III - MULTIVARIATE RANDOM VARIABLES

A random vector, or multivariate random variable, is a vector of $n$ scalar random variables. The random vector is written

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$$

A random vector is also said an $n$-dimensional random variable.

1 Joint ad marginal densities

Every random variable has a distribution function.

**DEFINITION III.1A - (DISTRIBUTION)** - The $n$-dimensional random variable $X$ has the distribution function

$$F(x_1, \ldots, x_n) = P\{X_1 \leq x_1, \ldots, X_n \leq x_n\}. \quad (2)$$

A random vector, $X$, is called absolutely continuous if there exists the joint probability density $f$ such that

$$F(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} f(t_1, \ldots, t_n) dt_1 \ldots dt_n. \quad (3)$$

A random vector, $X$, is called discrete if the set of its values is discrete, that is either finite or countable. It means that $X : \Omega \to R^n$, where $\Omega$ is discrete. In this case the joint probability function (or, less properly, the joint discrete density) is defined as

$$f(x_1, \ldots, x_n) = P\{X_1 = x_1, \ldots, X_n = x_n\}. \quad (4)$$

The joint distribution and the joint probability function are related by

$$F(x_1, \ldots, x_n) = \sum_{t_1 \leq x_1} \ldots \sum_{t_n \leq x_n} f(t_1, \ldots, t_n). \quad (5)$$
Ex. III.1B \((n = 2)\) The joint distribution function of the random vector \((X, Y)\) is defined by the relation

\[
F(x, y) = P(X \leq x, Y \leq y)
\]

The function \(F(x, y)\) is defined on \(R^2\) and, as a function each variable, is monotone non-decreasing and such that

\[
\lim_{x,y \to +\infty} F(x, y) = 1, \quad \lim_{x \to -\infty} F(x, y) = \lim_{y \to -\infty} F(x, y) = 0
\]

To know \(F(x, y)\) allows to know the probability that \((X, Y)\) falls in any given rectangle \((a, b] \times (c, d]\):

\[
P[(X, Y) \in (a, b] \times (c, d)] = F(b, d) - F(a, d) - F(b, c) + F(c, d).
\]

Since rectangles approximate any measurable set of \(R^2\), if we know \(F\) then we know \(P[(X, Y) \in A]\) for any measurable set \(A \subset R^2\); that is, the distribution function \(F(x, y)\) thoroughly describes the distribution measure of the couple \((X, Y)\).

If the random vector \((X, Y)\) is absolutely continuous, i.e. if there is an integrable function \(f(x, y)\) such that

\[
f(x, y) \geq 0, \quad \int_{R^2} f(x, y)dx\,dy = 1, \quad F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(\xi, \eta)d\xi\,d\eta,
\]

then

\[
\int_A f(x, y)dx\,dy = P[(X, Y) \in A], \quad \forall \text{ measurable set } A \subset R^2,
\]

which illustrates the meaning of the joint probability density \(f(x, y)\).

An example of an absolutely continuous random vector: \((X, Y)\) with probability density and distribution function

\[
f(x, y) = \frac{1}{2\pi} \exp[-\frac{1}{2}(x^2 + y^2)].
\]

For example \((X, Y)\) falls in the unit disk \(D = \{(x, y): \ x^2 + y^2 \leq 1\}\) with probability

\[
\int_D f(x, y)dx\,dy = \int_0^1 \int_0^{2\pi} \frac{1}{2\pi} e^{-r^2/2} r\,dr\,d\theta = \frac{2\pi}{2\pi} \left[-e^{-r^2/2}\right]_0^1 = 1 - e^{-1/2}.
\]

An example of a discrete random vector is \((X, Y)\) with \(X = \text{score got by tossing a die}, Y = \text{score got by tossing another die}. \) Then the joint probability
function has the value $1/36$ for any $(x_i, y_j) \in \{1, ..., 6\}^2$ and zero elsewhere. The distribution function

$$F(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} P(X = x_i, Y = y_j)$$

is computed by counting and dividing by 36. For example $F(2, 4) = 8/36$, $F(3, 3) = 9/36$, ..., $F(6, 6) = 1$. △

