The Normal Density

* The normal distribution plays a very important role in univariate statistics.

* The density function is of the form

$$f(x;\mu,\sigma^2) \;=\; \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\,\frac{(x-\mu)^2}{\sigma^2}\right\} \qquad \text{for any } x \in \mathbb{R}$$

* The parameters are the mean μ and variance σ^2.

* One reason for the importance of the normal is that many real-life quantities approximately follow such a distribution.

* Also, from the Central Limit Theorem, averages of random variables have a distribution that can be approximated by the normal for large n.

The Multivariate Normal Density

* The normal can be extended to a multivariate normal in a very natural way replacing x and μ with vectors and σ^2 with a matrix.

* The density of this distribution for any $\boldsymbol{x} \in \mathbb{R}^p$ is

$$f(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{- \frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right\}$$

* Note that when $p = 1$ this simplifies into the usual univariate normal density.

* The parameters are the $p \times 1$ mean vector $\boldsymbol{\mu}$ and the $p \times p$ symmetric positive-definite covariance matrix. $\boldsymbol{\Sigma}$.

The Standard Multivariate Normal

Definition 2

The random vector $\boldsymbol{Z}$ is said to have a standard normal distribution if the component random variables $Z_1, \ldots, Z_p$ are mutually independent and

$$Z_j \sim \textit{Normal}(0,1) \qquad j = 1, \ldots, p$$

* The mean vector of $\boldsymbol{Z}$ is $\mathbf{0}$ and the covariance matrix is $\boldsymbol{I}_p$.

* The density of this distribution for any $\boldsymbol{z} \in \mathbb{R}^p$ is

$$\begin{aligned} f(\boldsymbol{z}) &= \frac{1}{(2\pi)^{p/2}} \exp\left\{-\frac{\boldsymbol{z}^t \boldsymbol{z}}{2}\right\} \\ &= \frac{1}{(2\pi)^{p/2}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{p} z_j^2\right\} \end{aligned}$$

The General Multivariate Normal

Theorem 3

Suppose that $\boldsymbol{Z} \sim \mathsf{N}_p(\mathbf{0}, \boldsymbol{I}_p)$*. Now let* $\boldsymbol{\mu}$ *be any vector in* $\mathbb{R}^p$ *and let* $\boldsymbol{\Sigma}$ *be a positive definite matrix with matrix square root given by* $\boldsymbol{\Sigma}^{1/2}$*. Then the random vector*

$$\boldsymbol{X} \;=\; \boldsymbol{\Sigma}^{1/2}\boldsymbol{Z} + \boldsymbol{\mu} \;\sim\; \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Corollary 3.1

Suppose that the random vector $\boldsymbol{X} \;\sim\; \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *then*

$$\boldsymbol{Z} \;=\; \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu}) \;\sim\; \mathsf{N}_p(\mathbf{0}, \boldsymbol{I}_p)$$

The Bivariate Normal

* For simplicity we shall first look at the bivariate normal density ($p = 2$).

* The covariance matrix is

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$$

* Hence we have that

$$\begin{aligned} |\boldsymbol{\Sigma}| &= \sigma_{11}\sigma_{22} - \sigma_{12}^2 \\ &= \sigma_{11}\sigma_{22}(1 - \rho_{12}^2) \end{aligned}$$

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_{11}\sigma_{22}(1 - \rho_{12}^2)} \begin{pmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{pmatrix}$$

The Bivariate Normal

* Using these results and setting $\boldsymbol{\mu}^t = (\mu_1, \mu_2)$ we get that the bivariate normal density can be written as

$$
\begin{aligned}
f(x_1, x_2; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \;=\; & \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}} \\
& \times \;\exp\left\{-\,\frac{1}{2(1-\rho_{12}^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_{11}} + \frac{(x_2-\mu_2)^2}{\sigma_{22}}\right.\right. \\
& \left.\left. -2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right]\right\}
\end{aligned}
$$

for any point $(x_1, x_2)^t \in \mathbb{R}^2$.

Properties of the Bivariate Normal

Theorem 4

Suppose $(X_1, X_2)^t$ *is a bivariate normal random vector then* X_1 *and* X_2 *are independent if, and only if, their correlation* $\rho_{12} = 0$.

