STT 843: Multivariate Analysis

6. The Multivariate Normal Distribution (Chapter 4.4-4.6

Guanqun Cao

Department of Statistics and Probability
Michigan State University

Spring 2026

Outline

1. The Sampling Distribution of $\bar{X}$ and S
2. Large-Sample Behavior of $\bar{X}$ and S
3. Assessing the Assumption of Normality

Sufficient Statistics

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be a random sample from a multivariate normal population with mean μ and covariance Σ. Then

$$\overline{\boldsymbol{X}} \text{ and } \mathbf{S} = \frac{1}{n-1} \sum_{j=1}^{n} \left(\boldsymbol{X}_j - \overline{\boldsymbol{X}}\right) \left(\boldsymbol{X}_j - \overline{\boldsymbol{X}}\right)^{\mathsf{T}}$$

are sufficient statistics

- The importance of sufficient statistics for normal populations is that all of the information about $\boldsymbol{\mu}$ and Σ in the data matrix $\boldsymbol{X}$ is contained in $\overline{\boldsymbol{X}}$ and S, regardless of the sample size n.
- This generally is not true for nonnormal populations.
- Since many multivariate techniques begin with sample means and covariances, it is prudent to check on the adequacy of the multivariate normal assumption.
- If the data cannot be regarded as multivariate normal, techniques that depend solely on $\overline{\boldsymbol{X}}$ and $\mathbf{S}$ may be ignoring other useful sample information.

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

- The univariate case ($p = 1$)
- $\bar{X}$ is normal with mean $\mu =$ (population mean) and variance

$$\frac{1}{n}\sigma^2 = \frac{\text{population variance}}{\text{sample size}}$$

For the sample variance, recall that $(n-1)s^2 = \sum_{j=1}^{n}\left(X_j - \bar{X}\right)^2$ is distributed as σ^2 times a chi-square variable having $n-1$ degrees of freedom (d.f.).

- The chi-square is the distribution of a sum squares of independent standard normal random variables. That is, $(n-1)s^2$ is distributed as

$$\sigma^2\left(Z_1^2+\cdots+Z_{n-1}^2\right)=(\sigma Z_1)^2+\cdots+(\sigma Z_{n-1})^2 .$$

The individual terms σZ_i are independently distributed as $N(0,\sigma^2)$.

Wishart distribution

$$
\begin{aligned}
W_m(\cdot \mid \Sigma) &= \text{ Wishart distribution with } m \text{ d.f.} \\
&= \text{ distribution of } \sum_{j=1}^{n} \mathbf{Z}_j \mathbf{Z}_j^{\mathsf{T}}
\end{aligned}
$$

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

Properties of the Wishart Distribution

1. If $\mathbf{A}_1$ is distributed as $W_{m_1}(\mathbf{A}_1 \mid \Sigma)$ independently of $\mathbf{A}_2$, which is distributed as $W_{m_2}(\mathbf{A}_2 \mid \Sigma)$, then $\mathbf{A}_1 + \mathbf{A_2}$ is distributed as $W_{m_1+m_2}(\mathbf{A}_1 + \ \mathbf{A}_2 \mid \Sigma)$. That is, the the degree of freedom add.
2. If $\mathbf{A}$ is distributed as $W_m(\mathbf{A} \mid \Sigma)$, then $\mathbf{CAC}^{\mathsf{T}}$ is distributed as $W_m\left(\mathbf{CAC}' \mid \mathbf{C}\Sigma\mathbf{C}^{\mathsf{T}}\right)$.

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

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be a random sample size n from a p-variate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ. Then

1. $\overline{\boldsymbol{X}}$ is distributed as $N_p\left(\boldsymbol{\mu}, \frac{1}{n}\Sigma\right)$.
2. $(n-1)\mathbf{S}$ is distributed as a Wishart random matrix with $n-1$ d.f.
3. $\overline{\boldsymbol{X}}$ and $\mathbf{S}$ are independent.

Law of Large numbers

Let $Y_1, Y_2, \ldots, Y_n$ be independent observations from a population with mean $\mathrm{E}(Y_i) = \mu$, then

$$\bar{Y} = \frac{Y_1 + Y_2 + \cdots + Y_n}{n}$$

converges in probability to μ as n increases without bound. That is, for any prescribed accuracy $\varepsilon > 0$, $P[-\varepsilon < \bar{Y} - \mu < \varepsilon]$ approaches unity as $n \to \infty$.