For a sub-vector $S = (X_1, ..., X_k)^T$ ($k < n$) of the random vector $X$, the marginal density function is

$$f_S(x_1, ..., x_k) = \int_{-\infty}^{\infty} f(x_1, ..., x_n) \, dx_{k+1} \cdots dx_n$$

in the continuous case, and

$$f_S(x_1, ..., x_k) = \sum_{x_{k+1}} \cdots \sum_{x_n} f(x_1, ..., x_n) \, dx_{k+1} \cdots dx_n$$

if $X$ is discrete. In both cases the marginal distribution function is

$$F_S(x_1, ..., x_k) = F(x_1, ..., x_k, \infty, ..., \infty).$$

EX. III.1C

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{8}, & 0 \leq y \leq x \leq 4 \\ 0, & \text{else} \end{cases}$$

Then the marginal density function of $X$ is

$$f_X(x) = \int_R f_{X,Y}(x, y) \, dy = \begin{cases} \int_0^x \frac{1}{8} \, dy = \frac{1}{8} x, & x \in [0, 4] \\ 0, & x \notin [0, 4] \end{cases}$$

DEF. III.1D (INDEPENDENCE)

The random vectors $X$ and $Y$ are independent \(^1\) if

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y), \, \forall x, y \tag{9}$$

\(^1\)In most textbooks the $n$–dimensional and $m$–dimensional r.v. $X, Y$ are said to be independent if

$$P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B), \forall \text{ measurable } A \subset R^n, B \subset R^m.$$

A suitable choice of $A, B$ leads directly to the equivalent definition (10)
which corresponds to
\[ F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y), \forall x, y \] (10)

2 Moments of scalar random variables

Evolution presupposes the existence of something that can develop. The theory says nothing about where this "something" came from. Furthermore, questions about the being, essence, dignity, mission, meaning and wherefore of the world and man cannot be answered in biological terms. (YouCat 42)

An average of the possible values of \( X \), each value being weighted by its probability,
\[ E[X] = \sum_x x \cdot P\{X = x\}, \]
is \textit{ad litteram} the notion of expectation for a discrete variable \( X \) (provided that such series is absolutely convergent). For continuous variables, the sums are replaced by integrals.

DEFINITION III.2A (EXPECTATION)
The expectation (or mean value) of an absolutely continuous variable \( X \) with density function \( f_X \) is
\[ E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) \, dx \] (11)
provided that the integral is absolutely convergent, i.e. \( E[|X|] < \infty \).

If \( X \) is a random variable and \( g \) is a function, then \( Y = g(X) \) is also a random variable. To calculate the expectation of \( Y \), we could find \( f_Y \) and use (11). However the process of finding \( f_Y \) can be complicated; instead, we can simply use
\[ E[Y] = E[g(X)] = \int_R g(x)f_X(x) \, dx \] (12)
This provides a method for calculating the "moments" of a distribution.

DEF. III.2B (MOMENTS)
The \( n \)-th moment of \( X \) is
\[
E[X^n] = \int_R x^n f_X(x) \, dx
\]
and the \( n \)-th central moment is
\[
E[(X - E[X])^n] = \int_R (x - E[X])^n f_X(x) \, dx
\]

The second central moment is also called the variance, and the variance of \( X \) is given by
\[
\]

Now let \( X \) be an \( n \)-dimensional random variable and put \( Y = g(X) \), where \( g \) is a function. As in (12) we have
\[
E[Y] = E[g(X)] = \int_{\mathbb{R}^n} g(x_1, \ldots, x_n) f_X(x_1, \ldots, x_n) \, dx_1 \ldots dx_n
\]

This gives sense to the mean of a linear combination of variables, and to any type of moments. Since
\[
E[aX_1 + bX_2] = aE[X_1] + bE[X_2]
\]
the expectation operator is a linear operator.

