Multivariate Gaussian distribution

Woojoo Lee

- A generalization of univariate normal (Gaussian) distribution plays an important role in multivariate data analysis.
- Most of the multivariate techniques introduced in the textbook are based on the "multivariate" normal distribution assumption.
- Although real data are not exactly multivariate normal, it often gives a useful approximation to the true population.
- You should note that the sampling distribution of many multivariate statistics are approximately normal regardless of the population. The central limit theorem comes into play.

Univariate Gaussian density

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) \sim N(\mu, \sigma^2)$$

The importance of the normal distribution rests on its dual role
as both population model for certain natural phenomena and approximate
sampling distribution for many statistics.

Q) What is the meaning of μ and σ ?

See some analogy !

$$(x-\mu)(\sigma^2)^{-1}(x-\mu) \to (\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)$$

$$(2\pi\sigma^2)^{1/2} \to |2\pi\Sigma|^{1/2} = (2\pi)^{p/2}|\Sigma|^{1/2}$$

p-dimensional Gaussian density with mean vector μ and covariance matrix Σ (Assume that Σ is nonsingular):

$$\begin{aligned} f(\mathbf{x}) &= \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp(-(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)/2) \\ &\sim N_p(\mu, \Sigma) \end{aligned}$$

Example – Bivariate Gaussian density

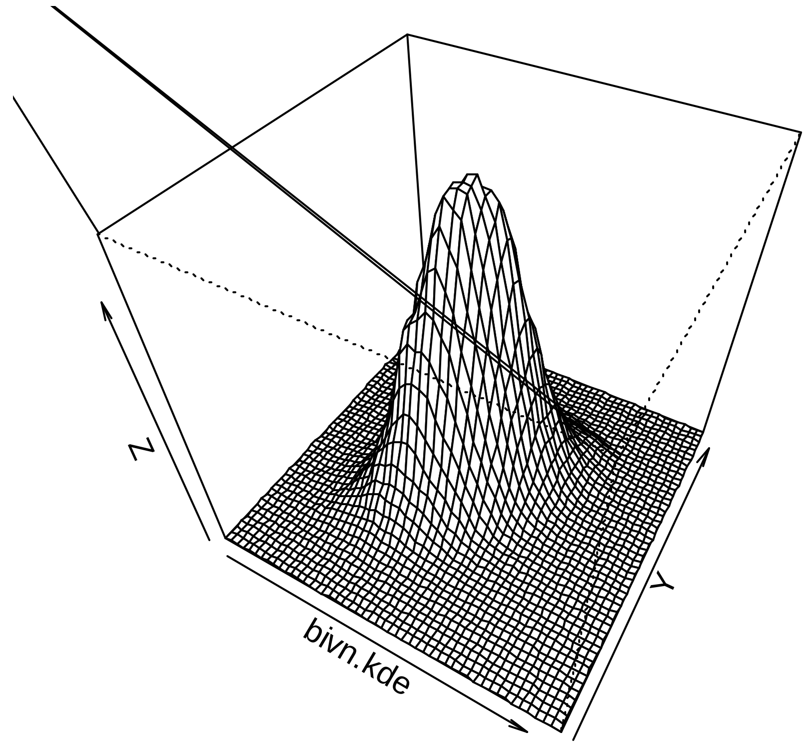
Derive the explicit expression for the bivariate normal density.