The central limit theorem

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

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

for large sample sizes. Here n should also be large relative to p.

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be independent observations from a population with mean $\boldsymbol{\mu}$ and finite (nonsingular) covariance Σ. Then

$$\sqrt{n}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}) \text{ is approximately } N_p(0, \Sigma)$$

and

$$n(\overline{\boldsymbol{X}} - \boldsymbol{\mu})^{\mathsf{T}} \mathbf{S}^{-1} (\overline{\boldsymbol{X}} - \boldsymbol{\mu}) \text{ is approximately } \chi_p^2$$

for $n - p$ large.

Assessing the Assumption of Normality

- Most of the statistical techniques discussed assume that each vector observation $\boldsymbol{X}_j$ comes from a multivariate normal distribution.
- In situations where the sample size is large and the techniques dependent solely on the behavior of $\overline{\boldsymbol{X}}$, or distances involve $\bar{X}$ of the form $n(\overline{\boldsymbol{X}} - \boldsymbol{\mu})^{\top}\mathbf{S}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})$, the assumption of normality for the individual observations is less crucial.
- But to some degree, the quality of inferences made by these methods depends on how closely the true parent population resembles the multivariate normal form.

Therefore, we address these questions:

1. Do the marginal distributions of the elements of $\boldsymbol{X}$ appear to be normal ? What about a few linear combinations of the components X_j ?
2. Do the scatter plots of observations on different characteristics give the elliptical appearance expected from normal population?
3. Are there any "wild" observations that should be checked for accuracy?

- Dot diagrams for smaller n and histogram for $n > 25$ or so help reveal situations where one tail of a univariate distribution is much longer than other.
- If the histogram for a variable X_i appears reasonably symmetric, we can check further by counting the number of observations in certain interval, for examples

A univariate normal distribution assigns probability 0.683 to the interval

$$(\mu_i - \sqrt{\sigma_{ii}}, \mu_i + \sqrt{\sigma_{ii}})$$

and probability 0.954 to the interval

$$(\mu_i - 2\sqrt{\sigma_{ii}}, \mu_i + 2\sqrt{\sigma_{ii}})$$

Consequently, with a large same size n, the observed proportion $\hat{p}_{i1}$ of the observations lying in the interval $\left(\bar{x}_i - \sqrt{s_{ii}}, \bar{x}_i + \sqrt{s_{ii}}\right)$ to be about 0.683 , and the interval $\left(\bar{x}_i - 2\sqrt{s_{ii}}, \bar{x}_i + 2\sqrt{s_{ii}}\right)$ to be about 0.954

Using the normal approximating to the sampling of $\hat{p}_i$, observe that either

$$|\hat{p}_{i1} - 0.683| > 3\sqrt{\frac{(0.683)(0.317)}{n}} = \frac{1.396}{\sqrt{n}}$$

or

$$|\hat{p}_{i2} - 0.954| > 3\sqrt{\frac{(0.954)(0.046)}{n}} = \frac{0.628}{\sqrt{n}}$$

would indicate departures from an assumed normal distribution for the i th characteristic.

- Plots are always useful devices in any data analysis. Special plots called $Q - Q$ plots can be used to assess the assumption of normality.

Let $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ represent these observations after they are ordered according to magnitude. For a standard normal distribution, the quantiles $q_{(j)}$ are defined by the relation

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

Here $p_{(j)}$ is the probability of getting a value less than or equal to $q_{(j)}$ in a single drawing from a standard normal population.

- The idea is to look at the pairs of quantiles $\left(q_{(j)}, x_{(j)}\right)$ with the same associated cumulative probability $\left(j - \frac{1}{2}\right)/n$. If the data arise from a normal population, the pairs $\left(q_{(j)}, x_{(j)}\right)$ will be approximately linear related, since $\sigma q_{(j)} + \mu$ is nearly expected sample quantile.