Of special interest is the covariance of \( X_i \) and \( X_j \):

\text{DEF: III.2C - Covariance - The covariance of two scalar r.v.'s \( X_{1,2} \) is}
\[
\text{Cov}[X_1, X_2] = E[(X_1 - E[X_1])(X_2 - E[X_2])]
\]

The covariance gives information about the simultaneous variation of two variables and is useful to find whether \( X_1, X_2 \) are dependent on each other. Since the expectation is linear, the covariance is a bilinear operator, i.e. linear in each of the two arguments, whence its calculation rule:
\[
\text{Cov}[aX_1 + bX_2, cX_3 + dX_4] =
\]
\[
= ac \text{ Cov}[X_1, X_3] + ad \text{ Cov}[X_1, X_4] + bc \text{ Cov}[X_2, X_3] + bd \text{ Cov}[X_2, X_4]
\]
for constants \( a, b, c, d \) and r.v. \( X_1, \ldots, X_4 \).
PROP. III.2D  - For two scalar random variables $X, Y$ we have the formulae:

$$\text{Cov}(X, Y) = E(XY) - \mu_x \mu_y$$

(18)

$$\text{Var}(X + Y) = \text{Var}(X) + 2 \cdot \text{Cov}(X, Y) + \text{Var}(Y)$$

(19)

Proof

$$E[(X - \mu_x) \cdot (Y - \mu_y)] = E[XY] - \mu_x E[Y] - \mu_y E[X] + \mu_x \mu_y =$$

$$E[XY] - 2\mu_x \mu_y + \mu_x \mu_y = E[XY] - \mu_x \mu_y$$

In particular, as we already know,

$$\text{Var}(X) \equiv \text{Cov}(X, X) = E[X^2] - \mu_x^2.$$

Besides,

$$E[(X + Y - (\mu_x + \mu_y))^2] = E[(X - \mu_x)^2 + 2(X - \mu_x)(Y - \mu_y) + (Y - \mu_y)^2] =$$

$$= \text{Var}(X) + 2 \cdot \text{Cov}(X, Y) + \text{Var}(Y) \quad \triangle$$

DEF. III.2D - Correlation coefficient - The correlation coefficient, or correlation, of two scalar r.v.’s $X_1, X_2$ is

$$\rho \equiv \rho(X_1, X_2) := E\left[\frac{X_1 - \mu_1}{\sigma_1} \cdot \frac{X_2 - \mu_2}{\sigma_2}\right]$$

THM. III.2D - The covariance and the correlation of two scalar r.v.’s $X, Y$ satisfy

$$\text{Cov}(X, Y)^2 \leq \text{Var}(X) \text{Var}(Y) \quad \text{(Schwarz’s inequality),}$$

$$-1 \leq \rho(X, Y) \leq 1$$

respectively.

Proof

Since the expectation of a nonnegative variable is always nonnegative,

$$0 \leq E[(\theta|X| + |Y|)^2] = \theta^2 E[X^2] + 2\theta E[|XY|] + E|Y|^2, \quad \forall \theta.$$
Thus a polynomial with degree 2 (in $\theta$) does not take negative values, so that the discriminant must be $\leq 0$:

$$E[|XY|^2] - E[X^2]E[Y^2] \leq 0, \quad \text{or} \quad E[|XY|^2] \leq E[X^2]E[Y^2].$$

Replacing $X$ and $Y$ by $X - \mu_x$ and $Y - \mu_y$, and using $|\int f(t)g(t)dt| \leq \int |f(t)g(t)|dt$, this implies

$$E[(X - \mu_x)(Y - \mu_y)]^2 \leq E[(X - \mu_x)(Y - \mu_y)]^2 \leq E[(X - \mu_x)^2]E[(Y - \mu_y)^2] = Var(X)Var(Y).$$

Since $\rho = \text{Cov}(X,Y)/\sqrt{Var(X)Var(Y)}$, this also means

$$[\rho(X,Y)]^2 \leq 1, \text{ i.e. } -1 \leq \rho \leq +1. \quad \triangle$$

**DEFINITION III.2E - Uncorrelation** - Two scalar random variables $X$ and $Y$ are said **uncorrelated**, or orthogonal, when $\text{Cov}(X,Y) = 0$.

Since $\text{Cov}(X,Y) = E(XY) - \mu_x\mu_y$, $X$ and $Y$ independent $\Rightarrow$ $X, Y$ uncorrelated

However

$X, Y$ uncorrelated does not imply $X, Y$ independent

as we see in the following example.

