ and $\mathbf{V}_2$ are joint normal

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

55

Proof of Result 4.8

$$\mathbf{X}'=[\mathbf{X}_1',\mathbf{X}_2',\cdots,\mathbf{X}_n']: N_{np}(\boldsymbol{\mu},\boldsymbol{\Sigma}_{\mathbf{X}})$$

$$\boldsymbol{\mu}=\begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\\ \vdots\\ \boldsymbol{\mu}_n\end{bmatrix},\quad \boldsymbol{\Sigma}_{\mathbf{X}}=\begin{bmatrix}\boldsymbol{\Sigma} & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Sigma} & \cdots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}\end{bmatrix}$$

$$\mathbf{A}=\begin{bmatrix}c_1\mathbf{I} & c_2\mathbf{I} & \cdots & c_n\mathbf{I}\\ b_1\mathbf{I} & b_2\mathbf{I} & \cdots & b_n\mathbf{I}\end{bmatrix},\mathbf{A}\mathbf{X}=\begin{bmatrix}\mathbf{V}_1\\ \mathbf{V}_2\end{bmatrix}: N_{2p}(\mathbf{A}\boldsymbol{\mu},\mathbf{A}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{A}')$$

block diagonal terms of $\mathbf{A}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{A}': (\sum_{j=1}^{n}c_j^2)\boldsymbol{\Sigma}, (\sum_{j=1}^{n}b_j^2)\boldsymbol{\Sigma}$

off − diagonal terms of $\mathbf{A}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{A}': (\sum_{j=1}^{n}c_jb_j)\boldsymbol{\Sigma}$

56

Example 4.8 Linear Combinations

$$\mathbf{X}_1,\mathbf{X}_2,\mathbf{X}_3,\mathbf{X}_4 : \text{independent identical } N_3(\boldsymbol{\mu},\boldsymbol{\Sigma})$$

$$\boldsymbol{\mu}=\begin{bmatrix}3\\-1\\1\end{bmatrix},\boldsymbol{\Sigma}=\begin{bmatrix}3&-1&1\\-1&1&0\\1&0&2\end{bmatrix}$$

$$\mathbf{a}'\mathbf{X}_1 : N(\mathbf{a}'\boldsymbol{\mu},\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$$

$$\mathbf{a}'\boldsymbol{\mu}=3a_1-a_2+a_3$$

$$\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}=3a_1^2+a_2^2+2a_3^2-2a_1a_2+2a_1a_3$$

57

Example 4.8 Linear Combinations

$$\mathbf{V}_1=\frac{1}{2}\mathbf{X}_1+\frac{1}{2}\mathbf{X}_2+\frac{1}{2}\mathbf{X}_3+\frac{1}{2}\mathbf{X}_4 : N_3(\boldsymbol{\mu}_{\mathbf{V}_1},\boldsymbol{\Sigma}_{\mathbf{V}_1})$$

$$\boldsymbol{\mu}_{\mathbf{V}_1}=\sum_{j=1}^{4}c_j\boldsymbol{\mu}_j=2\boldsymbol{\mu}=\begin{bmatrix}6\\-2\\2\end{bmatrix}$$

$$\boldsymbol{\Sigma}_{\mathbf{V}_1}=(\sum_{j=1}^{4}c_j^2)\boldsymbol{\Sigma}=\boldsymbol{\Sigma}=\begin{bmatrix}3&-1&1\\-1&1&0\\1&0&2\end{bmatrix}$$

$$\mathbf{V}_2=\mathbf{X}_1+\mathbf{X}_2+\mathbf{X}_3-3\mathbf{X}_4,\quad \text{Cov}(\mathbf{V}_1,\mathbf{V}_2)=(\sum_{j=1}^{4}c_jb_j)\boldsymbol{\Sigma}=\mathbf{0}$$

58

Outline

- Introduction
- The Multivariate Normal Density and Its Properties
- Sampling from a Multivariate Normal Distribution and Maximum Likelihood Estimation
- The Sampling Distribution of $\bar{X}$ and *S*
- Large-Sample Behavior of $\bar{X}$ and *S*

Questions

- What are random samples?
- What is the likelihood?
- How to estimate the mean and variance of a univariate normal distribution by the maximum-likelihood technique? (point estimates)
- What is the multivariate normal likelihood?

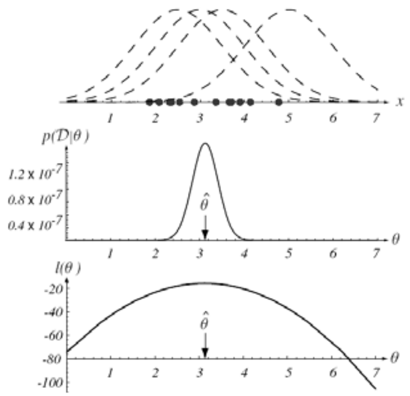
Questions

- What is the trace of a matrix?
- How to compute the quadratic form using the trace of the matrix? (Result 4.9)
- How to express the trace of a matrix by its eigenvalues? (Result 4.9)
- Result 4.10

Questions

- How to estimate the mean and covariance matrix of a multivariate normal vector? (Result 4.11)
- What is the invariance property of the maximum likelihood estimates?
- What is the sufficient statistics?