Example 3.9 (Constructing a Q-Q plot) A sample of $n = 10$ observation gives the values in the following table:

Ordered observations $x_{(j)}$	Probability levels $\left(j - \frac{1}{2}\right)/n$	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

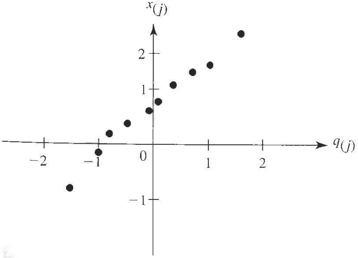

Figure 3.5 A $Q-Q$ plot for the data in Example 3.9.

The steps leading to a $\mathrm{Q}-\mathrm{Q}$ plot are as follows:

1. Order the original observations to get $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ and their corresponding probability values $\left(1-\frac{1}{2}\right)/n, \left(2-\frac{1}{2}\right)/n, \ldots, \left(n-\frac{1}{2}\right)/n$;
2. Calculate the standard quantiles $q_{(1)}, q_{(2)}, \ldots, q_{(n)}$ and
3. Plot the pairs of observations $\left(q_{(1)}, x_{(1)}\right), \left(q_{(2)}, x_{(2)}\right), \ldots, \left(q_{(n)}, x_{(n)}\right)$, and examine the "straightness" of the outcome.

Example 4.10 (A Q-Q plot for radiation data)

The quality -control department of a manufacturer of microwave ovens is required by the federal government to monitor the amount of radiation emitted when the doors of the ovens are closed. Observations of the radiation emitted through closed doors of $n = 42$ randomly selected ovens were made. The data are listed in the following table.

Table 4.1 Radiation Data

Oven no.	Radiation	Oven no.	Radiation	Oven no.	Radiation
1	.15	16	.10	31	.10
2	.09	17	.02	32	.20
3	.18	18	.10	33	.11
4	.10	19	.01	34	.30
5	.05	20	.40	35	.02
6	.12	21	.10	36	.20
7	.08	22	.05	37	.20
8	.05	23	.03	38	.30
9	.08	24	.05	39	.30
10	.10	25	.15	40	.40
11	.07	26	.10	41	.30
12	.02	27	.15	42	.05
13	.01	28	.09		
14	.10	29	.08		
15	.10	30	.18		

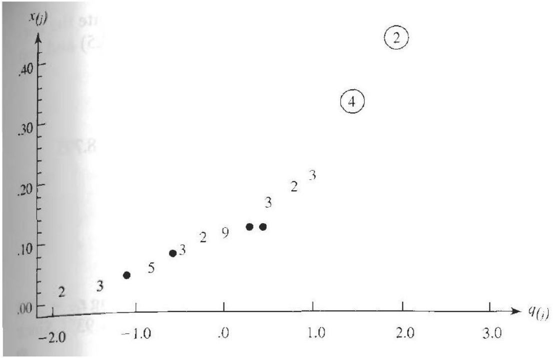

Figure 4.6 A Q–Q plot of the radiation data (door closed) from Example 4.10. (The integers in the plot indicate the number of points occupying the same location.)

The straightness of the $Q - Q$ plot can be measured by calculating the correlation coefficient of the points in the plot. The correlation coefficient for the Q-Q plot is defined by

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

and a powerful test of normality can be based on it. Formally we reject the hypothesis of normality at level of significance α if r_Q fall 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	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
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

Example 3.11 (A correlation coefficient test for normality)

Let us calculate the correlation coefficient r_Q from Q-Q plot of Example 3.9 and test for normality.

Linear combinations of more than one characteristic can be investigated. Many statistician suggest plotting

$$\widehat{\mathbf{e}}_1^{\mathsf{T}}\mathbf{x}_j \quad \text{where } \widehat{\mathbf{S}}\widehat{\mathbf{e}}_1 = \widehat{\lambda}_1\widehat{\mathbf{e}}_1$$

in which $\hat{\lambda}_1$ is the largest eigenvalue of $\mathbf{S}$. Here $\mathbf{x}_j^{\mathsf{T}} = [x_{j1}, x_{j2}, \ldots, x_{jp}]$ is the j th observation on p variables $X_1, X_2, \ldots, X_p$. The linear combination $\hat{\mathbf{e}}_p\mathbf{x}_j$ corresponding to the smallest eigenvalue is also frequently singled out for inspection