Theorem 5

Suppose that $(X_1, X_2)^t$ *is a bivariate normal random vector with mean vector and covariance matrix*

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$$

Then the marginal distributions are such that

$$X_j \sim \textit{Normal}\big(\mu_j, \sigma_{jj}\big) \qquad j = 1, 2.$$

Conditional Distribution

Theorem 6

Suppose that $(X_1, X_2)^t$ *is a bivariate normal random vector with mean vector and covariance matrix*

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$$

Then the conditional distribution of X_1 *given* $X_2 = x_2$ *is such that*

$$X_1 \mid X_2 = x_2 \sim \textit{Normal}\left(\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(x_2 - \mu_2),\ \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}\right).$$

Linear Combinations

Theorem 7

Suppose that $\boldsymbol{X} = (X_1, X_2)^t$ is a bivariate normal random vector with mean vector and covariance matrix

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$$

Let $\boldsymbol{a} = (a_1, a_2)^t$ be a vector of constants then

$$\boldsymbol{a}^t\boldsymbol{X} = a_1X_1 + a_2X_2 \sim \textit{Normal}\left(\boldsymbol{a}^t\boldsymbol{\mu},\ \boldsymbol{a}^t\boldsymbol{\Sigma}\boldsymbol{a}\right).$$

The converse of this is that $\boldsymbol{X}$ is bivariate normal if, and only if,

$$\boldsymbol{a}^t\boldsymbol{X} \sim \textit{Normal}\left(\boldsymbol{a}^t\boldsymbol{\mu},\ \boldsymbol{a}^t\boldsymbol{\Sigma}\boldsymbol{a}\right).$$

for every $\boldsymbol{a} \in \mathbb{R}^2$.

General Results for the Multivariate Normal

* The results from the previous slides can be extended to the multivariate case.

* The four main results are
  1. Linear combinations of normal random vectors are (multivariate) normal.
  2. If two components of a normal random vector have correlation equal to 0 then those components are independent.
  3. All subsets of a normal random vector are (multivariate) normal.
  4. The conditional distribution of a subset of $\boldsymbol{X}$ conditional on the rest of $\boldsymbol{X}$ is normal.

* In the next slides we will give more formal statements of these results.

Linear Combinations of Normals

Lemma 1

If $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{C}$ is a non-singular $p \times p$ matrix of constants. Let $\boldsymbol{Y}$ be the p-dimensional random vector $\boldsymbol{Y} = \boldsymbol{C}\boldsymbol{X}$. Then

$$\boldsymbol{Y} \sim N_p\left(\boldsymbol{C}\boldsymbol{\mu}, \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}^t\right).$$

Theorem 8

If $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{A}$ is a $q \times p$ matrix of constants then the q linear combinations

$$\boldsymbol{A}\boldsymbol{X} = \begin{pmatrix} a_{11}X_1 + \cdots + a_{1p}X_p \\ a_{21}X_1 + \cdots + a_{2p}X_p \\ \vdots \\ a_{q1}X_1 + \cdots + a_{qp}X_p \end{pmatrix} \sim N_q\left(\boldsymbol{A}\boldsymbol{\mu}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^t\right).$$

Also if $\boldsymbol{b}$ is a vector of constants in $\mathbb{R}^p$ then

$$\boldsymbol{X} + \boldsymbol{b} \sim N_p(\boldsymbol{\mu} + \boldsymbol{b}, \boldsymbol{\Sigma}).$$

Correlation and Independence

Theorem 9

Suppose that

$$\begin{pmatrix} \boldsymbol{X}_1 \\ \hdashline \boldsymbol{X}_2 \end{pmatrix} \sim N_{q_1+q_2}\left(\begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \hdashline \boldsymbol{\mu}^{(2)} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hdashline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right)$$

The $\boldsymbol{X}_1$ *and* $\boldsymbol{X}_2$ *are independent if, and only if,* $\boldsymbol{\Sigma}_{12} = \mathbf{0}_{q_1 \times q_2}$.