Maximum-likelihood Estimation



63

Multivariate Normal Likelihood

$\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$: random sample from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$\left\{ \begin{matrix} \text{Joint density of} \\ \mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n \end{matrix} \right\} = \frac{1}{(2\pi)^{np/2} |\boldsymbol{\Sigma}|^{n/2}} e^{-\sum_{j=1}^{n} (\mathbf{x}_j - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_j - \boldsymbol{\mu})/2}$$

as a function of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ for fixed $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$

$\Rightarrow$ likelihood

Maximum likelihood estimation

Maximum likelihood estimates

64

Trace of a Matrix

$$\underset{(k\times k)}{\mathbf{A}} = \{a_{ij}\} \Rightarrow \mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{k} a_{ii}; \quad c \text{ is a scalar}$$

(a) $\mathrm{tr}(c\mathbf{A}) = c\,\mathrm{tr}(\mathbf{A})$

(b) $\mathrm{tr}(\mathbf{A} \pm \mathbf{B}) = \mathrm{tr}(\mathbf{A}) \pm \mathrm{tr}(\mathbf{B})$

(c) $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$

(d) $\mathrm{tr}(\mathbf{B}^{-1}\mathbf{AB}) = \mathrm{tr}(\mathbf{A})$

(e) $\mathrm{tr}(\mathbf{AA}') = \sum_{i=1}^{k}\sum_{j=1}^{k} a_{ij}^2$

65

Result 4.9

$\mathbf{A} : k \times k$ symetric matrix

$\mathbf{x} : k \times 1$ vector

(a) $\mathbf{x}'\mathbf{Ax} = \mathrm{tr}(\mathbf{x}'\mathbf{Ax}) = \mathrm{tr}(\mathbf{Axx}')$

(b) $\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{k} \lambda_i$

66

Proof of Result 4.9 (a)

$\mathbf{B} : m \times k$ matrix, $\quad \mathbf{C} : k \times m$ matrix

$\mathrm{tr}(\mathbf{BC}) = \mathrm{tr}(\mathbf{CB})$

$$\because \mathrm{tr}(\mathbf{BC}) = \sum_{i=1}^{m}\left(\sum_{j=1}^{k} b_{ij}c_{ji}\right)$$

$$\mathrm{tr}(\mathbf{CB}) = \sum_{j=1}^{k}\left(\sum_{i=1}^{m} c_{ji}b_{ij}\right) = \sum_{i=1}^{m}\left(\sum_{j=1}^{k} b_{ij}c_{ji}\right) = \mathrm{tr}(\mathbf{BC})$$

$\Rightarrow \mathrm{tr}(\mathbf{x}'\mathbf{Ax}) = \mathrm{tr}((\mathbf{Ax})\mathbf{x}') = \mathrm{tr}(\mathbf{Axx}')$

67

Proof of Result 4.9 (b)

$\mathbf{A} = \mathbf{P}'\mathbf{\Lambda P}, \quad \mathbf{P}'\mathbf{P} = \mathbf{I}$

$\mathbf{\Lambda} = diag\{\lambda_1, \lambda_2, \cdots, \lambda_k\}$

$\mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\mathbf{P}'\mathbf{\Lambda P})$

$$= \mathrm{tr}(\mathbf{\Lambda PP}') = \mathrm{tr}(\mathbf{\Lambda}) = \sum_{i=1}^{k} \lambda_i$$

68

Likelihood Function

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

$$\sum_{j=1}^{n}(\mathbf{x}_j-\boldsymbol{\mu})(\mathbf{x}_j-\boldsymbol{\mu})'$$

$$=\sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}}+\bar{\mathbf{x}}-\boldsymbol{\mu})(\mathbf{x}_j-\bar{\mathbf{x}}+\bar{\mathbf{x}}-\boldsymbol{\mu})'$$

$$=\sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})(\mathbf{x}_j-\bar{\mathbf{x}})'+n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'$$

$$L(\boldsymbol{\mu},\boldsymbol{\Sigma})=\frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}}e^{-\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})(\mathbf{x}_j-\bar{\mathbf{x}})'+n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'\right)\right]/2}$$

69

Result 4.10

$\mathbf{B}$: $p\times p$ symmetric positive definite matrix

b : positive scalar

$$\frac{1}{|\boldsymbol{\Sigma}|^{b}}e^{-\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{B})/2}\le\frac{1}{|\mathbf{B}|^{b}}(2b)^{pb}e^{-bp}$$

for all positive definite $\underset{(p\times p)}{\boldsymbol{\Sigma}}$, with equality holding only for $\boldsymbol{\Sigma}=(1/2b)\mathbf{B}$

