—— WORKSHOP ——

Applied Classical and Modern Multivariate Statistical Analysis

Module 3: Multivariate Normal and Related Distributions

Weixing Song, Juan Du

Department of Statistics
Kansas State University

August 5, 2018

Outline

Outline

Univariate Normal Density

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{\sigma^2}}, \quad x \in \mathbb{R}, \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+.$$

If a random variable X has the above density function, then $X \sim N(\mu, \sigma^2)$.

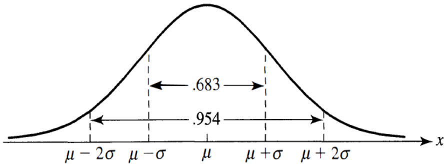


Figure 4.1 A normal density with mean μ and variance σ^2 and selected areas under the curve.

Multivariate Normal Density

Let $X = (X_1, \ldots, X_p)$ be a p-dimensional vector.

We say $X \sim N_p(\mu, \Sigma)$, if it's density function has the form

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

where

- $-\infty < x_i < \infty, i = 1, 2, \ldots, p$;
- $-\infty < \mu_i < \infty, i = 1, 2, \ldots, p$;
- $\Sigma > 0$.

Example (Bivariate normal density) Let $p = 2$.

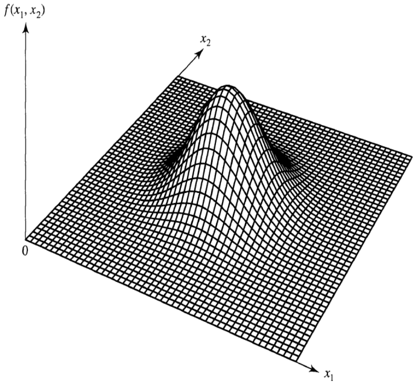
- $\mu_1 = EX_1$, $\mu_2 = EX_2$;
- $\sigma_{11} = \text{Var}(X_1)$, $\sigma_{22} = \text{Var}(X_2)$, $\sigma_{12} = \text{Cov}(X_1, X_2)$;
- $\rho_{12} = \text{Corr}(X_1, X_2)$.

The density function of $X = (X_1, X_2)'$ has the form

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

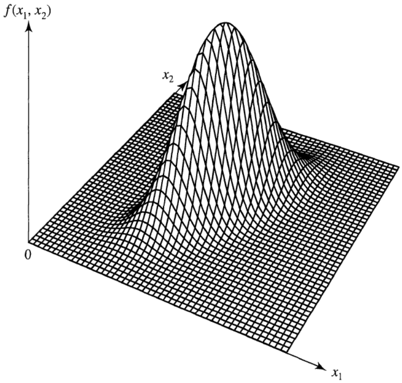
Example (Bivariate Normal Density Plots)

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



Example (Bivariate Normal Density Plots)

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



Contours of Normal Density

Contours of Constant Normal Density

Contours of constant density for the p-dimensional normal distribution are ellipsoids defined by $\mathbf{x}$ such that

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

These ellipsoids are centered at μ and have axes $\pm c\sqrt{\lambda_i}e_i$, where $\Sigma e_i=\lambda_i e_i$, $i=1,2,\ldots,p$.

Example (Contours of the bivariate normal density) Suppose we have a bivariate normal distribution with $\sigma_{11}=\sigma_{22}$. The axes of the contour ellipses of constant density are determined by

$$\pm c\sqrt{\sigma_{11}+\sigma_{12}}\begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix} \quad \text{and} \quad \pm c\sqrt{\sigma_{11}-\sigma_{12}}\begin{bmatrix}\frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}}\end{bmatrix}.$$

Outline

Linear Combination of Random Variables from An MVN Random Vector

If X is distributed as $N_p(\mu, \Sigma)$, then any linear combination of variables $\mathbf{a}'X = a_1X_1 + \cdots + a_pX_p$ is distributed as $N(\mathbf{a}'\mu, \mathbf{a}'\Sigma\mathbf{a})$. Also, if $\mathbf{a}'X$ is distributed as $N(\mathbf{a}'\mu, \mathbf{a}'\Sigma\mathbf{a})$ for every $\mathbf{a}$, then X must be $N_p(\mu, \Sigma)$.

Several Linear Combinations of Random Variables from An MVN Random Vector

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

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

are distributed as $N_q(A\mu, A\Sigma A')$. Also for a $p \times 1$ vector $\mathbf{d}$, $X + d$ is distributed as $N_p(\mu + d, \Sigma)$.

Subvector of An MVN Random Vector

All subsets of a multivariate normal random vector X are normally distributed. If we respectively partition $\mathbf{X}$, its mean vector μ, and its covariance matrix Σ as

$$\mathbf{X}_{p\times 1} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}, \quad \mu_{p\times 1} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix},$$

and

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

then $\mathbf{X}_1$ is distributed as $N_q(\mu_1, \Sigma_{11})$.

