# 7. RANDOM VECTORS

It is possible to extend all the previous definitions to vectors of functions on $\Omega$:

$$X(\omega) = \begin{pmatrix} X_1(\omega) \\ \vdots \\ X_n(\omega) \end{pmatrix}$$

<u>DEFINITION</u> $X : \Omega \to \mathbb{R}^n$ is a random vector on $(\Omega, \mathcal{F}, P)$ if

$$\{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$$

$\forall\ B \in \mathcal{B}(\mathbb{R}^n)$

<u>REMARK</u> Since a set of generators for $\mathcal{B}(\mathbb{R}^n)$ is

$$\{(-\infty, a_1) \times \dots \times (-\infty, a_n),\ a_1, \dots, a_n \in \mathbb{R}\}$$

it is sufficient to consider in the above definition $B$ to be any set like

$$(-\infty, a_1) \times \dots \times (-\infty, a_n)$$

**DEFINITION**. The joint distribution function $F_{X_1,\dots,X_n}(x_1,\dots,x_n)$ of a random vector $X(\omega)=\begin{pmatrix} X_1(\omega) \\ \vdots \\ X_n(\omega) \end{pmatrix}$ is

$$F_{X_1\dots X_n}(x_1,\dots,x_n)=$$
$$P\{\omega\in\Omega:\ X_i(\omega)\in(-\infty,x_i),\ i=1,\dots,n\}$$

The functions $F_{X_i}(x_i)$ are called MARGINAL distribution functions. If $F_X(x)$ is differentiable:

$$dF_X(x)=dF_{X_1\dots X_n}(x_1,\dots,x_n)=$$
$$=\frac{\partial F_{X_1\dots X_n}}{\partial x_1\dots\partial x_n}(x_1,\dots,x_n)\,dx_1\dots dx_n$$
$$=\frac{\partial F_X}{\partial x}\,dx$$

and

DEFINITION. The joint probability density function $p_X(x)$ of a random vector $X(\omega)$ is

$$p_X(x) = p_{X_1 \dots X_n}(x_1, \dots, x_n) = \frac{\partial F_X}{\partial x}(x)$$

The functions $p_{X_i}(x_i)$ are called MARGINAL probability density functions.

The marginal distributions and densities can be obtained from the joint distribution and density as follows:

$$P\{\omega \in \Omega : X_i(\omega) < x_i\} = F_{X_i}(x_i) =$$

$$= P\{\omega \in \Omega : X_i(\omega) < x_i,\ X_j(\omega) \in (-\infty, +\infty),\ j \neq i\}$$

$$= F_X(+\infty, \dots, +\infty, x_i, +\infty, \dots, +\infty)$$

and

$$p_{X_i}(x_i) = \frac{dF_{X_i}(x_i)}{dx_i} =$$

$$= \frac{d}{dx_i} F_X(+\infty, \dots, +\infty, x_i, +\infty, \dots, +\infty) =$$

$$= \frac{d}{dx_i} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{x_i} \cdots \int_{-\infty}^{\infty} p_{X_1 \ldots X_n}(x_1, \ldots, x_n)\, dx_1 \ldots dx_n$$

$$= \underbrace{\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty}}_{n-1 \text{ times}} p_{X_1 \ldots X_n}(x_1, \ldots, x_n)\, dx_1 \ldots dx_{i-1}\, dx_{i+1} \ldots dx_n$$

<u>DEFINITION</u>. The expectation of a random vector $X(\omega)$ is

$$E\{X(\omega)\} = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} x \, dF_{X_1 \ldots X_n}(x_1, \ldots, x_n)$$

$$= \begin{pmatrix} \int_{-\infty}^{\infty} x_1 \, dF_{X_1 \ldots X_n}(x_1, \ldots, x_n) \\ \vdots \\ \int_{-\infty}^{\infty} x_n \, dF_{X_1 \ldots X_n}(x_1, \ldots, x_n) \end{pmatrix}$$

$$= \begin{pmatrix} E\{X_1(\omega)\} \\ \vdots \\ E\{X_n(\omega)\} \end{pmatrix}$$

REMARK. Each $E\{X_i(\omega)\}$ can be calculated by using the marginal distribution $F_{X_i}(x_i)$:

$$E\{X_i(\omega)\} = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} x_i \, dF_{X_1 \ldots X_n}(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

$$\underset{\substack{\uparrow \\ \text{if a density} \\ \text{is available}}}{=} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} x_i \, p_{X_1 \ldots X_n}(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

$$= \int_{-\infty}^{+\infty} x_i \Big\{ \underbrace{\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty}}_{n-1 \text{ times}} p_{X_1 \ldots X_n}(x_1, \ldots, x_n) \, dx_1 \ldots dx_{i-1} \, dx_{i+1} \ldots dx_n \Big\} dx_i$$

$$= \int_{-\infty}^{+\infty} x_i \, p_{X_i}(x_i) dx_i = \int_{-\infty}^{+\infty} x_i \, dF_{X_i}(x_i)$$