70

Proof of Result 4.10

$$\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{B})=\mathrm{tr}[(\boldsymbol{\Sigma}^{-1}\mathbf{B}^{1/2})\mathbf{B}^{1/2}]=\mathrm{tr}[\mathbf{B}^{1/2}\boldsymbol{\Sigma}^{-1}\mathbf{B}^{1/2}]$$

η_i : eigenvalues of $\mathbf{B}^{1/2}\boldsymbol{\Sigma}^{-1}\mathbf{B}^{1/2}$, all positive

$$\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{B})=\sum_{i=1}^{p}\eta_i,\quad |\boldsymbol{\Sigma}^{-1}\mathbf{B}|=\prod_{i=1}^{p}\eta_i=|\mathbf{B}|/|\boldsymbol{\Sigma}|$$

$$\frac{1}{|\boldsymbol{\Sigma}|^{b}}e^{-\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{B})/2}=\frac{\left(\prod_{i=1}^{p}\eta_i\right)^{b}}{|\mathbf{B}|^{b}}e^{-\sum_{i=1}^{p}\eta_i/2}=\frac{1}{|\mathbf{B}|^{p}}\prod_{i=1}^{p}\eta_i^{b}e^{-\eta_i/2}$$

$\eta^{b}e^{-\eta/2}$ has a maximum $(2b)^{b}e^{-b}$ at $\eta=2b$ $\therefore \frac{1}{|\boldsymbol{\Sigma}|^{b}}e^{-\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{B})/2}\le\frac{1}{|\mathbf{B}|^{b}}(2b)^{pb}e^{-bp}$

upper bound is attained when $\boldsymbol{\Sigma}=(1/2b)\mathbf{B}$ such that $\mathbf{B}^{1/2}\boldsymbol{\Sigma}^{-1}\mathbf{B}^{1/2}=2b\mathbf{I}$

71

Result 4.11 Maximum Likelihood Estimators of μ and Σ

$\mathbf{X}_1,\mathbf{X}_2,\cdots,\mathbf{X}_n$: random sample from $N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$

$$\hat{\boldsymbol{\mu}}=\bar{\mathbf{X}}$$

$$\hat{\boldsymbol{\Sigma}}=\frac{1}{n}\sum_{j=1}^{n}(\mathbf{X}_j-\bar{\mathbf{X}})(\mathbf{X}_j-\bar{\mathbf{X}})'=\frac{n-1}{n}\mathbf{S}$$

72

Proof of Result 4.11

Exponent of $L(\boldsymbol{\mu},\boldsymbol{\Sigma})$:

$$-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})(\mathbf{x}_j-\bar{\mathbf{x}})'\right)\right]-\frac{1}{2}n(\bar{\mathbf{x}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu})$$

$$\Rightarrow \hat{\boldsymbol{\mu}}=\bar{\mathbf{x}}$$

$$L(\hat{\boldsymbol{\mu}},\boldsymbol{\Sigma})=\frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}}e^{-\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})(\mathbf{x}_j-\bar{\mathbf{x}})'\right)\right]}$$

$$\Rightarrow \hat{\boldsymbol{\Sigma}}=\frac{1}{n}\sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})(\mathbf{x}_j-\bar{\mathbf{x}})'=\frac{n-1}{n}\mathbf{S}$$

73

Invariance Property

$\hat{\theta}$: maximum likelihood estimator of θ

$h(\hat{\theta})$: maximum likelihood estimator of $h(\theta)$

Examples:

MLE of $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}=\hat{\boldsymbol{\mu}}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}$

MLE of $\sqrt{\sigma_{ii}}=\sqrt{\hat{\sigma}_{ii}}$

$$\hat{\sigma}_{ii}=\frac{1}{n}\sum_{j=1}^{n}(X_{ji}-\bar{X}_i)^2=\text{MLE of }\mathrm{Var}(X_i)$$

74

Sufficient Statistics

$$\begin{Bmatrix}\text{Joint density of}\\ \mathbf{X}_1,\mathbf{X}_2,\cdots,\mathbf{X}_n\end{Bmatrix}=\frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}}e^{-\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})(\mathbf{x}_j-\bar{\mathbf{x}})'+n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'\right)\right]/2}$$

depends on the whole set of observations $\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_n$ through $\bar{\mathbf{x}}$ and $\mathbf{S}$

$\therefore$ $\bar{\mathbf{x}}$ and $\mathbf{S}$ are sufficient statistics of a multivariate normal population

75

Outline

- Introduction
- The Multivariate Normal Density and Its Properties
- Sampling from a Multivariate Normal Distribution and Maximum Likelihood Estimation
- The Sampling Distribution of $\bar{X}$ and S
- Large-Sample Behavior of $\bar{X}$ and S

Questions

- What is the distribution of sample mean for multivariate normal samples?
- What is the distribution of sample covariance matrix for multivariate normal samples?