Independence of MVN Random Vectors

(a). If $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent, then $\mathrm{Cov}(X_1, X_2) = 0$.

(b). If

$$\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \sim N_{q_1+q_2}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right),$$

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

Conditional Distributions

Let

$$\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \sim N_{q_1+q_2}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right)$$

and $\Sigma_{22} > 0$. Then the conditional distribution of $\mathbf{X}_1$, given that $\mathbf{X}_2 = \mathbf{x}_2$, is normal and has

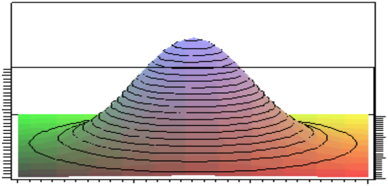
$$\begin{aligned} \text{Mean} &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{X}_2 - \mu_2), \\ \text{Covariance} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \end{aligned}$$

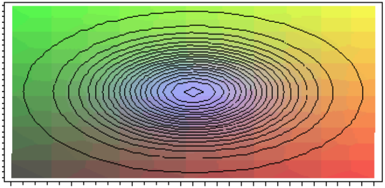
Quadratic Forms of MVN Random Vectors

Let $\mathbf{X}$ be distributed as $N_p(\mu, \Sigma)$ with $|\Sigma| > 0$. Then

(a). $(\mathbf{X} - \mu)'\Sigma^{-1}(\mathbf{X} - \mu)$ is distributed as χ_p^2, where χ_p^2 denotes the chi-square distribution with p degrees of freedom.

(b). The $N_p(\mu, \Sigma)$ distribution assigns probability $1 - \alpha$ to the solid ellipsoid $\{\mathbf{x} : (\mathbf{x} - \mu)'\Sigma^{-1}(\mathbf{x} - \mu) \leq \chi_p^2(\alpha)\}$, where $\chi_p^2(\alpha)$ denotes the upper 100α-th percentile of the χ_p^2 distribution.





Now let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be n random vectors with dimension $p \times 1$.

Consider the following linear combinations

$$V_1 = c_1\mathbf{X}_1 + c_2\mathbf{X}_2 + \cdots + c_n\mathbf{X}_n, \quad V_2 = b_1\mathbf{X}_1 + b_2\mathbf{X}_2 + \cdots + b_n\mathbf{X}_n$$

Linear Combinations of Several MVN Random Vectors

Suppose $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are mutually independent with $\mathbf{X}_j \sim N_p(\mu_j, \Sigma)$. Then

$$V_1 \sim N_p\left(\sum_{j=1}^n c_j\mu_j, \left(\sum_{j=1}^n c_j^2\right)\Sigma\right).$$

Moreover,

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \sim N_{2p}\left(\begin{bmatrix} \sum_{j=1}^n c_j\mu_j \\ \sum_{j=1}^n b_j\mu_j \end{bmatrix}, \begin{bmatrix} \left(\sum_{j=1}^n c_j^2\right)\Sigma & (b'c)\Sigma \\ (b'c)\Sigma & \left(\sum_{j=1}^n b_j^2\right)\Sigma \end{bmatrix}\right)$$

Outline

The Multivariate Normal Likelihood

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be n $p \times 1$ vectors from $N_p(\mu, \Sigma)$.

The joint density function has the form of

$$\prod_{j=1}^{n} \left\{ \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{-(\mathbf{x}_j-\mu)'\Sigma^{-1}(\mathbf{x}_j-\mu)/2} \right\} \tag{1}$$

$$= \frac{1}{(2\pi)^{np/2}|\Sigma|^{n/2}} e^{-\sum_{j=1}^{n}(\mathbf{x}_j-\mu)'\Sigma^{-1}(\mathbf{x}_j-\mu)/2} \tag{2}$$

As a function of μ and Σ, the joint density function is called the likelihood function. Denoted by $L(\mu, \Sigma)$.

The parameter values that maximize the likelihood function is called the maximum likelihood estimates (MLE).

MLE of μ and Σ

MLE of μ and Σ

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

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

are the maximum likelihood estimators of μ and Σ, respectively. Their observed values $\bar{\mathbf{x}}$ and $(1/n)\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'$ are called the maximum likelihood estimates of μ and Σ.

Remarks: The value of the maximum likelihood is

$$L(\hat{\mu}, \hat{\Sigma}) = \frac{1}{(2\pi e)^{np/2}}|\hat{\Sigma}|^{-n/2} = \text{constant} \times (\text{Genralized Variance})^{-n/2}.$$

Outline

1. Multivariate Normal Density
2. Properties of Multivariate Normal Distribution
3. Maximum Likelihood Estimation
4. The Sampling Distribution of MLE
5. Assessing the Normality Assumption
6. Detecting Outliers

Wishart Distribution

Definition: Suppose $\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_m$ are i.i.d. from $N_p(0, \Sigma)$. Then

$$\sum_{j=1}^{m} \mathbf{Z}_i \mathbf{Z}_i' \sim W_m(\cdot|\Sigma),$$

where $W_m(\cdot|\Sigma)$ denotes the Wishart distribution with degrees of freedom m.