Remark) If the correlation is zero, note that

$$f(x_1, x_2) = f(x_1)f(x_2)$$

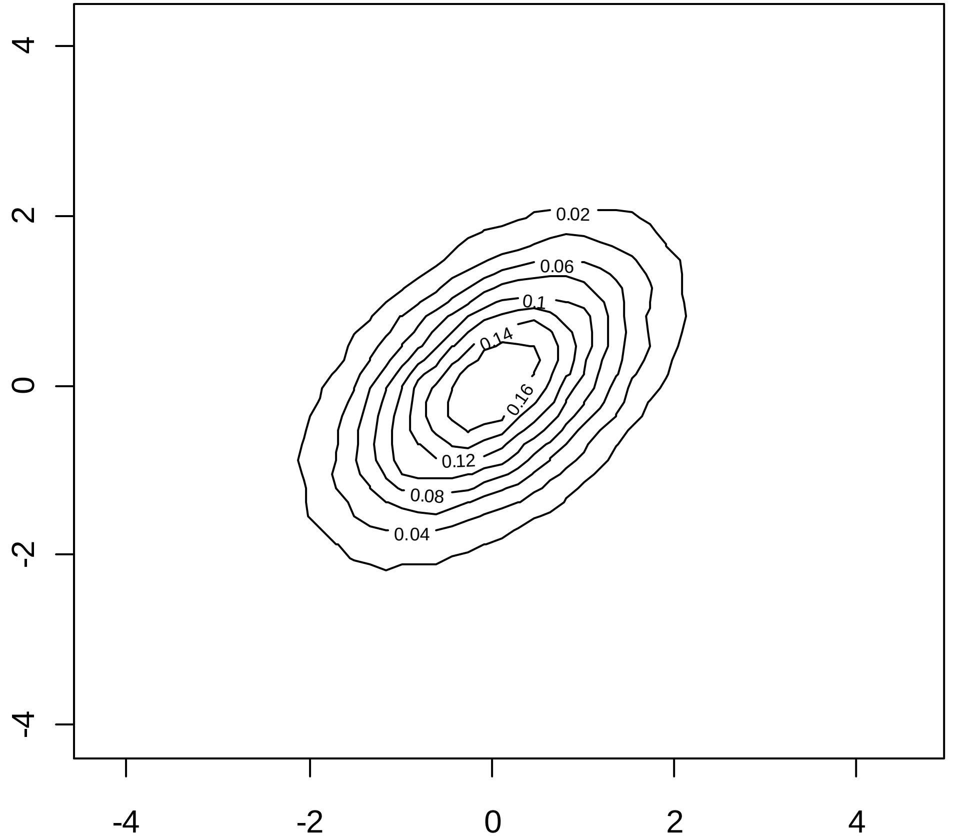
So, zero correlation implies the independence of the two Gaussian random variables.

e.g. $cor(X_1, X_2) = 0.5$



Consider the contour having equal density values :

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



Determine the major axis and the minor axis !

In summary,
contours of constant density for the p-dim'l Gaussian distribution are "ellipsoid" defined by $\mathbf{x}$ such that

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

These ellipsoids are centered at μ and have axes $\pm c\sqrt{\lambda_i}\mathbf{e}_i$, where $\Sigma \mathbf{e}_i = \lambda_i \mathbf{e}_i$ for $i = 1 \cdots, p$.

Example: Consider a bivariate Gaussian distribution with mean 0 and covariance Σ:

$$\Sigma = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

Note that

$$\Sigma = \begin{pmatrix} cos(\pi/4) & -sin(\pi/4) \\ sin(\pi/4) & cos(\pi/4) \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} cos(\pi/4) & -sin(\pi/4) \\ sin(\pi/4) & cos(\pi/4) \end{pmatrix}^T$$

Q) Describe the shape of the corresponding density function.

Basic properties of Multivariate Gaussian distribution

1. Linear combinations of the components of X are normally distributed.
2. All subsets of the components of X have a (multivariate) normal distribution.
3. Zero covariance implies that the corresponding components are independently distributed.
4. The conditional distributions of the components are (multivariate) normal.

$$\begin{aligned} \mathbf{x} \sim N(\mu, \Sigma) \quad &\rightarrow \quad a^T \mathbf{x} \sim N(a^T \mu, a^T \Sigma a) \\ &\rightarrow \quad A\mathbf{x} \sim N(A\mu, A\Sigma A^T) \end{aligned}$$

Remark) $a^T \mathbf{x} \sim N(a^T \mu, a^T \Sigma a)$ for all $a \quad \rightarrow \quad \mathbf{x} \sim N(\mu, \Sigma)$

Remark) $\mathbf{x}_1 \sim N(\mu_1, \Sigma_1)$ and $\mathbf{x}_2 \sim N(\mu_2, \Sigma_2)$ are independent.
Then, $\mathbf{x}_1 + \mathbf{x}_2 \sim N(\mu_1 + \mu_2, \Sigma_1 + \Sigma_2)$

Q) Suppose that

$$\mathbf{x} \sim N(\mu, \Sigma).$$

Then, what is the distribution of $\Sigma^{-1/2}(X - \mu)$?

Show the following statement:

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

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

has probability $1-\alpha$.

Q) When $\mathbf{x} \sim N_3(\mu, \Sigma)$,
find the distribution of $(X_1 - X_2, X_2 - X_3)^T$.

Q) If $\mathbf{x}$ follows $N_5(\mu, \Sigma)$, find the distribution of $(X_2, X_4)^T$.

Q) Let $\mathbf{x}$ be $N_3(\mu, \Sigma)$ with

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

Then, are X_1 and X_2 independent ? How about (X_1, X_2) and X_3 ?

Remark) In general, for Gaussian distribution, $\mathbf{x}_1$ and $\mathbf{x}_2$ are independent if and only if $\Sigma_{12} = 0$.