Distribution of Sample Mean

$\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$: random sample from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Univariate case : $p = 1$

$\bar{X} : N(\mu, \sigma^2 / n)$

Multivariate case :

$\bar{\mathbf{X}} : N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma} / n)$

cf. Result 4.8

78

Sampling Distribution of S

$\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$: random sample from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Univariate case : $p = 1$

$$(n-1)s^2 = \sum_{j=1}^{n} \left(X_j - \bar{X}\right)^2 : \sigma^2 \chi^2_{n-1}$$

$$(n-1)s^2 = \sigma^2 \sum_{j=1}^{n} Z_j^2, \quad \sigma Z_j : N(0, \sigma^2)$$

Multivariate case :

$\mathbf{Z}_j = \mathbf{X}_j - \bar{\mathbf{X}} : N_p(\mathbf{0}, \boldsymbol{\Sigma})$

$(n\text{-}1)\mathbf{S} = \sum_{j=1}^{n} \mathbf{Z}_j \mathbf{Z}_j'$: Wishart distribution $W_{n-1}((n\text{-}1)\mathbf{S} \mid \boldsymbol{\Sigma})$

79

Wishart Distribution

$$w_{n-1}(\mathbf{A} \mid \boldsymbol{\Sigma}) = \frac{|\mathbf{A}|^{(n-p-2)/2} e^{-\mathrm{tr}\left[\mathbf{A}\boldsymbol{\Sigma}^{-1}\right]/2}}{2^{p(n-1)/2} \pi^{p(p-1)/4} |\boldsymbol{\Sigma}|^{(n-1)/2} \prod_{i=1}^{p} \Gamma\left(\frac{1}{2}(n-i)\right)}$$

$\mathbf{A}$: positive definite

Properties :

$\mathbf{A}_1 : W_{m_1}(\mathbf{A}_1 \mid \boldsymbol{\Sigma}), \quad \mathbf{A}_2 : W_{m_2}(\mathbf{A}_2 \mid \boldsymbol{\Sigma}) \Rightarrow$

$\mathbf{A}_1 + \mathbf{A}_2 : W_{m_1+m_2}(\mathbf{A}_1 + \mathbf{A}_2 \mid \boldsymbol{\Sigma})$

$\mathbf{A} : W_m(\mathbf{A} \mid \boldsymbol{\Sigma}) \Rightarrow \mathbf{CAC}' : W_m(\mathbf{CAC}' \mid \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}')$

80

Outline

- Introduction
- The Multivariate Normal Density and Its Properties
- Sampling from a Multivariate Normal Distribution and Maximum Likelihood Estimation
- The Sampling Distribution of $\overline{X}$ and *S*
- Large-Sample Behavior of $\overline{X}$ and *S*

Questions

- What is the univariate central limit theorem?
- What is the law of large numbers, for the univariate case and the multivariate case? (Result 4.12)
- What is the multivariate central limit theorem? (Result 4.13)

Questions

- What is the limit distribution for the square of statistical distance?

Univariate Central Limit Theorem

X : determined by a large number of independent causes $V_1, V_2, \cdots, V_n$

V_i : random variables having approximately the same variability

$X = V_1 + V_2 + \cdots + V_n$

$\Rightarrow X$ has a nearly normal distribution

$\overline{X}$ is also nearly normal for large sample size

84

Result 4.12 Law of Large Numbers

$Y_1, Y_2, \cdots, Y_n$: independent observations from a population (may not be normal) with $E(Y_i) = \mu$

$\Rightarrow$

$\bar{Y} = \dfrac{Y_1 + Y_2 + \cdots + Y_n}{n}$ converges in probability to μ

That is, for any prescribed $\varepsilon > 0$,

$P[-\varepsilon < \bar{Y}-\mu < \varepsilon] \to 1$ as $n \to \infty$

85

Result 4.12 Multivariate Cases

$\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ independent observations from population (may not be multivariate normal) with mean $E(\mathbf{X}_i) = \boldsymbol{\mu} \Rightarrow$

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

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

86

Result 4.13 Central Limit Theorem

$\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$: independent observation from a population with mean $\boldsymbol{\mu}$ and finite covariance $\boldsymbol{\Sigma}$

$\Rightarrow \sqrt{n}(\bar{\mathbf{X}} - \boldsymbol{\mu})$ is approximately $N_p(\mathbf{0}, \boldsymbol{\Sigma})$ for large sample size $n >> p$

(quite good approximation for moderate n when the parent population is nearly normal)

87

Limit Distribution of Statistical Distance

$\bar{\mathbf{X}}$: nearly $N_p(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma})$ for large sample size $n >> p$

$n(\bar{\mathbf{X}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu})$: approximately χ^2_p for large n-p

$\mathbf{S}$ close to $\boldsymbol{\Sigma}$ with high probability when n is large

$\therefore n(\bar{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu})$: approximately χ^2_p for large n-p

88

Outline

- Assessing the Assumption of Normality
- Detecting Outliers and Cleaning Data
- Transformations to Near Normality

Questions

- How to determine if the samples follow a normal distribution?
- What is the Q-Q plot? Why is it valid?
- How to measure the straightness in a Q-Q plot?