**EX. III.2F** - Let $U$ be uniformly distributed in $[0,1]$ and consider the r.v.’s

$$X = \sin 2\pi U, \quad Y = \cos 2\pi U.$$ 

$X$ and $Y$ are not independent since $Y$ has only two possible values if $X$ is known. However they are orthogonal or **uncorrelated**:

$$E(X) = \int_0^1 \sin 2\pi t \, dt = 0, \quad E(Y) = \int_0^1 \cos 2\pi t \, dt = 0,$$

so that

$$\text{Cov}(X,Y) = E(X \cdot Y) = \int_0^1 \sin 2\pi t \cos 2\pi t \, dt = \frac{1}{2} \int_0^1 \sin 4\pi t \, dt = 0.$$ 

In this case the covariance is zero, but the variables are not independent.
3 Moments of multivariate random variables

We are interested in defining a (vector) first moment and a (matrix) second moment of a multivariate random variable.

**DEF. III.3A (EXPECTATION OF A RANDOM VECTOR)**
The expectation (or the mean value) of the random vector $X$ is

$$
\mu = E[X] = \begin{pmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{pmatrix}
$$

(20)

**DEFINITION III.3B (COVARIANCE matrix)** The covariance matrix of the random vector $X$ is

$$
\Sigma_X = Var[X] = E[(X - \mu)(X - \mu)^T] = 
\begin{pmatrix} Var[X_1] & Cov[X_1, X_2] & \ldots & Cov[X_1, X_n] \\ Cov[X_2, X_1] & Var[X_2] & \ldots & Cov[X_2, X_n] \\ \vdots & \vdots & \ddots & \vdots \\ Cov[X_n, X_1] & Cov[X_n, X_2] & \ldots & Var[X_n] \end{pmatrix}
$$

(21)

Sometimes we shall use the notation

$$
\begin{pmatrix} \sigma_1^2 & \sigma_{12} & \ldots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \ldots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \ldots & \sigma_n^2 \end{pmatrix}
$$

Both $\sigma_{ii}$ and $\sigma_i^2$ are possible notations for the variance of $X_i$.

By the calculation rule (17), the variance of a linear combination of $X_1, X_2$ is known from the variance matrix of $(X_1, X_2)$:

$$
Var[z_1X_1 + z_2X_2] = z_1^2 Var(X_1) + 2z_1z_2 Cov(X_1, X_2) + z_2^2 Var(X_2).
$$

and in general, for any $n \geq 2$, it is the the quadratic form of $\Sigma$ calculated in the coefficients $z_1, \ldots, z_n$:

$$
Var[z_1X_1 + \ldots + z_nX_n] = z^T \Sigma z.
$$
According to III.2C, the correlation of two random variables $X_i$, $X_j$ is:

$$
\rho_{ij} = \frac{Cov[X_i, X_j]}{\sqrt{Var[X_i]Var[X_j]}} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}
$$

Equations (21, 23) lead to the definition

**DEFINITION III.3C**

The correlation matrix for $X$ is

$$
\begin{pmatrix}
1 & \rho_{12} & \cdots & \rho_{1n} \\
\rho_{21} & 1 & \cdots & \rho_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
\rho_{n1} & \rho_{n2} & \cdots & 1
\end{pmatrix}
$$

**THEOREM III.3D**

The covariance matrix $\Sigma$ and the correlation matrix $R$ are a) symmetric and b) nonnegative.

**Proof.** (a) The symmetry is obvious. (b) By definition of variance, $Var(z_1X_1 + \ldots + z_nX_n) \geq 0$ for any $z \in \mathbb{R}^n$. Hence the quadratic form $z^T \Sigma z = \nonumber Var[z^T X] \geq 0$, $\forall z$, i.e. the matrix $\Sigma$ is nonnegative semidefinite. $\triangle$.

**REMARK III.3E** - Very often $\Sigma$ is positive definite, i.e.

$$
z^T \Sigma z \geq 0 \quad \text{and} \quad z^T \Sigma z = 0 \implies z = (0, \ldots, 0)
$$

When $\Sigma$ is positive definite, we write $\Sigma > 0$. For our purposes, we always assume $\Sigma > 0$.