**DEFINITION** The COVARIANCE MATRIX of a random vector $X(\omega)$ is

$$\Psi_X \triangleq E\left\{ (X - E\{X\})(X - E\{X\})^T \right\}$$

$$= \begin{pmatrix} [\Psi_X]_{11} & \cdots & [\Psi_X]_{1,n} \\ \vdots & & \\ [\Psi_X]_{n,1} & \cdots & [\Psi_X]_{n,n} \end{pmatrix}$$

where

$$[\Psi_X]_{i,j} = E\left\{ (X_i - E\{X_i\})(X_j - E\{X_j\}) \right\}$$

The CROSS-COVARIANCE MATRIX of two random $X$ and $Y$ is:

$$\Psi_{XY} \triangleq E\left\{ (X - E\{X\})(Y - E\{Y\})^T \right\}$$

**REMARK** Notice that $[\Psi_X]_{i,j} = [\Psi_X]_{j,i}\ \forall j,i$,

$[\Psi_X]_{i,i} = \sigma^2_{X_i}$ and

$[\Psi_X]_{i,j} = \sigma_{X_i X_j}$ for $i \neq j$

Since $\forall z \in \mathbb{R}^n$:

$$z^T \Psi_X z = z^T E\{(X - E\{X\})(X - E\{X\})^T\} z$$

$$= E\{((X - E\{X\})^T z)^T ((X - E\{X\})^T z)\}$$

$$= E\{\| (X - E\{X\})^T z \|^2\} \geq 0$$

then

$\Psi_X$ is a positive semidefinite (symmetric) matrix.

DEFINITION. The VARIANCE of a random vector $X(\omega)$ is:

$$\sigma_X^2 \triangleq \sum_{i=1}^{n} \sigma_{X_i}^2$$

$$= E\{\| X - E\{X\} \|^2\}$$

$$= \mathrm{Tr}\, \Psi_X$$

($\mathrm{Tr}(\Psi_X)$ is the trace of $\Psi_X$, the sum of the diagonal elements of $\Psi_X$).

<u>REMARK</u>. For random vectors

$$X(\omega) = \begin{pmatrix} X_1(\omega) \\ \vdots \\ X_n(\omega) \end{pmatrix}$$ for which

$X_i(\omega)$ and $X_j(\omega)$ are independent for all $i, j = 1, .., n$:

$$F_{X_1 \cdots X_n}(x_1, \dots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i)$$

$$\phi_{X_1 \cdots X_n}(x_1, \dots, x_n) = \prod_{i=1}^{n} \phi_{X_i}(x_i)$$

so that

$$E\{X_i \cdot X_j\} = E\{X_i\} E\{X_j\}$$

$\forall i, j$ ◭

# 7.1. GAUSSIAN RANDOM VECTORS

**DEFINITION.** A random vector $X(\omega)$ is GAUSSIAN if

$$\phi_X(x) = \frac{1}{(2\pi)^{\frac{n}{2}}[\det \Psi]^{1/2}} \cdot e^{-\frac{1}{2}(x-m)^T \Psi^{-1}(x-m)}$$

for some symmetric positive definite $\Psi$ and $m \in \mathbb{R}^n$.

We have
$$\boxed{\begin{aligned} E\{X\} &= m \\ \Psi_X &= \Psi \end{aligned}}$$
(see remark).

REMARK

$$[\Psi_X]_{i,j} = \int_{\mathbb{R}^n} (x_i - m)(x_j - m_j)\, p_X(x)\, dx$$

$$= \frac{1}{(2\pi)^{n/2}[\det \Psi]^{1/2}} \int_{\mathbb{R}^n} y_i y_j e^{-\frac{1}{2} y^T \Psi^{-1} y} dy =$$

(with $y \triangleq x - m$)

$$= \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^n} (R^{-1}z)_i (R^{-1}z)_j e^{-\frac{1}{2}z^T z} dz$$

(with $z \triangleq Ry$, $R = \psi^{-1/2}$ and $\psi^{-1/2}$ is the unique matrix such that $\psi^{-1/2} \cdot \psi^{-1/2} = \psi^{-1}$)

$$= \frac{1}{(2\pi)^{\frac{n}{2}}} \sum_{k=1}^{n} \sum_{h=1}^{n} [R^{-1}]_{ik} [R^{-1}]_{jh} \cdot \int_{\mathbb{R}^n} z_k z_h e^{-\frac{1}{2}\sum_{r=1}^{n} z_r^2} dz \qquad (*)$$