Questions

- How to use Result 4.7 to check if the samples are taken from a multivariate normal population?
- What is the chi-square plot? How to use it?

Q-*Q* Plot

$x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$: observations on X_i

Let $x_{(j)}$ be distict and n moderate to large, e.g., $n \ge 20$

Portion of $x \le x_{(j)}$: $j/n \to (j - \frac{1}{2})/n$

$$P[Z \le q_{(j)}] = \int_{-\infty}^{q_{(j)}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \frac{j - 1/2}{n}$$

Plot $\left(q_{(j)}, x_{(j)}\right)$ to see if they are approximately linear, since $x_{(j)} \approx \sigma q_{(j)} + \mu$ if the data are from a normal distribution

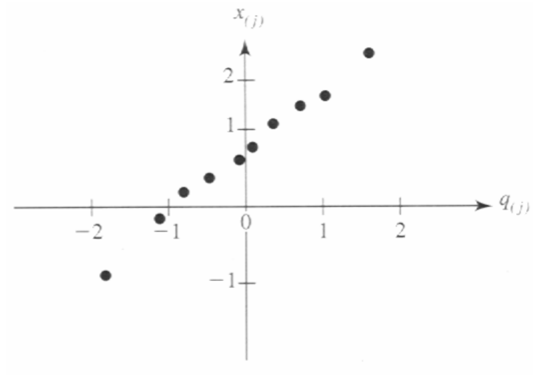
92

Example 4.9

Ordered observations $x_{(j)}$	Probability levels $(j - \frac{1}{2})/n$	Standard normal quantiles $q_{(j)}$
−1.00	.05	−1.645
−.10	.15	−1.036
.16	.25	−.674
.41	.35	−.385
.62	.45	−.125
.80	.55	.125
1.26	.65	.385
1.54	.75	.674
1.71	.85	1.036
2.30	.95	1.645

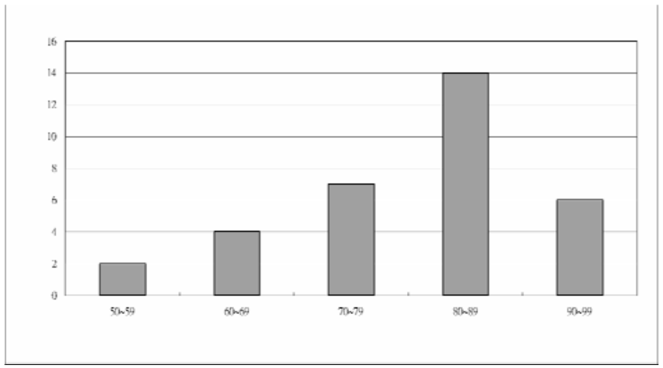
93

Example 4.9



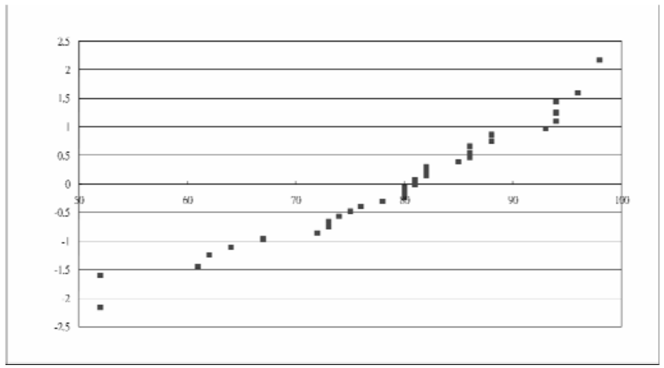
94

Histogram of MidTerm Scores of Students of This Course in 2006



95

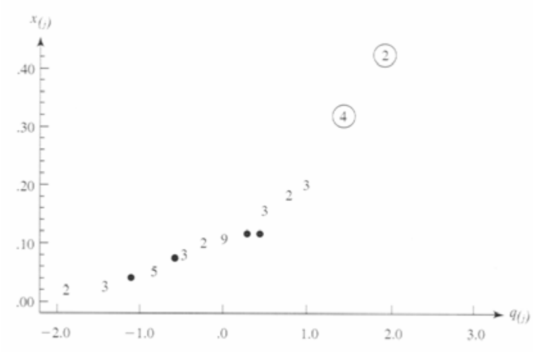
Q-Q Plot of MidTerm Scores of Students of This Course in 2006



n = 33, r_Q = 0.946652

96

Example 4.10 Radiation Data of Closed-Door Microwave Oven



97

Measurement of Straightness

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

Reject the normality hypothesis at level of significance α if r_Q falls below the appropriate value in Table 4.2