Conditional distributions of multivariate normal

Suppose that

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \sim N_p(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}) \text{ where } |\Sigma_{22}| > 0.$$

Q) Find the conditional distribution of $\mathbf{x}_1$, given $\mathbf{x}_2$.

Trick:

$$A = \begin{pmatrix} \mathrm{I} & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & \mathrm{I} \end{pmatrix}$$

Sampling distribution of $\bar{\mathbf{x}}$ and S from a multivariate normal population

Before going to the multivariate case, consider the univariate case first. Suppose that

$$X_1, \cdots, X_n$$

constitute a random sample from a univariate normal population with mean μ and variance σ^2.

- What is the sampling distribution of $\bar{X}$?

- Explain why the distribution of $\bar{X}$ cannot be used directly to make inferences about μ.

- What is the sampling distribution of $(n-1)s^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2$?

- Note that the distribution s^2 does not depend on μ.

Properties of $\bar{\mathbf{x}}$

Suppose that

$$\mathbf{x}_1, \cdots, \mathbf{x}_n$$

constitute a random sample from a multivariate normal population with mean μ and covariance Σ.

Q) What is the distribution of $\bar{\mathbf{x}}$?

Properties of $(n-1)S$: Wishart distribution

- This is a generalization of χ^2 distribution.
- This is a probability distributions on random (symm positive definite) matrices.
- This probability density function exists only when $n > p - 1$.
- Two mathematical properties:

  Let $W_p(m, \Sigma)$ be a Wishart distribution with m d.f. and parameter Σ. This is defined as the sum of independent products of multivariate normal random vectors: $\sum_{j=1}^{m} \mathbf{z}_j \mathbf{z}_j^T$ where $\mathbf{z}_j \sim N_p(0, \Sigma)$.

$$
\begin{aligned}
1) A_1 \sim W_p(m_1, \Sigma) \quad & \perp \quad A_2 \sim W_p(m_2, \Sigma) \\
& \rightarrow \quad A_1 + A_2 \sim W_p(m_1 + m_2, \Sigma) \\
2) A \sim W_p(m, \Sigma) \quad & \rightarrow \quad CAC^T \sim W_p(m, C\Sigma C^T)
\end{aligned}
$$

Q) Derive the previous two properties of Wishart distribution.

Large sample behavior of $\bar{\mathbf{x}}$ and S

Can we derive some meaningful asymptotic conclusions without the MVN assumption ? Assume that $\mathbf{x}_i$ $(i = 1, \cdots, n)$ are i.i.d. p-dimensional random vectors with mean μ and covariance Σ.

1) Law of large numbers

$$\bar{\mathbf{x}} \rightarrow_p \mu$$
$$S \rightarrow_p \Sigma$$

2) Central limit theorem

$$\sqrt{n}(\bar{\mathbf{x}} - \mu) \sim N_p(0, \Sigma)$$

Q) Explain “convergence in probability”.
Remark) Vector convergence is just the result from elementwise convergence.
Remark) Matrix convergence is also the result from elementwise convergence.

When n is large and $n >> p$,
Step 1)

$$n(\bar{\mathbf{x}} - \mu)^T \Sigma^{-1} (\bar{\mathbf{x}} - \mu) \rightarrow \chi_p^2$$

Step 2)

$$n(\bar{\mathbf{x}} - \mu)^T S^{-1} (\bar{\mathbf{x}} - \mu) \rightarrow \chi_p^2$$

Assessing the assumption of normality

To some degree, the quality of inferences made by asymptotic methods depends on how closely the true parent population resembles the MVN form.

→It is imperative that procedures exist for detecting extreme departure cases from MVN.

→We will concentrate on one or two dimensional examinations of normality, but we have to pay a price.

The following questions are addressed:

- Do the marginal distribution of the elements of $\mathbf{x}$ appear to be normal ?
- Do the scatter plots of pairs of observations give the elliptical appearance ?