Properties:

- If $\mathbf{X}$ is distributed as $W_{m_1}(\mathbf{X}|\Sigma)$, $\mathbf{Y}$ is distributed as $W_{m_2}(\mathbf{Y}|\Sigma)$, $\mathbf{X}$ and $\mathbf{Y}$ are independent, then $\mathbf{X} + \mathbf{Y}$ is distributed as $W_{m_1+m_2}(X_1 + X_2|\Sigma)$.
- If $\mathbf{X}$ is distributed as $W_m(\mathbf{X}|\Sigma)$, then $\mathbf{CXC}'$ is distributed as $W_m(\mathbf{CXC}'|\mathbf{C}\Sigma\mathbf{C}')$.

Sampling Distribution of $\bar{\mathbf{X}}$ and $\mathbf{S}$

Sampling Distribution of MLE

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be random sample of size n from a p-variate normal distribution with mean μ and covariance matrix Σ. Then

- $\bar{\mathbf{X}}$ is distributed as $N_p(\mu, \Sigma/n)$.
- $(n-1)\mathbf{S}$ is distributed as a Wishart distribution with $n-1$ degrees of freedom.
- $\bar{\mathbf{X}}$ and $\mathbf{S}$ are independent.

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

Law of Large Numbers

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be independent observations from a population $\mathbf{X}$ with mean $E\mathbf{X} = \mu$. Then $\bar{\mathbf{X}}$ converges in probability to μ as n increases as $n \to \infty$.

The Central Limit Theorem

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be independent observations from a population $\mathbf{X}$ with mean $E\mathbf{X} = \mu$ and finite covariance Σ. Then $\sqrt{n}(\bar{\mathbf{X}} - \mu)$ has an approximate $N_p(0, \Sigma)$ distribution for large sample sizes. here n should also be large relative to p.

As a consequence of the central limit theorem, we have

$$n(\bar{\mathbf{X}} - \mu)'\mathbf{S}^{-1}(\bar{\mathbf{X}} - \mu) \quad \text{is approximately} \quad \chi^2_p.$$

Outline

Evaluating Normality: Univariate Case

Evaluating Normality: Univariate Case

- Dot diagrams, Histogram: Symmetry?
- Numerical Checking: 3σ-rule.

  When the sample size is large,

  - $[\bar{x}_i - \sqrt{s_{ii}}, \bar{x}_i + \sqrt{s_{ii}}]$ has 68.3% observations;
  - $[\bar{x}_i - 2\sqrt{s_{ii}}, \bar{x}_i + 2\sqrt{s_{ii}}]$ has 95.4% observations;
  - $[\bar{x}_i - 3\sqrt{s_{ii}}, \bar{x}_i + 3\sqrt{s_{ii}}]$ has 99.7% observations;
- QQ-plot, PP-plot, etc.

Straightness Checking

The straightness of the QQ-plot can be measured by the correlation coefficient for the QQ-plot:

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

The hypothesis of normality will be rejected at level of significance α if r_Q falls below the appropriate value in the following Table.

Table 4.2 Critical Points for the Q-Q Plot Correlation Coefficient Test for Normality

Sample size n	Significance levels α .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

Assessing Normality: Multivariate Case

Note that under the normality assumption

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

has probability $1 - \alpha$.

So, we expect roughly $1 - \alpha$ of sample observations to lie in the ellipsoid given by

$$\{\mathbf{x} : (\mathbf{x} - \bar{\mathbf{x}})'S^{-1}(\mathbf{x} - \bar{\mathbf{x}}) \leq \chi_p^2(\alpha)\}.$$

Example (Checking bivariate normality): Consider the following data set

The World's 10 Largest Companies[1]

Company	x_1 = sales (billions)	x_2 = profits (billions)	x_3 = assets (billions)
Citigroup	108.28	17.05	1,484.10
General Electric	152.36	16.59	750.33
American Intl Group	95.04	10.91	766.42
Bank of America	65.45	14.14	1,110.46
HSBC Group	62.97	9.52	1,031.29
ExxonMobil	263.99	25.33	195.26
Royal Dutch/Shell	265.19	18.54	193.83
BP	285.06	15.73	191.11
ING Group	92.01	8.10	1,175.16
Toyota Motor	165.68	11.13	211.15

[1]From www.Forbes.com partially based on *Forbes* The Forbes Global 2000, April 18, 2005.