98

Table 4.2 *Q-Q* Plot Correlation Coefficient Test

Sample size	Significance levels α		
n	.01	.05	.10
5	.8299	.8788	.9032
10	.8801	.9198	.9351
15	.9126	.9389	.9503
20	.9269	.9508	.9604
25	.9410	.9591	.9665
30	.9479	.9652	.9715
35	.9538	.9682	.9740
40	.9599	.9726	.9771
45	.9632	.9749	.9792
50	.9671	.9768	.9809
55	.9695	.9787	.9822
60	.9720	.9801	.9836
75	.9771	.9838	.9866
100	.9822	.9873	.9895
150	.9879	.9913	.9928
200	.9905	.9931	.9942
300	.9935	.9953	.9960

99

Example 4.11

For data from Example 4.9, $\bar{x} = 0.770, \bar{q} = 0$

$$\sum_{j=1}^{10}\left(x_{(j)} - \bar{x}\right)q_{(j)} = 8.584, \quad \sum_{j=1}^{10}\left(x_{(j)} - \bar{x}\right)^2 = 8.472$$

$$\sum_{j=1}^{10} q_{(j)}^2 = 8.795, \quad r_Q = 0.994$$

$$n = 10, \quad \alpha = 0.10$$

$r_Q > 0.9351 \Rightarrow$ Do not reject normality hypothesis

100

Evaluating Bivariate Normality

Check if roughly 50% of sample observations lie in the ellipse given by

$$\left\{\text{all } \mathbf{x} \text{ such that } (\mathbf{x}-\overline{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x}-\overline{\mathbf{x}}) \le \chi_2^2(0.5)\right\}$$

101

Example 4.12

Company	x_1 = sales (millions of dollars)	x_2 = profits (millions of dollars)	x_3 = assets (millions of dollars)
General Motors	126,974	4,224	173,297
Ford	96,933	3,835	160,893
Exxon	86,656	3,510	83,219
IBM	63,438	3,758	77,734
General Electric	55,264	3,939	128,344
Mobil	50,976	1,809	39,080
Philip Morris	39,069	2,946	38,528
Chrysler	36,156	359	51,038
Du Pont	35,209	2,480	34,715
Texaco	32,416	2,413	25,636

Source: "Fortune 500," *Fortune*, **121** (April 23, 1990), 346–367. © 1990 Time Inc. All rights reserved.

102

Example 4.12

$$\overline{\mathbf{x}} = \begin{bmatrix} 62.309 \\ 2927 \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} 10{,}005.20 & 255.76 \\ 255.76 & 14.30 \end{bmatrix} \times 10^5$$

$$\chi_2^2(0.5) = 1.39$$

$$d^2 = \begin{bmatrix} x_1 - 62.309 \\ x_2 - 2927 \end{bmatrix}' \begin{bmatrix} 0.000184 & -0.003293 \\ -0.003293 & 0.128831 \end{bmatrix} \begin{bmatrix} x_1 - 62.309 \\ x_2 - 2927 \end{bmatrix} \times 10^{-5}$$

$[x_1, x_2]' = [126.974, 4224] \Rightarrow d^2 = 4.34 > 1.39$

Seven out of 10 observations are with $d^2 < 1.39$

Greater than 50% $\Rightarrow$ reject bivariate normality

However, sample size ($n = 10$) is too small to reach the conclusion

103

Chi-Square Plot

$d^2 = (x-\overline{x})S^{-1}(x-\overline{x})$: squared distance

Order the squared distance $d_{(1)}^2 \le d_{(2)}^2 \le \cdots \le d_{(n)}^2$

$q_{c,p}((j-\frac{1}{2})/n)$: $100(j-\frac{1}{2})/n$ quantile of the chi-square distribution with p degrees of freedom

Graph all $(q_{c,p}((j-\frac{1}{2})/n), d_{(j)}^2)$

The plot should resemble a straight line through the origin having slope 1

Note that $q_{c,p}((j-\frac{1}{2})/n) = \chi_p^2(1-(j-\frac{1}{2})/n)$

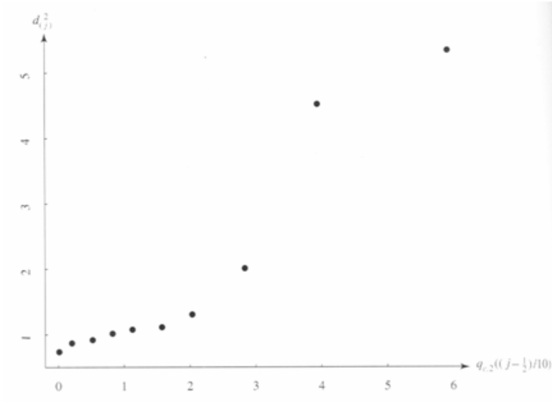
104

Example 4.13 Chi-Square Plot for Example 4.12

j	$d^2_{(j)}$	$q_{c,2}\left(\frac{j-\frac{1}{2}}{10}\right)$
1	.59	.10
2	.81	.33
3	.83	.58
4	.97	.86
5	1.01	1.20
6	1.02	1.60
7	1.20	2.10
8	1.88	2.77
9	4.34	3.79
10	5.33	5.99