**COROLLARY III.3F** - Linear transformations of random vectors

Let $X = (X_1, \ldots, X_n)^T$ have mean $\mu$ and covariance $\Sigma_X$. Now we introduce the new random variable $Y = (Y_1, \ldots, Y_n)^T$ by the transformation

$$
Y = a + BX
$$

where $a$ is $(n \times 1)$ vector and $B$ is a $(n \times n)$ matrix. Then the expectation and variance matrix of $Y$ are:

$$
E[Y] = E[a + BX] = a + BE[X] \quad (24)
$$

$$
\Sigma_Y = Var[a + BX] = Var[BX] = BVar[X]B^T \quad (25)
$$
Proof
The first formula is obvious since the expectation operator is linear. As for the variance of \( B X \) we have:

\[
(\text{Var}[a + BX])_{i,j} = \text{Cov}[(BX)_i, (BX)_j] = \text{Cov} \left[ \sum_l b_{i,l} X_l, \sum_m b_{j,m} X_m \right] = \sum_{l,m} b_{i,l} b_{j,m} \text{Cov}(X_l, X_m) = (B \Sigma X B^T)_{i,j}.
\]

\( \triangle \)

4 The multivariate normal distribution

We assume that \( X_1, ..., X_n \) are independent random variables with means \( \mu_1, ..., \mu_n \), and variances \( \sigma_1^2, ..., \sigma_n^2 \). We write \( X_i \in N(\mu_i, \sigma_i^2) \). Now, define the random vector \( X = (X_1, ..., X_n)^T \). Because the random variables are independent, the joint density is the product of the marginal densities:

\[
f_X(x_1, ..., x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)
\]

\[
= \prod_{i=1}^n \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left[ -\frac{(x_i - \mu_i)^2}{2\sigma_i^2} \right]
\]

\[
\frac{1}{(\prod_{i=1}^n \sigma_i)(2\pi)^{n/2}} \exp \left[ -\frac{1}{2} \sum_{i=1}^n \left( \frac{x_i - \mu_i}{\sigma_i} \right)^2 \right].
\]

By introducing the mean \( \mu = (\mu_1, ..., \mu_n)^T \), and the covariance

\[
\Sigma_X = \text{Var}[X] = \text{diag}(\sigma_1^2, ..., \sigma_n^2),
\]

this is written

\[
f_X(x) = \frac{1}{(2\pi)^{n/2} \sqrt{\det \Sigma_X}} \exp \left[ -\frac{1}{2} (x - \mu)^T \Sigma_X^{-1} (x - \mu) \right] \quad (26)
\]

A generalization to the case where the covariance matrix is a full matrix leads to the following

DEF. III.4A - The multivariate normal distribution -

The joint density function for the \( n \)-dimensional random variable \( X \) with mean \( \mu \) and covariance matrix \( \Sigma \) is
\[ f_X(x) = \frac{1}{(2\pi)^{n/2} \sqrt{\det \Sigma_X}} \exp \left[ -\frac{1}{2}(x - \mu)^T \Sigma_X^{-1} (x - \mu) \right] \]  

(27)

where \( \Sigma > 0 \). We write \( X \in N(\mu, \Sigma) \). If \( X \in N(0, I) \) we say that \( X \) is standardized normally distributed.

THM. III.4B - Within the class of normal variables, \( X_1, \ldots, X_n \) are uncorrelated if and only if they are independent.

Proof
\[ \text{Cov}(X_i, X_j) = 0, \ \forall i \neq j \] implies that the variance matrix is diagonal. Thus the quadratic form in the exponent of \( f \) is a diagonal form. So the density is the product of the marginal densities, hence the \( X_i \)'s are independent. \( \triangle \)

THM. III.4C - Principal components of \( N(\mu; \Sigma) \).

Let \( X \in N(\mu, \Sigma) \), with values in \( \mathbb{R}^n \). Each level surface (or "contour") of the joint density \( f(x_1, \ldots, x_n) \) is an ellipsoid
\[ \{ x : (x - \mu)^T \Sigma^{-1} (x - \mu) = c^2 \} \]  

(28)
in \( \mathbb{R}^n \), \( c \) being a constant. The axes of the ellipsoid are
\[ \pm c \sqrt{\lambda_i} e_i \]

where \((\lambda_i, e_i)\) are the eigenvalue-eigenvector pairs of the covariance matrix \( \Sigma \) \((i = 1, \ldots, n)\).

If the matrix \( G \) is defined with the normalized eigenvectors \( e_i \) as columns, then the equation
\[ Y = G^T (X - \mu), \]  

(29)
determines independent and normally distributed variables \( Y_i \) with mean zero and variance \( \lambda_i \). Such \( Y_i \)'s are said "principal components" of the normal vector \( X \).