Evaluating the normality of the univariate marginal distribution

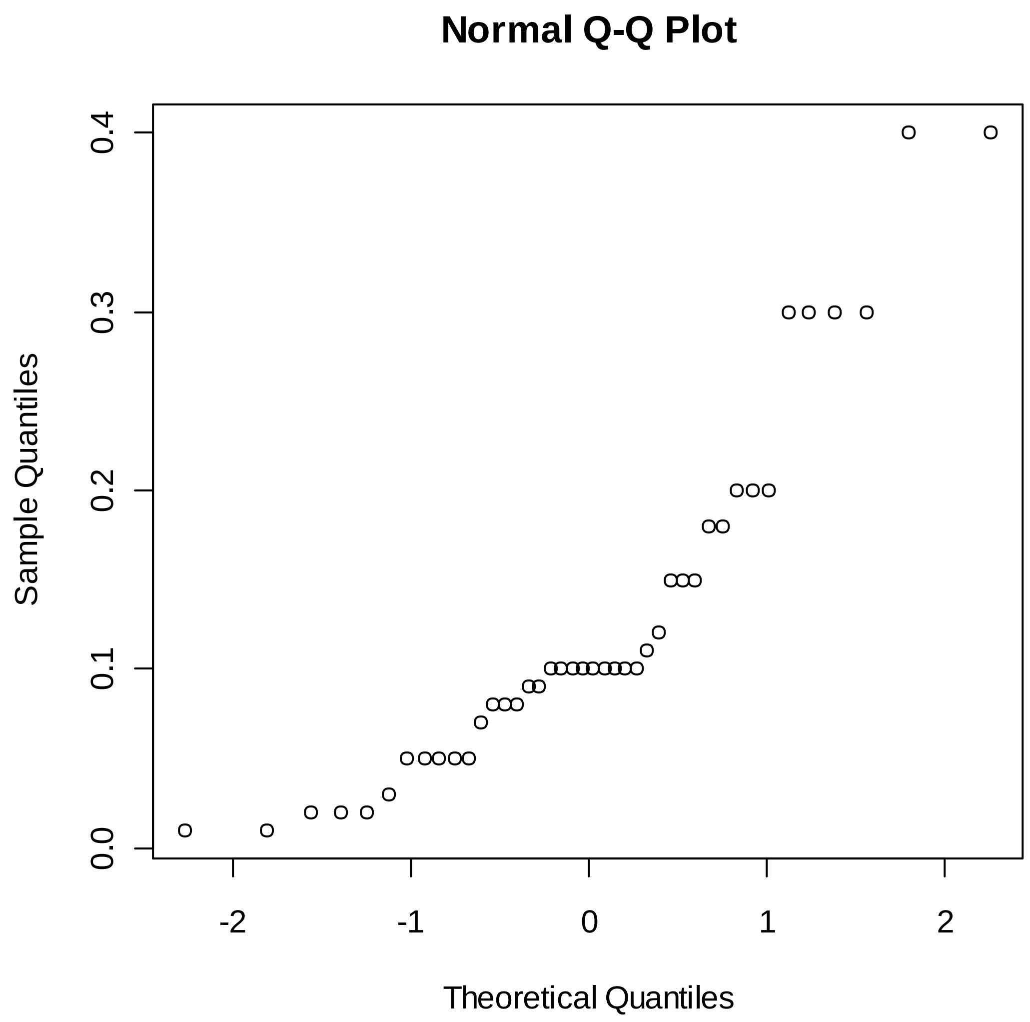
The basic idea is that
"the marginal distribution of the elements of $\mathbf{x}$ appear to be normal".

Normal Q-Q plot : the plot of the sample quantile versus the quantile one would expect to observe if the observations were normally distributed (useful when $n > 20$)

- Order the observations (X_i) and find their corresponding probability values.
- Compute the standard normal quantile (q)
- plot the pairs of (q, x) and examine the straightness

Q) If the data arise from a normal distribution, what do you expect for the shape of (q, x) ?

Example of QQ plot (real data)



Checking multivariate normality

This can be done by using

$$d_j^2 = (\mathbf{x}_j - \bar{\mathbf{x}})^T S^{-1} (\mathbf{x}_j - \bar{\mathbf{x}}).$$

We expect that if the population is multivariate normal and n and $n-p$ are greater than 30, $d_1^2, \cdots, d_n^2$ will behave like a χ_p^2.

To construct the chi-square plot :

- Order the d_j^2 and find their corresponding probability values.
- Compute the χ_p^2 quantile (q)
- plot the pairs of (q, x) and examine the straightness

Detecting outliers

- Draw scatter plots
- Examine standardized values
- Examine $(\mathbf{x}_i - \bar{\mathbf{x}})^T S^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})$

But, you should be aware that these methods are not perfect !

Transformations to near normality

If normality is not a viable assumption, what is the next step ?

Some helpful transformations

- count$(y) \to \sqrt{y}$
- proprotions $(p) \to logit(p) = \frac{1}{2}\log(\frac{p}{1-p})$
- correlation $(r) \to Fisher'z(r) = \frac{1}{2}\log(\frac{1+r}{1-r})$

In addition, the reciprocal and logrithm transformation are also often used.

For multivariate observations, just make each marginal distribution approximately normal. Although this is not sufficient to ensure that the joint distribution is normal, this is often good enough in practice.

See that this transformation is beneficial with computer experiment !.