The Multivariate Normal Distribution

Prof. M. Salicrú,
Prof. S. Civit

Multivariate Analysis
MSc. in Statistics & Operational Research
UB & UPC

Multivariate Normal Distribution

The univariate normal distribution has a generalized form in p dimensions. The p-dimensional normal density function is $N(\mu, \Sigma)$.

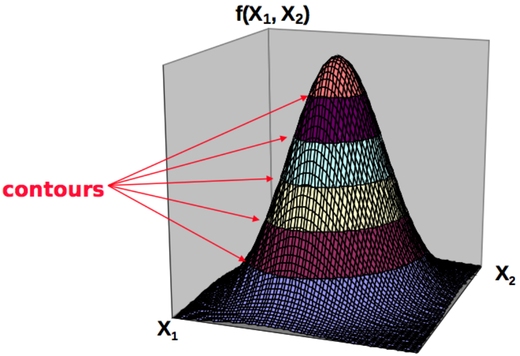
$$f(x) = \frac{1}{|\Sigma|^{\frac{1}{2}}(2\pi)^{\frac{p}{2}}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}$$

where $-\infty \leq x_i \leq \infty \;\; i = 1, \ldots, p$

$$\mu = (\mu_1, \mu_2, \ldots, \mu_p)'$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix}$$

Graphically, the bivariate normal probability density function,



Properties of the Multivariate Normal Distribution

For any multivariate normal random vector,

1. The density

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

   has a maximum value at

$$\mu = (\mu_1, \mu_2, \ldots, \mu_p)$$

2. Linear combinations of the components of X are normally distributed.
3. All subsets of the components of X have a (multivariate) normal distribution.
4. Cov$= 0$ implies that the correponding components of X are independently distributed.
5. Conditional distributions of the components of X are (multivariate) normal.

1. If $X \sim N_p(\mu, \Sigma)$, then any linear combination

$$a'X = \sum_{i=1}^{p} a_i X_i \sim N(a'\mu, a'\Sigma a)$$

Furthermore, if $a'X \sim N(a'\mu, a'\Sigma a) \quad \forall a$, then

$$X \sim N_p(\mu, \Sigma)$$

2. If $X \sim N_p(\mu, \Sigma)$, then any set of q linear combinations

$$A \cdot X = \begin{bmatrix} \Sigma a_{1i} X_i \\ \Sigma a_{2i} X_i \\ \vdots \\ \Sigma a_{qi} X_i \end{bmatrix} \sim N_q(A\mu, A\Sigma A')$$

Furthermore, if d is a vector of constants, then
$X + d \sim N_p(\mu + d, \Sigma)$

3. If $X \sim N_p(\mu, \Sigma)$, the all subsets of X are (multivariate) normally distributed,

$$X = \begin{bmatrix} X_1 \\ (q \times 1) \\ \\ X_2 \\ ((p-q) \times 1) \end{bmatrix} \quad \mu = \begin{bmatrix} \mu_1 \\ (q \times 1) \\ \\ \mu_2 \\ ((p-q) \times 1) \end{bmatrix} \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ (q \times q) & (q \times (p-q)) \\ \\ \Sigma_{21} & \Sigma_{22} \\ ((p-q) \times q) & (p-q) \times (p-q) \end{bmatrix}$$

then $X_1 \sim N_q(\mu_1, \Sigma_{11})$ and $X_2 \sim N_{p-q}(\mu_2, \Sigma_{22})$

4. a) If X_1 and X_2 are independent, then Cov$(X_1, X_2) = 0$

   b) If

$$X = \begin{bmatrix} X_1 \\ (q_1) \\ \\ X_2 \\ (q_2) \end{bmatrix} \sim N_{q_1+q_2}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right)$$

   then X_1 and X_2 are independent if $\Sigma_{12} = 0$.