Proof
The covariance matrix \( \Sigma \) is symmetric and definite positive. Thus by the spectral decomposition theorem,
\[ G^T \Sigma G = \Lambda, \]  

(or \( G^T \Sigma^{-1} G = \Lambda^{-1} \))
where $\Lambda = \text{Diag}(\lambda_1, \ldots, \lambda_n)$ is the diagonal matrix of (positive) eigenvalues of $\Sigma$, while $G$ is an orthogonal matrix whose columns are the corresponding normalized eigenvectors. By the principal component transformation

$$y = G^T(x - \mu), \quad \text{i.e.} \quad x - \mu = Gy$$

the ellipsoid becomes

$$(Gy)^T \Sigma^{-1} Gy = c^2, \quad \text{i.e.} \quad y^T G^T \Sigma^{-1} G y = c^2,$$

i.e. $y^T \Lambda^{-1} y = c^2 \quad \text{i.e.} \quad \sum_{i=1}^n \frac{1}{\lambda_i} y_i^2 = c^2.$

This is the equation of an ellipsoid with semiaxes of length $c \sqrt{\lambda_i}$. $y_1, \ldots, y_n$ represent the proper axes of the ellipsoid. The direction of the axes are given by the eigenvectors of $\Sigma$ since $y = G^T(x - \mu)$ is a rotation and $G$ is orthogonal with eigenvectors of $\Sigma$ as columns. △

**EX. III.4D Principal components and contours of a two-dimensional normal vector.**

To obtain contours of a bivariate normal density we have to find the eigenvalues and eigenvectors of $\Sigma$. Suppose

$$\mu_1 = 5, \mu_2 = 8, \Sigma = \begin{pmatrix} 12 & 9 \\ 9 & 25 \end{pmatrix}.$$  

Here the eigenvalue equation becomes

$$0 = |\Sigma - \lambda I| = \det \begin{pmatrix} 12 - \lambda & 9 \\ 9 & 25 - \lambda \end{pmatrix} = (12 - \lambda)(25 - \lambda) - 81 = 0.$$  

The equation gives $\lambda_1 = 29.601$, $\lambda_2 = 7.398$. The corresponding normalized eigenvectors are

$$e_1 = \begin{pmatrix} .455 \\ .890 \end{pmatrix}, \quad e_2 = \begin{pmatrix} .890 \\ -.455 \end{pmatrix}.$$  

The ellipse of constant probability density (??) has center at $\mu = (5, 8)^T$ and axes $\pm 5.44 \, c \, e_1$ and $\pm 2.72 \, c \, e_2$. The eigenvector $e_1$ lies along the line which makes an angle of $\cos^{-1}(0.455) = 62.92^\circ$ with the $X_1$ axis and passes through the point $(5, 8)$ in the $(X_1, X_2)$ plane. By principal component transformation $y = G^T(x - \mu)$ where $G = (e_1, e_2)$, the ellipse (??) has the equation

$$\{y : \sum_{i=1}^n \frac{1}{\lambda_i} y_i^2 = c^2\}.$$
The ellipse as center at \((y_1 = 0, y_2 = 0)\) and axes \(\pm c \cdot \sqrt{\lambda_i} f_i\), where \(f_i, i = 1, 2,\) are the eigenvectors of \(L = Diag(\lambda_1, \lambda_2)\). In terms of the new co-ordinate axes \(Y_1, Y_2\), the eigenvectors are \(f_1 = (1, 0)^T, f_2 = (0, 1)^T\). The lengths of the axes are \(c \sqrt{\lambda_i}, i = 1, 2\). The axes \(Y_1, Y_2\) are the Principal Components of \(\Sigma\).

```r
> Sigma <- matrix(c(12,9,9,25), nrow=2, ncol=2)
> eigen(Sigma)$values
[1] 29.601802  7.398198
> eigen(Sigma)$vectors
   [,1]       [,2]
[1,] 0.4552524 0.8903624
[2,] 0.8903624-0.4552524
```

THM. III.4E (n-dimensional standardization)

Let \(X \sim N(\mu, \Sigma)\), with dimension \(n\). Then there exists a square matrix \(C\) such that

\[
U := C^{-1}(X - \mu) \sim N(0; I).
\]

i.e. the components \((U_1, \ldots, U_n)\) are independent \(N(0; 1)\) variables.