Example 4.12 (Checking bivariate normality) Although not a random sample, data consisting of the pairs of observations (x_1 = sales, x_2 = profits) for the 10 largest companies in the world are listed in Exercise 1.4. These data give

$$\bar{\mathbf{x}} = \begin{bmatrix} 155.60 \\ 14.70 \end{bmatrix}, \qquad \mathbf{S} = \begin{bmatrix} 7476.45 & 303.62 \\ 303.62 & 26.19 \end{bmatrix}$$

so

$$\mathbf{S}^{-1} = \frac{1}{103{,}623.12} \begin{bmatrix} 26.19 & -303.62 \\ -303.62 & 7476.45 \end{bmatrix}$$

$$= \begin{bmatrix} .000253 & -.002930 \\ -.002930 & .072148 \end{bmatrix}$$

From Table 3 in the appendix, $\chi_2^2(.5) = 1.39$. Thus, any observation $\mathbf{x}' = [x_1, x_2]$ satisfying

$$\begin{bmatrix} x_1 - 155.60 \\ x_2 - 14.70 \end{bmatrix}' \begin{bmatrix} .000253 & -.002930 \\ -.002930 & .072148 \end{bmatrix} \begin{bmatrix} x_1 - 155.60 \\ x_2 - 14.70 \end{bmatrix} \le 1.39$$

is on or inside the estimated 50% contour. Otherwise the observation is outside this contour.

The first pair of observations in Exercise 1.4 is $[x_1, x_2]' = [108.28, 17.05]$. In this case

$$\begin{bmatrix} 108.28 - 155.60 \\ 17.05 - 14.70 \end{bmatrix}' \begin{bmatrix} .000253 & -.002930 \\ -.002930 & .072148 \end{bmatrix} \begin{bmatrix} 108.28 - 155.60 \\ 17.05 - 14.70 \end{bmatrix}$$
$$= 1.61 > 1.39$$

and this point falls outside the 50% contour. The remaining nine points have generalized distances from $\bar{\mathbf{x}}$ of .30, .62, 1.79, 1.30, 4.38, 1.64, 3.53, 1.71, and 1.16, respectively. Since four of these distances are less than 1.39, a proportion, .40, of the data falls within the 50% contour. If the observations were normally distributed, we would expect about half, or 5, of them to be within this contour. This difference in proportions might ordinarily provide evidence for rejecting the notion of bivariate normality; however, our sample size of 10 is too small to reach this conclusion.

Chi-Square Plot

- For each observation $\mathbf{x}_j$, calculate

$$d_j^2 = (\mathbf{x}_j - \bar{\mathbf{x}})' S^{-1} (\mathbf{x}_j - \bar{\mathbf{x}}), \quad j = 1, 2, \ldots, n$$

- Order d_j^2 from smallest to the largest as

$$d_{(1)}^2 \leq d_{(2)}^2 \leq \cdots \leq d_{(n)}^2.$$

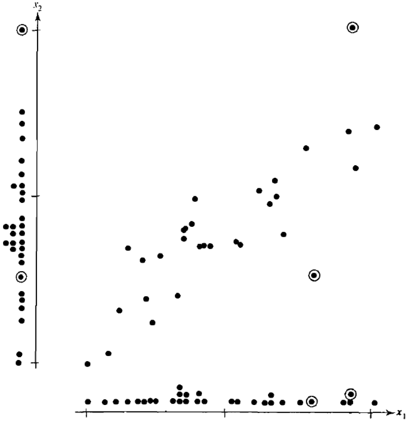
- Caculate $q_{c,p}((j-0.5)/n)$ for $j = 1, 2, \ldots, n$, where $q_{c,p}((j-0.5)/n)$ is the $100(j-0.5)/n$ quantile of the χ_p^2.
- Graph the pairs $(q_{c,p}((j-0.5)/n), d_j^2)$, $j = 1, 2, \ldots, n$.

If the normality assumption holds, then the plot should resemble a straight line through the origin having slope 1.

Outline

What are outliers?

Outliers are unusual observations that do not seem to belong to the pattern of variability produced by the other observations.



Detecting Multidimensional Outliers

- Graphical Tools
  dot plots, scatter plots,...
- Numerical Tools
  - Standardized values
    Calculate

$$z_{jk} = \frac{x_{jk} - \bar{x}_k}{\sqrt{s_{kk}}}, \quad j = 1, 2, \ldots, n; \quad k = 1, 2, \ldots, p.$$

    Guidelines: the observations with standardized values less than $-3(3.5)$ or greater than $3(3.5)$ might be considered as outliers.
  - Generalized squared distances
    Calculate

$$d_j = (\mathbf{x}_j - \mathbf{x})'\mathbf{S}^{-1}(\mathbf{x}_j - \mathbf{x}), \quad j = 1, 2, \ldots, n.$$

    Guideline: Very large d_j's imply outliers.