   c) If $X_1 \sim N_{q_1}(\mu_1, \Sigma_{11})$ and $X_2 \sim N_{q_2}(\mu_2, \Sigma_{22})$ are independent, then,

$$X = \begin{bmatrix} X_1 \\ (q_1) \\ \\ X_2 \\ (q_2) \end{bmatrix} \sim N_{q_1+q_2}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix} \right)$$

5. If $X \sim N_p(\mu, \Sigma)$

$$X = \begin{bmatrix} X_1 \\ (q_1) \\ \\ X_2 \\ (q_2) \end{bmatrix} \sim N_p \left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right)$$

with $|\Sigma_{22}| > 0$, then

$$X_1 / X_2 = x_2 \sim N \left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \right)$$

6. If $X \sim N_p(\mu, \Sigma)$ and $|\Sigma| > 0$ then,

$$(X-\mu)'\Sigma^{-1}(X-\mu) \sim \chi^2_p$$

with p degrees of freedom

$$P(x : (x-\mu)'\Sigma^{-1}(x-\mu) \leq c) = P(\chi^2_p \leq c)$$

for $p = 2$,

$(x-\mu)'\Sigma^{-1}(x-\mu) = c$ define an ellipse

7. Let $X_i \sim N_p(\mu_i, \Sigma_i) \quad i = 1 \ldots, n$ be mutually independent, then,

a)

$$V = \sum_{i=1}^{n} c_i X_i \sim N_p\left(\sum_{i=1}^{n} c_i \mu_i, \left(\sum_{i=1}^{n} c_i^2 \Sigma_i\right)\right)$$

being c_i real values

b)

$$S = \sum_{i=1}^{n} A_i X_i \sim N_s\left(\sum_{i=1}^{n} A_i \mu_i, \sum_{i=1}^{n} A_i \Sigma_i A_i'\right)$$

being A_i s x p matrix

c)

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} \sim N_p(\mu, \Sigma/n)$$

where

$$c_1 = c_2 =, \ldots = c_n = \frac{1}{n}$$

or

$$A_1 = A_2 = \ldots = A_n = diag\left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$$

Linear combination: Example (1/2)

Suppose $X \sim N_3(\mu, \Sigma)$, being,

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

Based on the *following linear combination*,

$$Y_1 = 2X_1 + 0X_2 + 1X_3$$
$$Y_2 = 0X_1 + 0X_2 + 2X_3$$
$$Y_3 = 1X_1 - 1X_2 + 0X_3$$

Find the probability density function of (Y_1, Y_2, Y_3)

Linear combination: Example (2/2)

According to *property 2*

$$Y = A \cdot X$$

$$\mu_Y = A \cdot \mu_X,$$

$$\Sigma_Y = A \cdot \Sigma \cdot A'$$

$$\mu_Y = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 0 & 2 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ -3 \end{bmatrix}$$

$$\Sigma_Y = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 0 & 2 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 2 & -1 & 1 \\ -1 & 3 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 1 \\ 0 & 0 & -1 \\ 1 & 2 & 0 \end{bmatrix} = \begin{bmatrix} 5 & -2 & 3 \\ 2 & 0 & 2 \\ 3 & -4 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 1 \\ 0 & 0 & -1 \\ 1 & 2 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 13 & 6 & 7 \\ 6 & 4 & 2 \\ 7 & 2 & 7 \end{bmatrix}$$

Conditional distribution: Example (1/2)

Suppose $X \sim N_3(\mu, \Sigma)$, being,

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

Find the $(X_1, X_2)|X_3 = 1$ probability density function

According to *property 5*

$$(X_1, X_2|X_3 = 1) = N\left(\mu_{(X_1,X_2)} + \Sigma_{12}\Sigma_{22}^{-1}(x_3 - \mu_3), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

Conditional distribution: Example (2/2)

$$\mu = \begin{bmatrix} -1 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \cdot 1 \cdot (1-1) = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix} - \begin{bmatrix} -1 \\ 0 \end{bmatrix} \cdot 1 \cdot (1,0) = \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 3 \end{bmatrix}$$

Marginal distribution: Example (1/2)

Suppose $X \sim N_3(\mu, \Sigma)$, being,

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

Find the marginal probability density function of (X_1, X_3) and X_2
Are (X_1, X_3) and X_2 independent?