The random vector $Z = R(X - m)$ is such that

$$E\{Z\} = \psi^{-1/2} E\{X - m\} = 0$$

$$\psi_Z = \psi^{-1/2} E\{(X-m)(X-m)^T\} \psi^{-1/2}$$

$$= \psi^{-1/2} \psi \cdot \psi^{-1/2} = I$$

Moreover, $Z$ is gaussian since

$$p_Z(z) = p_X(R^{-1}z + m) \left|\det(\psi^{-1/2})\right|$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}z^T z} = \prod_{r=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_r^2}$$

(here we used the fact that:

$$\boxed{\begin{array}{l} Z = f(X),\ f \text{ differentiable and} \\ f^{-1} \text{ exists} \Rightarrow \\ \quad p_Z(z) = p_X(f^{-1}(z)) \left| \det \dfrac{\partial f^{-1}}{\partial z}(z) \right| \end{array}}$$

).

Therefore, $p_Z(z) = \prod_{i=1}^{n} p_{Z_i}(z_i)$

and $Z_1, \ldots, Z_n$ are independent and gaussian. Moreover, following from $(*)$

$$[\Psi_X]_{ij} = \sum_{k=1}^{n} \sum_{h=1}^{n} [R^{-1}]_{ik} [R^{-1}]_{jh} E\{z_k z_h\}$$

$$= \sum_{k=1}^{n} \sum_{h=1}^{n} [R^{-1}]_{ih} [R^{-1}]_{jh}$$

$$= [R^{-1} R^{-1}]_{ij} = [\Psi]_{ij}$$

In conclusion, $Z_1, \ldots, Z_n$ are gaussian, independent, with zero expectation and unitary variance. It also follows that each $X_i$ is gaussian (X_i and X_j are not independent for $i \neq j$ except when $\Psi_X = I$) ◄

For a gaussian random $X(\omega)$ vector with mean $m_X$ and covariance $\Psi_X$ we will write

$$X \sim \mathcal{N}(m_X, \Psi_X)$$

Since $Z = \Psi_X^{1/2}(X - m_X)$ is also gaussian with mean 0 and covariance I :

$$Z \sim \mathcal{N}(0, I)$$

In this case, Z is said to be <u>standard</u>.

REMARK. If $X_i$ and $X_j$, $i \neq j$, are uncorrelated, i.e. $\sigma_{X_i X_j} = 0$ or which is the same $[\Psi_X]_{i,j} = 0$, then

$$p_X(x) = \frac{1}{(2\pi)^{\frac{n}{2}}} \frac{1}{(\det \Psi_X)^{1/2}} e^{-\frac{1}{2}(x-m)^T \Psi_X^{-1}(x-m)}$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}} \frac{1}{\prod_{r=1}^{n} (\det [\Psi_X]_{i,i})^{1/2}} \prod_{r=1}^{n} e^{-\frac{1}{2}\frac{(x_i-m_i)^2}{[\Psi_X]_{i,i}}}$$

$$= \prod_{r=1}^{n} \frac{1}{\sqrt{2\pi}} \frac{1}{(\det [\Psi_X]_{i,i})^{1/2}} \cdot e^{-\frac{1}{2}\frac{(x_i-m_i)^2}{[\Psi_X]_{i,i}}}$$

$$= \prod_{r=1}^{n} p_{X_i}(x_i)$$

so that $X_i$ and $X_j$, $i \neq j$, are also independent:

INDEPENDENCE $\Leftrightarrow$ UNCORRELATION for gaussian vectors

# 8. INEQUALITIES FOR EXPECTATIONS

- HÖLDER INEQUALITY

$$\boxed{E\{|XY|\} \leq E\{|X|^p\}^{1/p} E\{|Y|^q\}^{1/q}}$$

for all $p, q \geq 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$

REMARKS. if $Y = 1 \Rightarrow$

$$\boxed{E\{|X|^p\} \leq E^p\{|X|\}}, \forall p \geq 1$$

if $p = q = 2 \Rightarrow$

$$\boxed{E\{|XY|\}^2 \leq E\{X^2\} E\{Y^2\}}$$

(CAUCHY-SCHWARTZ INEQUALITY)

- MINKOWSKI INEQUALITY

$$\boxed{E\{|X+Y|^p\}^{1/p} \leq E\{|X|^p\}^{1/p} + E\{|Y|^p\}^{1/p} \quad \forall p \geq 1}$$

- JENSEN INEQUALITY :

If $f$ is convex on $\mathbb{R}$ :

$$\boxed{f(E\{X\}) \leq E\{f(X)\}}$$

- CHEBYSHEV INEQUALITY :

$$\boxed{\mathbb{P}\{\omega \in \Omega ; |X(\omega)| \geq b\} \leq \frac{E\{f(X(\omega))\}}{f(b)}}$$

$\forall\ b > 0$, $\forall\ f$ continuous on $\mathbb{R}$, $f(x) \geq 0$, $f(x) = f(-x)\ \forall\ x \in \mathbb{R}$, $f$ non decreasing on $(0, +\infty)$.