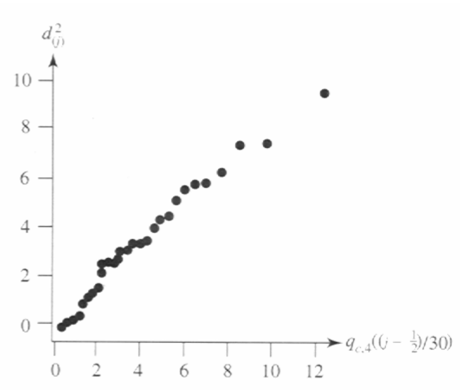
105

Example 4.13 Chi-Square Plot for Example 4.12



106

Chi-Square Plot for Computer Generated 4-variate Normal Data



107

Outline

- Assessing the Assumption of Normality
- Detecting Outliers and Cleaning Data
- Transformations to Near Normality

Steps for Detecting Outliers

- Make a dot plot for each variable
- Make a scatter plot for each pair of variables
- Calculate the standardized values. Examine them for large or small values
- Calculated the squared statistical distance. Examine for unusually large values. In chi-square plot, these would be points farthest from the origin.

109

Outline

- Assessing the Assumption of Normality
- Detecting Outliers and Cleaning Data
- Transformations to Near Normality

Questions

- How to transform sample counts, proportion, and correlation, such that the new variable is more near to a univariate normal distribution?
- What is Box and Cox's univariate transformation?
- How to extend Box and Cox's transformation to the multivariate case?

Questions

- How to deal with data including large negative values?

Helpful Transformation to Near Normality

Original Scale	Transformed Scale
Counts, y	$\sqrt{y}$
Proportions, $\hat{p}$	$\text{logit}(\hat{p}) = \frac{1}{2}\log\left(\frac{\hat{p}}{1-\hat{p}}\right)$
Correlations, r	Fisher's $z(r) = \frac{1}{2}\log\left(\frac{1+r}{1-r}\right)$

113

Box and Cox's Univariate Transformations

$$x^{(\lambda)} = \begin{cases} \dfrac{x^{\lambda}-1}{\lambda}, & \lambda \neq 0 \\ \ln x, & \lambda = 0 \end{cases}$$

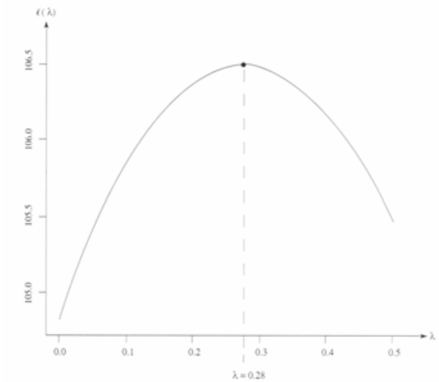
Choose λ to maximize

$$\ell(\lambda) = -\frac{n}{2}\ln\left[\frac{1}{n}\sum_{j=1}^{n}\left(x_j^{(\lambda)} - \overline{x^{(\lambda)}}\right)^2\right] + (\lambda-1)\sum_{j=1}^{n}\ln x_j$$

$$\overline{x^{(\lambda)}} = \frac{1}{n}\sum_{j=1}^{n} x_j^{(\lambda)}$$

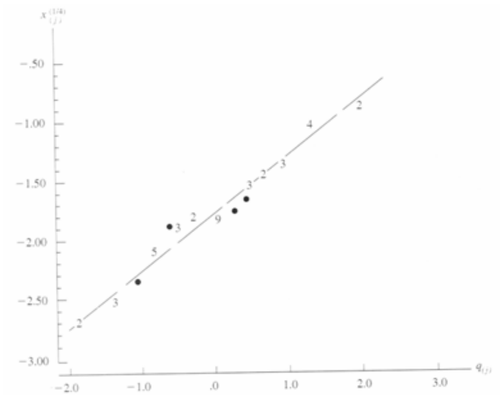
114

Example 4.16 $\ell(\lambda)$ vs. λ



115

Example 4.16 *Q-Q* Plot



116

Transforming Multivariate Observations

$\lambda_1, \lambda_2, \cdots, \lambda_p$: power transformations for the p characteristics