Conditional distribution: Example (2/2)

According to *property 3*

$$(X_1, X_3) \sim N\left(\begin{bmatrix}-1, & 1\end{bmatrix}, \begin{bmatrix}2 & 1\\ 1 & 1\end{bmatrix}\right)$$

$$X_2 \sim N(2, 3)$$

According to *property 4a, 4c*

$$\Sigma_{(X_1,X_3),X_2} = (-1, 0) \neq (0, 0)$$

then X_1, X_3, X_2 are dependent

$$(X_1, X_3, X_2) \sim N\left(\begin{bmatrix}-1, & 1, & 2\end{bmatrix}, \begin{bmatrix}2 & 1 & -1\\ 1 & 1 & 0\\ -1 & 0 & 3\end{bmatrix}\right)$$

Sampling from a Multivariate Normal Distribution

Let $X_i \sim N_p(\mu, \Sigma) \quad i = 1, \ldots, n$ represent a random sample.

Since the X_j's are mutually independent and each have $N_p(\mu, \Sigma)$ distribution, their joint density is the product of their marginal densities.

Joint distribution

$$f(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} \left\{ \frac{1}{|\Sigma|^{1/2}(2\pi)^{p/2}} e^{-\frac{1}{2}(x_i-\mu)'\Sigma^{-1}(x_i-\mu)} \right\}$$

Maximum Likelihood Estimation

For a random sample $X_i \sim N_p(\mu, \Sigma) \quad i = 1, \ldots, n$ from a normal population, the MLE for μ and Σ are,

$\hat{\mu}_{\text{ML}}$ & $\hat{\Sigma}_{\text{ML}}$

$$\hat{\mu} = \bar{X} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_p \end{bmatrix}$$

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})' = \frac{(n-1)}{n} S$$

$$S_{ij} = \frac{1}{n-1} \sum_{k=1}^{n} (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)$$

Large sample behavior of $\bar{X}$ and S

Let $X_i \quad i = 1, \ldots, n$ be iid observations with mean μ and covariance Σ, then,

$$\bar{X} \sim N_p(\mu, (1/n)\Sigma)$$

for n large relative to p. This can be restated as

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

For n large, Σ can be substituted by S so,

$$n(\bar{X} - \mu)' S^{-1} (\bar{X} - \mu) \simeq \chi^2_p$$

Based on the properties of the Multivariate Normal Distribution, we know

- All linear combinations of the individual normal are normal.
- The contours of the multivariate normal density are concentric ellipsoids

These facts suggest investigation of the following questions (in one or two dimensions):

- Do the marginal distributions of the elements of X appear normal? What about a few linear combinations?
- Do the bivariate scatterplots appear ellipsoidal?
- Are there any unusual looking observations (outliers)?

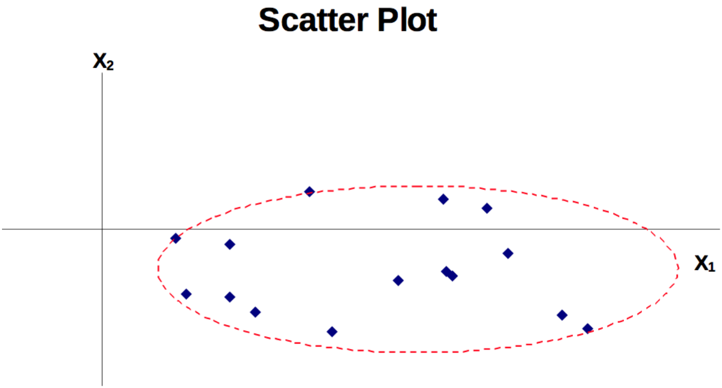
Example

Suppose we had the following fifteen (ordered) sample observations on some random variables X_1 and X_2:

x_{j1}	x_{j2}
1.43	-0.69
1.62	-5.00
2.46	-1.13
2.48	-5.20
2.97	-6.39
4.03	2.87
4.47	-7.88
5.76	-3.97
6.61	2.32
6.68	-3.24
6.79	-3.56
7.46	1.61
7.88	-1.87
8.92	-6.60
9.42	-7.64

- Do these data support the assertion that they were drawn from a bivariate normal parent population?