REMARK. If $f(X) = |X|^p$, $p > 0$ :

$$\mathbb{P}\{\omega : |X(\omega)| \geq b\} \leq \frac{E\{|X(\omega)|^p\}}{b^p}.$$

On the other hand, if $p = 2$

$$\mathbb{P}\{\omega : |X(\omega) - E\{X(\omega)\}| \geq b\} \leq \frac{E\{|X(\omega) - E\{X(\omega)\}|^2\}}{b^2}$$

$$= \frac{\sigma_X^2}{b^2}$$

# 9. CONVERGENCE OF RANDOM VARIABLES.

## 9.1 ALMOST SURE CONVERGENCE

A sequence of random variables $\{X_n\}$ CONVERGES ALMOST SURELY or ALMOST EVERYWHERE (A.S.) to a random variable $X$ if

$$\boxed{\lim_{n\to+\infty} X_n(\omega) = X(\omega)}$$

$\forall \omega \in \Omega \setminus N$, where $N \in \mathcal{F}$ is such that $P(N) = 0$.

We write $X_n \xrightarrow{A.S.} X$.

N & S CONDITION:

$X_n \xrightarrow{A.S.} X$ if and only if

$\forall \epsilon > 0$:

$$\lim_{r \to +\infty} \mathcal{P}\{|X_n - X| \leq \epsilon, \forall n \geq r\} = 1$$

or, equivalently,

$$\lim_{r \to +\infty} \mathcal{P}\{\exists n \geq r : |X_n - X| > \epsilon\} = 0.$$

## 9.2. CONVERGENCE IN PROBABILITY

A sequence of random variables $\{X_n\}$ CONVERGES IN PROBABILITY to a random variable $X$ if $\forall \epsilon > 0$

$$\boxed{\lim_{n \to +\infty} \mathcal{P}\{|X_n - X| > \epsilon\} = 0}.$$

We write $X_n \xrightarrow{\mathcal{P}} X$.

FACTS.

$$X_n \xrightarrow{A.S.} X \Rightarrow X_n \xrightarrow{P} X$$

$X_n \xrightarrow{P} X \Rightarrow \exists$ a subsequence $\{X_{n_k}\}$ such that $X_{n_k} \xrightarrow{A.S.} X$

## 9.3. CONVERGENCE IN $p$-MEAN

A sequence of random variables $\{X_n\}$ CONVERGES IN $p$-mean, $p>0$, to a random variable $X$ if

$$\boxed{\lim_{n\to+\infty} E\{|X_n - X|^p\} = 0}$$

We write $X_n \xrightarrow{p} X$.

If $p=2$, we also say that $X_n$ converges to $X$ in QUADRATIC MEAN.

FACTS.

$$X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{\mathcal{P}} X$$

(by Chebyshev inequality)

$$X_n \xrightarrow{\mathcal{P}} X \Rightarrow X_n \xrightarrow{p} X$$

if $|X_n|, |X| \leq Y, \forall n$, with

$E\{|Y|^p\} < +\infty$ .

DOMINATED CONVERGENCE THEOREM.

For a sequence of random variables $\{X_n\}$ such that $X_n \xrightarrow{A.S.} X$ and $|X_n| \leq Y$ $\forall n$, with $\int_\Omega Y(\omega)\, d\mathcal{P}(\omega) < +\infty$ :

$$\boxed{\lim_{n \to +\infty} \int_A X_n(\omega)\, d\mathcal{P}(\omega) = \int_A X(\omega)\, d\mathcal{P}(\omega)}$$

$\forall A \in \mathcal{F}$. In particular, if $A = \Omega$

$$\boxed{\lim_{n \to +\infty} E\{X_n\} = E\{X\}}$$

# 10. THE (HILBERT) SPACE OF RANDOM VARIABLES.

Given $(\Omega, \mathcal{F}, P)$ consider the space of random variables $X(\omega)$. This is a linear space $\mathcal{L}$ on the field of real numbers $\mathbb{R}$. Define as $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ the subset of $\mathcal{L}$ of random variable $X(\omega)$ such that $E\{|X|^p\} < +\infty$. This space is also linear on $\mathbb{R}$ since by Minkowski inequality

$$E\{|a_1 X_1 + a_2 X_2|^p\} \leq \left(|a_1| E\{|X_1|^p\}^{1/p} + |a_2| E\{|X_2|^p\}^{1/p}\right)^p < +\infty$$