Proof

Due to symmetry of the positive matrix \(\Sigma\), there always exists a non singular real matrix \(C\) such that \(\Sigma = CC^T\) (a matrix "square root" of \(\Sigma\)). Setting \(U = C^{-1}(X - \mu)\), by means of Corollary III.3G we find:

\[
E[U] = C^{-1}E[X - \mu] = 0
\]

\[
Var[U] = C^{-1}Var[X](C^{-1})^T = C^{-1}CC^T(C^T)^{-1} = I. \quad \triangle
\]

5 Distributions derived from the normal distribution

Most of the test quantities used in Statistics are based on distributions derived from the normal one.
Any linear combination of normally distributed random variables is normal. If, for instance, \( X \in N(\mu; \Sigma) \), then the linear transformation \( Y = a + BX \) defines a normally distributed random variable

\[
Y \in N(a + B\mu, B\Sigma B^T)
\]  

(30)

Compare with Corollary III.3E, where the transformation rule \( \Sigma \rightarrow B\Sigma B^T \) of the variance of \( X \), for any affine map \( X \rightarrow a + BX \), is established.

Let \( U = (U_1, ..., U_n)^T \) be a vector of independent \( N(0; 1) \) random variables. The (central) \( \chi^2 \) distribution with \( n \) degrees of freedom is obtained as the squared sum of \( n \) independent \( N(0; 1) \) random variables, i.e.

\[
X^2 = \sum_{i=1}^{n} U_i^2 = U^T U \sim \chi^2(n)
\]  

(31)

From this it is clear that if \( Y_1, ..., Y_n \) are independent \( N(\mu_i; \sigma^2_i) \) r.v.'s, then

\[
X^2 = \sum_{i=1}^{n} \left( \frac{Y_i - \mu_i}{\sigma_i} \right)^2 \sim \chi^2(n)
\]

since \( U_i = (Y_i - \mu_i)/\sigma_i \) is \( N(0; 1) \) distributed.

THM. III.5A If \( X_i \), for \( i = 1, ..., n \), are independent normal variables from a population \( N(\mu; \sigma^2) \), then the sample mean and the sample variance follow a normal and a chi squared distribution, respectively:

\[
\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \Rightarrow \bar{X} \sim N(\mu, \frac{\sigma^2}{n})
\]

\[
S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2 \Rightarrow \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).
\]

Proof.
\( \bar{X} \) is normal as a linear transformation of normal variables. By linearity \( E[\bar{X}] = \frac{1}{n} \sum_i E[X_i] = \frac{1}{n} \cdot n\mu = \mu \) and by independence

\[
Var\left[\frac{1}{n} \sum_{i=1}^{n} X_i \right] = \frac{1}{n^2} \sum_i Var[X_i] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.
\]

whence \( \bar{X} \sim N(\mu, \sigma^2/n) \).

Now using this fact and the definition of \( \chi^2 \),

\[
\sum_i \left( \frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(n), \quad \frac{(\bar{X} - \mu)^2}{\sigma^2/n} = \frac{(\bar{X} - \mu)^2}{\sigma^2/n} \sim \chi^2(1)
\]
Now writing $X_i - \mu = X_i - \bar{X} + \bar{X} - \mu$, the following decomposition can be verified:

$$
\sum_{i=1}^{n} \left( \frac{X_i - \mu}{\sigma} \right)^2 = \sum_{i=1}^{n} \left( \frac{X_i - \bar{X}}{\sigma^2} \right)^2 + n\left( \frac{\bar{X} - \mu}{\sigma} \right)^2 = \frac{(n-1)S^2}{\sigma^2} + n\left( \frac{\bar{X} - \mu}{\sigma} \right)^2.
$$

Therefore the random variable

$$\frac{(n-1)S^2}{\sigma^2}$$

is the difference between a $\chi^2(n)$ r.v. and a $\chi^2(1)$ r.v., that is a r. v. with distribution $\chi^2(n-1)$. △.