Select λ_k to maximize

$$\ell_{\mathrm{k}}(\lambda) = -\frac{n}{2}\ln\left[\frac{1}{n}\sum_{j=1}^{n}\left(x_{jk}^{(\lambda)} - \overline{x_k^{(\lambda)}}\right)^2\right] + (\lambda-1)\sum_{j=1}^{n}\ln x_{jk}$$

$$\overline{x_k^{(\lambda)}} = \frac{1}{n}\sum_{j=1}^{n} x_{jk}^{(\lambda)}$$

$$\mathbf{x}_j^{(\hat{\lambda})}{}' = \begin{bmatrix} \dfrac{x_{j1}^{(\hat{\lambda}_1)} - 1}{\hat{\lambda}_1} & \dfrac{x_{j2}^{(\hat{\lambda}_2)} - 1}{\hat{\lambda}_2} & \cdots & \dfrac{x_{jp}^{(\hat{\lambda}_p)} - 1}{\hat{\lambda}_p} \end{bmatrix}$$

More Elaborate Approach

$\lambda_1, \lambda_2, \cdots, \lambda_p$: power transformations for the p characteristics

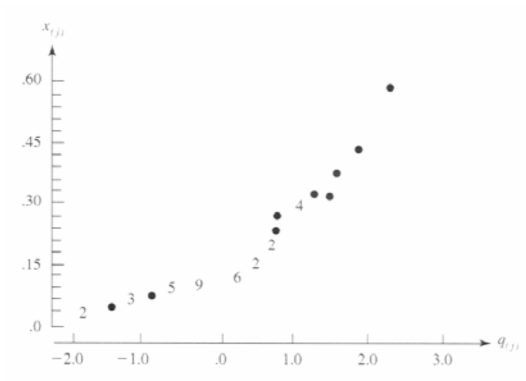
Select $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \cdots, \lambda_p]'$ to maximize

$$\ell(\lambda_1, \lambda_2, \cdots, \lambda_p) = -\frac{n}{2}\ln|\mathbf{S}(\boldsymbol{\lambda})| + \sum_{k=1}^{p}(\lambda_k - 1)\sum_{j=1}^{n}\ln x_{jk}$$

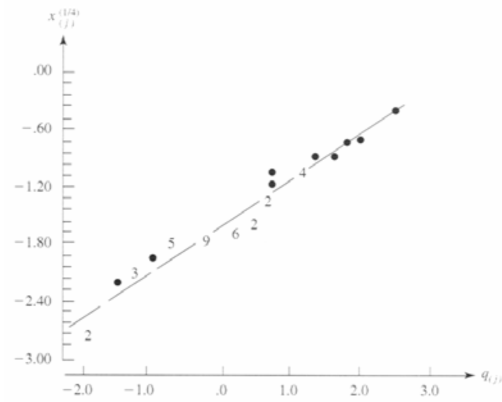
$\mathbf{S}(\boldsymbol{\lambda})$ is computed from

$$\mathbf{x}_j^{(\boldsymbol{\lambda})}{}' = \begin{bmatrix} \dfrac{x_{j1}^{(\lambda_1)} - 1}{\lambda_1} & \dfrac{x_{j2}^{(\lambda_2)} - 1}{\lambda_2} & \cdots & \dfrac{x_{jp}^{(\lambda_p)} - 1}{\lambda_p} \end{bmatrix}$$

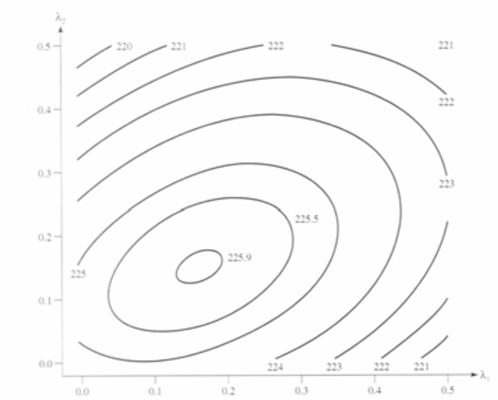
Example 4.17 Original *Q-Q* Plot for Open-Door Data



Example 4.17 *Q-Q* Plot of Transformed Open-Door Data



Example 4.17 Contour Plot of $\ell(\lambda_1, \lambda_2)$ for Both Radiation Data



121

Transform for Data Including Large Negative Values

$$x^{(\lambda)} = \begin{cases} \{(x+1)^{\lambda} - 1\}/\lambda & x \ge 0, \lambda \ne 0 \\ \log(x+1) & x \ge 0, \lambda = 0 \\ -\{(-x+1)^{2-\lambda} - 1\}/(2-\lambda) & x < 0, \lambda \ne 2 \\ -\log(-x+1) & x < 0, \lambda = 2 \end{cases}$$

122