Example



Example

To create a Chi-Square plot, we will need to calculate the squared generalized distance from the centroid for each observation.

$$d_i^2 = (x_i - \bar{X})' S^{-1} (x_i - \bar{X})$$

x_{j1}	x_{j2}	d_j^2
1.43	-0.69	2.400
1.62	-5.00	2.279
2.46	-1.13	1.336
2.48	-5.20	1.548
2.97	-6.39	1.739
4.03	2.87	2.976
4.47	-7.88	2.005
5.76	-3.97	0.090
6.61	2.32	2.737
6.68	-3.24	0.281
6.79	-3.56	0.333
7.46	1.61	2.622
7.88	-1.87	0.138
8.92	-6.60	2.686
9.42	-7.64	3.819

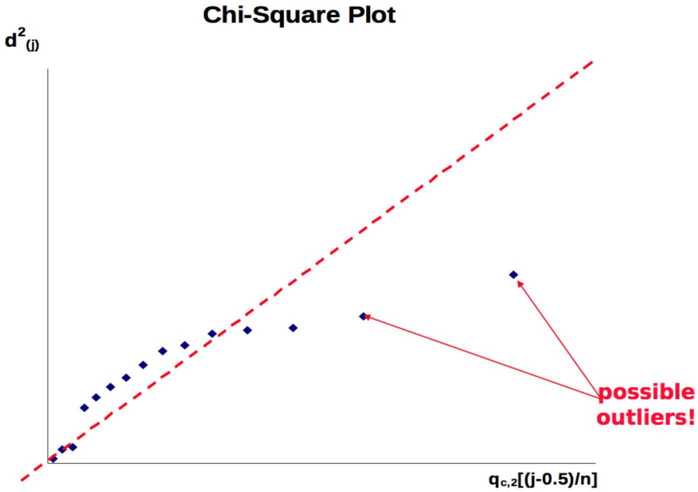
x_{j1}	x_{j2}	d_j^2
5.76	-3.97	0.090
6.68	-3.24	0.281
6.79	-3.56	0.333
7.88	-1.87	0.138
2.46	-1.13	1.336
2.48	-5.20	1.548
2.97	-6.39	1.739
4.47	-7.88	2.005
1.62	-5.00	2.279
1.43	-0.69	2.400
7.46	1.61	2.622
8.92	-6.60	2.686
6.61	2.32	2.737
4.03	2.87	2.976
9.42	-7.64	3.819

Example

The corresponding $(j-1/2)/n$ percentile of the Chi-Square distribution with p degrees of freedom,

x_{j1}	x_{j2}	d_j^2	$(j-0.5)/n$	$q_{c,2}[(j-0.5)/n]$
5.76	-3.97	0.090	0.033	0.068
6.68	-3.24	0.281	0.100	0.211
6.79	-3.56	0.333	0.167	0.365
7.88	-1.87	0.138	0.233	0.531
2.46	-1.13	1.336	0.300	0.713
2.48	-5.20	1.548	0.367	0.914
2.97	-6.39	1.739	0.433	1.136
4.47	-7.88	2.005	0.500	1.386
1.62	-5.00	2.279	0.567	1.672
1.43	-0.69	2.400	0.633	2.007
7.46	1.61	2.622	0.700	2.408
8.92	-6.60	2.686	0.767	2.911
6.61	2.32	2.737	0.833	3.584
4.03	2.87	2.976	0.900	4.605
9.42	-7.64	3.819	0.967	6.802

Example



Outlier detection

Detecting outliers (extreme or unusual observations) in $p > 2$ dimensions is very tricky. Consider the following situation: