Multivariate normal distribution in a nutshell

Definition: The most covenient definition for computation purposes and for purposes of proving facts about this distribution is the following:

Take independent standard normal random variables $Z_1, Z_2, \ldots, Z_n$ and assemble them into a vector $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)^T$. The distribution of any vector $\mathbf{X}$ defined as

$$\mathbf{X} = \mathbf{A}\mathbf{Z} + \mathbf{b} \tag{$*$}$$

where $\mathbf{A}(n \times m)$ is an arbitrary matrix and $\mathbf{b}(m \times 1)$ and arbitrary vector will be called multivariate normal.

Elementary properties: The properties of multivariate normal distribution that follow easily form the definiton are the following:

(i) The marginal distributions, i.e. the distributions of the components $X_1, X_2, \ldots, X_m$ are normal random variables (in general not independent). Also, the distribution of any subcollection of components is again a multivariate normal by the definition given above.

(ii) Any linear combination $\sum_{i=1}^n c_i X_i$ of the components of a multivariate normal is a normal random variable.

(iii) Any affine transformation $\mathbf{B}\mathbf{X} + \mathbf{c}$ of a multivariate normal vector $\mathbf{X}$ is again a multivariate normal vector (of possibly different dimension).

Characteristic functions: It has been shown in class that the characteristic function of the vector $\mathbf{X}$ defined by (*) equals

$$\phi_{\mathbf{x}}(\mathbf{t}) = e^{i\mathbf{t}^T\mathbf{b}} e^{-1/2\mathbf{t}^T\mathbf{A}\mathbf{A}^T\mathbf{t}}.$$

Since $E(\mathbf{X}) = \mathbf{b}$ and $\mathrm{var}(\mathbf{X}) = \mathbf{A}\mathbf{A}^T$ it is natural to use the notation $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for the multivariate normal in analogy with the univariate normal. Note that the expectation and the variance matrix completely determine the characteristic function of the multivariate normal distribution and hence of the distribution itself. A mathematically more elegant and compact definition of the multivariate normal $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for arbitrary $\boldsymbol{\mu}$ and non–negative definite $\boldsymbol{\Sigma}$ could be that its characteristic function is of the form

$$\phi_{\mathbf{x}}(\mathbf{t}) = e^{i\mathbf{t}^T\boldsymbol{\mu}} e^{-1/2\mathbf{t}^T\boldsymbol{\Sigma}^T\mathbf{t}}$$

but this gives a less intuitive picture.

In the above notation let $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and let $A(q \times p)$ be an arbitrary amtrix and $\boldsymbol{\nu}$ an arbitrary q–vector. Then $\mathbf{Y} = \mathbf{A}\mathbf{X} + \boldsymbol{\nu} \sim N_q(\mathbf{A}\boldsymbol{\mu} + \boldsymbol{\nu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T)$ which is a more precise statement of the fact that under affine transformations multivariate normal distribution transforms “nicely”.

Uncorrelated components are independent: In general it is not true that uncorrelated random variables are independent. If they are jointly multivariate normal, however, the assertion is true and follows easily from the characteristic functions above. More precisely, if $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where the variance matrix is of the form

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

where $\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{22}$ are $r \times r$ and $s \times s$ matrices ($r + s = p$) then it is easy to see that

$$\phi_{\mathbf{X}}(\mathbf{t}) = \phi_{\mathbf{X}_1}(\mathbf{t}_1)\phi_{\mathbf{X}_2}(\mathbf{t}_2)$$

with $\mathbf{X}_1$ denoting the vector of the first r components of $\mathbf{X}$ and similar notation is used for other vectors. This implies independence.

Densities: There was no mention of densities so far in the definition of the multivariate normal distribution. This is because the family of multivariate normal distributions contains also distributions that are concentrated on affine subspaces of the Euclidian space and therefore do not have a density in the usual sense. If $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and the variance matrix is invertible then $\mathbf{X}$ also has a density of the form

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-p/2}(det(\boldsymbol{\Sigma}))^{-1/2} \exp\{-1/2\,\mathbf{x}^T\boldsymbol{\Sigma}^{-1}\mathbf{x}\}.$$

This density can be obtained by writing $\mathbf{X} = \boldsymbol{\Sigma}^{1/2}\mathbf{Z} + \boldsymbol{\mu}$ with $\mathbf{Z}$ having independent standard normal components. The next step is then to use the standard tool of change of variables in more dimensions to show that the densities of $\mathbf{Z}$ and $\mathbf{X}$ are related as

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{Z}}(\boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu}))det(\boldsymbol{\Sigma})^{-1/2}.$$

Conditional distributions: Let $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Partition $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)^T$ and similarly $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)^T$ and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

The conditional distribution of $\mathbf{X}_1$ given $\mathbf{X}_2 = \mathbf{x}_2$ is multivariate normal, i.e.

$$\mathbf{X_1}|\mathbf{X}_2 = \mathbf{x}_2 \sim N_r(\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}).$$

The assertion can be derived easily from the fact that the two vectors

$$\mathbf{Y} = \mathbf{X}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{X}_2 \quad \text{and} \quad \mathbf{X}_2$$

are independent. This in fact can be shown quite easily by computing the correlation.