A Useful Lemma

Lemma 2

Suppose that

$$\left(\begin{array}{c}\boldsymbol{X}_1\\ \hdashline \boldsymbol{X}_2\end{array}\right) \sim N_{q_1+q_2}\left(\left(\begin{array}{c}\boldsymbol{\mu}^{(1)}\\ \hdashline \boldsymbol{\mu}^{(2)}\end{array}\right), \left(\begin{array}{c:c}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \hdashline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{array}\right) \right)$$

The the random vectors $\boldsymbol{X}_1$ *and* $\boldsymbol{X}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{X}_1$ *are independent.*

Marginal Distributions

Theorem 10

Suppose that $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *and we partition* $\boldsymbol{X}$ *as*

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \hdashline \boldsymbol{X}_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ \vdots \\ X_q \\ \hdashline X_{q+1} \\ \vdots \\ X_p \end{pmatrix}$$

Let the corresponding partitions of $\boldsymbol{\mu}$ *and* $\boldsymbol{\Sigma}$ *be*

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \hdashline \boldsymbol{\mu}^{(2)} \end{pmatrix} \qquad \boldsymbol{\Sigma} = \left(\begin{array}{c:c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hdashline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array}\right)$$

Then $\boldsymbol{X}_1 \sim N_q(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$.

Conditional Distribution

Theorem 11

Suppose that

$$\begin{pmatrix} \boldsymbol{X}_1 \\ \hdashline \boldsymbol{X}_2 \end{pmatrix} \sim N_p \left(\begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \hdashline \boldsymbol{\mu}^{(2)} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hdashline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right)$$

The the conditional distribution of the q-dimensional random vector $\boldsymbol{X}_1$ given that $\boldsymbol{X}_2 = \boldsymbol{x}_2$ is multivariate normal with mean vector

$$\boldsymbol{\mu}_{1.2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2)$$

and covariance matrix

$$\boldsymbol{\Sigma}_{1.2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$$

Contours of the Multivariate Normal Density

* A contour is a set of points for which the density is constant

$$\{\boldsymbol{x} \in \mathbb{R}^p : f(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \;=\; k\}$$

* Since $\boldsymbol{x}$ only appears in the exponent of the density we only need to consider this part and we can write a contour as

$$\left\{\boldsymbol{x} \in \mathbb{R}^p : (\boldsymbol{x} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \;=\; c^2\right\}$$

* That is a contour of the multivariate normal random density describes an ellipsoid in $\mathbb{R}^p$.

* If the covariance matrix is diagonal, then this ellipsoid becomes a spheroid.

Contours of the Multivariate Normal Density

* The quantity $(\boldsymbol{x}-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ is often referred to as the Mahalanobis Distance of the point $\boldsymbol{x}$ from $\boldsymbol{\mu}$.

* The axes of the ellipsoid will be oriented along the directions of the normalized eigenvectors of $\boldsymbol{Sigma}$.

* The following result is useful for finding probabilities within the ellipsoid.

Theorem 12

Suppose that $\boldsymbol{X}$ is distributed as a Normal$_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$ random vector. Then the random variable

$$(\boldsymbol{X}-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu}) \;\sim\; \chi^2_p.$$

Axes of the Multivariate Normal Density

* All of the contours of the multivariate normal are ellipses and have the same centre ($\boldsymbol{\mu}$) and orientation.

* The orientation of an ellipse is given by its largest (principal) axis.

* The orientation of the principal axis is given by the eigenvector of $\boldsymbol{\Sigma}$ corresponding to the largest eigenvalue, λ_1.

* For the ellipse

$$\left\{\boldsymbol{x} \in \mathbb{R}^p : (\boldsymbol{x}-\boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu}) \;=\; c^2\right\}$$

  the half-length of the principal axis is $c\sqrt{\lambda_1}$.

* Eigenvectors corresponding to other eigenvalues give the orientations of the other $p-1$ axes in order of length.

Axes in the Independent Case

* We saw that the components of $\boldsymbol{X}$ are independent if, and only if, Σ is a diagonal matrix.

* In that case the eigenvalues will be equal to the diagonal elements (in order) and the eigenvectors will be of the form

$$e_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

* Thus a characterization of independence is that the ellipsoid

$$\left\{ \boldsymbol{x} \in \mathbb{R}^p : (\boldsymbol{x} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) = c^2 \right\}$$

is oriented parallel to the co-ordinate axes.

* If $\Sigma = \sigma^2 \boldsymbol{I}_p$ then all eigenvalues will be equal to σ^2 and the eigenvectors will be still be of the form above so the contours become circles centred at $\boldsymbol{\mu}$.