DEF. III.5B -

The Student’s t distribution with $n$ degrees of freedom is obtained as

$$T = \frac{Z}{(K/n)^{1/2}} \sim t(n) \quad (32)$$

where $Z \in N(0;1)$, $K \sim \chi^2(n)$, and $Z$ and $K$ are independent.

The Fisher-Snedecor F distribution with $(n_1,n_2)$ degrees of freedom appears as the following ratio

$$F = \frac{K_1/n_1}{K_2/n_2} \sim F(n_1,n_2) \quad (33)$$

where $K_1 \sim \chi^2(n_1)$, and $K_2 \sim \chi^2(n_2)$, and $K_1, K_2$ are independent. It is clearly seen from (32) that $T^2 \sim F(1,n_1)$.

PROP. III.5C - For $i = 1,\ldots,n$, let the $X_i$ be i.i.d. variables from a normal population $N(\mu;\sigma^2)$. If $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$ then the estimated standardization of the mean

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

obeys a Student t distribution with $n - 1$ degrees of freedom.
Figure 1: Chi-squared density with 5 degrees of freedom

Figure 2: Student’s t density with 8 degrees of freedom
If two independent samples \( \{X_i\} \) and \( \{Y_j\} \) \((i = 1,\ldots,n_1; \ j = 1,\ldots,n_2)\) are taken from normal populations with the same variance, the ratio

\[
F = \frac{S_1^2}{S_2^2}
\]

obeys a Fisher's distribution with \( n_1 - 1, n_2 - 1 \) degrees of freedom.

**Proof** Since \( \bar{X} \sim N(\mu; \sigma^2/n) \), we have:

\[
T = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \cdot \frac{\sigma/\sqrt{n}}{S/\sqrt{n}} = \frac{U}{S/\sigma} = \frac{U}{\sqrt{K/(n-1)}}
\]

where

\[
U = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}, \quad K = \frac{S^2 \cdot (n-1)}{\sigma^2}
\]

are independent with \( U \sim N(0;1) \) and \( K \sim \chi^2(n-1) \).

Finally, under the assumptions of equal variance \( \sigma^2 \),

\[
\frac{S_1^2}{S_2^2} = \frac{\frac{1}{n_1-1}S_1^2 \cdot (n_1-1)/\sigma^2}{\frac{1}{n_2-1}S_2^2 \cdot (n_2-1)/\sigma^2} = \frac{K_1/(n_1-1)}{K_2/(n_2-1)}
\]

where \( K_1 \) and \( K_2 \) are independent, \( K_1 \sim \chi^2(n_1-1), \ K_2 \sim \chi^2(n_2-1) \).
\[\triangle\]

In the language and environment R, each distribution above described has four functions, as described in below (cited from: R reference card).

Distributions

\begin{itemize}
  \item \texttt{rnorm(n, mean=0, sd=1)} Gaussian (normal)
  \item \texttt{rt(n, df)} Student (t)
  \item \texttt{rf(n, df1, df2)} Fisher\-Snedecor (F) (c2)
  \item \texttt{rchisq(n, df)} Pearson
  \item \texttt{rbinom(n, size, prob)} binomial
  \item \texttt{runif(n, min=0, max=1)} uniform
\end{itemize}

All these functions can be used by replacing the letter \texttt{r} with \texttt{d}, \texttt{p} or \texttt{q} to get, respectively, the probability density (\texttt{dfunc(x, ...)}), the cumulative probability density (\texttt{pfunc(x, ...)}), and the value of quantile (\texttt{qfunc(p, ...)}, with \(0 < p < 1\)).

Some examples:

\begin{verbatim}
> pnorm(5.8, mean=2, sd=3)  # funz. distribuzione N(2;9) in 5.8
[1] 0.8973627

> dnorm(2.7,mean=0,sd=5)    # funz.densita' N(0;25) in 2.7
[1] 0.0689636

> qt(0.95,df=13)            # quantile 0.95 di "t" di Student
[1] 1.770933                 # con 13 gradi di liberta'

> qchisq(0.99,df=10)        # quantile 0.99 di "chi quadro"
[1] 23.20925                 # con 10 gradi di liberta'

> z <- seq( 0,40,length=1000)
> plot(z, dchisq(z,df=10),ty="l")  # densita' di chi quadro
                                  # con 10 gradi di liberta'
\end{verbatim}