Quadratic forms: Often it is necessary to consider random variables of the form

$$\mathbf{X}^T\mathbf{C}_i\mathbf{X}$$

for given matrices $\mathbf{C}_i, i = 1, 2, \ldots, k$. The motivation comes mainly from testing problems. The two questions that will be asked are (i) what are the distributions of individual terms? (b) when are such random variables independent?

Theorem: (Cochran 1934) If $\mathbf{X} \sim N_p(\mathbf{0}, \mathbf{I})$ and $\mathbf{C}$ is an idempotent matrix then $\mathbf{X}^T\mathbf{C}\mathbf{X}$ has the $\chi^2(r)$ distribution where $r = rank(\mathbf{C})$.

Proof: See lecture notes or Mardia, Kent, Bibby, Multivariate analysis, Academic Press 1979.

The question about independence is answered by

Theorem: (Craig 1943) Suppose $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \mathbf{I})$ and $C_i, i = 1, 2, \ldots, k$ are symmetric matrices. The random variables $\mathbf{X}^T\mathbf{C}\mathbf{X}$ are independent if and only if $\mathbf{C}_i\mathbf{C}_j = \mathbf{0}$ for all $i \neq j$.

Proof: See lecture notes or Mardia, Kent, Bibby, Multivariate analysis, Academic Press 1979.

A few sample problems

1. If $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}^T (q \times q)$ is non–singular, then, given that $\mathbf{Q}\mathbf{x} = \mathbf{q}$, show that the conditional distribution of $\mathbf{x}$ is normal with mean $\boldsymbol{\mu} + \boldsymbol{\Sigma}\mathbf{Q}^T(\mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}^T)^{-1}(\mathbf{q} - \mathbf{Q}\boldsymbol{\mu}$ and (singular) covariance matrix $\boldsymbol{\Sigma} - \boldsymbol{\Sigma}\mathbf{Q}^T(\mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}^T)^{-1}\mathbf{Q}\boldsymbol{\Sigma}$.

2. Suppose $\mathbf{X}$ and $\mathbf{Y}$ are p–dimensional random vectors such that

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim N_{2p}(\mathbf{0}, \boldsymbol{\Sigma})$$

where the covariance matrix is of the form

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathbf{I} & \rho\mathbf{1}\mathbf{1}^T \\ \rho\mathbf{1}\mathbf{1}^T & \mathbf{I} \end{pmatrix}.$$

The matrix $\mathbf{I}$ represents the $p \times p$ identity matrix, $\mathbf{1} = (1, 1, \ldots, 1)^T$ and ρ is a scalar constant such that $|\rho| \leq 1/\sqrt{p(p-1)}$.

a. Compute $E(\mathbf{Y}\mathbf{Y}^T|\mathbf{X})$.

Hint: For a random vector $\mathbf{W}$ *the covariance matrix equals* $E(\mathbf{W}\mathbf{W}^T) - E(\mathbf{W})E(\mathbf{W})^T$.

b. Assume $p \geq 4$. Compute

$$E\left(\frac{\sum_{i=1}^p X_i Y_i}{\sum_{i=1}^p (X_i - \bar{X})^2}\right).$$

Hint: $E(U) = E\left[E(U|\mathbf{X})\right]$ *for any* U. *Then use the fact that if* $W \sim F_{m,n}$ *then* $E(W) = n/(n-2)$ *if* $n > 2$.

3. Suppose $\mathbf{X} = (X_1, X_2, \ldots, X_n) \sim N(\boldsymbol{\mu}, \sigma^2\boldsymbol{\Sigma})$ where σ^2 is an unknown parameter and $\boldsymbol{\Sigma}$ is a known invertible matrix.

a. Suppose the expectation $\boldsymbol{\mu}$ is known and you have one observation $\mathbf{X}_1$. How would you estimate σ^2? Is your estimate unbiased? Is it the UMVU estimate of σ^2? What is the variance of the estimate you found?

Hint: What is the distribution of $\boldsymbol{\Sigma}^{-1/2}\mathbf{X}$*?*

b. How would you go about the questions in **a.** if $\boldsymbol{\mu}$ was not known but you knew that all components of $\boldsymbol{\mu}$ were the same?