In $\mathcal{L}_p(\Omega, \mathcal{F}, P)$ we can consider the norm $\|\cdot\|_p$:

$$\boxed{\|X\|_p \triangleq E\{|X|^p\}^{1/p}}$$

If $p=2$, $\|X\|_2 = \sigma_X$ as long as $E\{X\}=0$. In this case, the norm $\|\cdot\|_2$ is induced by the scalar product

$$\boxed{\langle X, Y\rangle \triangleq E\{XY\}}$$

which in particular gives

$$\boxed{\|X\|_2 \triangleq \sqrt{E\{X^2\}} = \sqrt{\langle X, X\rangle}}$$

By Cauchy-Schwartz inequality

$$\langle X, Y\rangle \leq E\{|XY|\} \leq E\{X^2\}^{1/2} E\{Y^2\}^{1/2}$$

$$= \|X\|_2 \|Y\|_2 .$$

Moreover, we identify $X_1, X_2 \in \mathcal{L}_2(\Omega, \mathcal{F}, P)$ if $X_1(\omega) = X_2(\omega)$ $\forall \omega \in \Omega \setminus N$, $P(N)=0$, since in doing so

$$\|X\|_2 = \langle X, X\rangle = 0 \iff X = 0$$

(which is one of the axioms to be satisfied by a norm).

Moreover, by the dominated convergence theorem, it can be seen that any sequence $\{X_n\}$, $X_n \in \mathcal{L}_2(\Omega, \mathcal{F}, P)$, convergent in the norm $\|\cdot\|_2$, has a limit $X \in \mathcal{L}_2(\Omega, \mathcal{F}, P)$. Convergence in the norm $\|\cdot\|_2$ is convergence in quadratic mean.

Two random variables $X, Y \in \mathcal{L}_2(\Omega, \mathcal{F}, P)$ are ORTHOGONAL if $\langle X, Y \rangle = 0$ (we write $X \perp Y$).

Notice that

$X, Y$ uncorrelated $\Longleftrightarrow$

$$\langle X - E\{X\}, Y - E\{Y\} \rangle = 0$$

The space $\mathcal{L}_2(\Omega, \mathcal{F}, P)$ is infinite-dimensional and can be decomposed as follows. Consider any $\mathcal{M} \subset \mathcal{L}_2(\Omega, \mathcal{F}, P)$ and the set $\mathcal{M}^\perp$ of $X \in \mathcal{L}_2(\Omega, \mathcal{F}, P)$ such that $\langle X, Y \rangle = 0$ for all $Y \in \mathcal{M}$. $\mathcal{M}^\perp$ is the <u>orthogonal space</u> of $\mathcal{M}$. Moreover, any $X \in \mathcal{L}_2(\Omega, \mathcal{F}, P)$ can be uniquely decomposed as

$$\boxed{\begin{array}{l} X = X_1 + X_2, \\ X_1 \in \mathcal{M},\ X_2 \in \mathcal{M}^\perp \end{array}}$$

Notice that $\mathcal{M} \cap \mathcal{M}^\perp = \{0\}$ since $\langle X, X \rangle = 0 \Leftrightarrow X = 0$.

The random variable $X_1$ is the <u>ORTHOGONAL PROJECTION of X</u> on $\mathcal{M}$. We write $X_1 = \Pi(X|\mathcal{M})$.

Notice that

$$\boxed{\Pi(X|\mathcal{M}) : \langle X - \Pi(X|\mathcal{M}), Y \rangle = 0 \ \forall Y \in \mathcal{M}.}$$

In the case that $\mathcal{M}$ has finite dimension $m$, it is possible to find a basis of $\mathcal{M}$

$$\{Y_1, \dots, Y_m\}$$

and

$$\Pi(X|\mathcal{M}) = \sum_{i=1}^{m} c_i Y_i, \quad c_i \in \mathbb{R}.$$

Let find the coefficients $c_i$, $i=1,\dots,m$.

Since it must be

$$< X - \Pi(X|\mathcal{M}); Y_j > = 0$$

if the basis $\{Y_1, \dots, Y_m\}$ is orthonormal, i.e. $<Y_i, Y_j> = \delta_{ij} \triangleq \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases}$,

we obtain

$$<X, Y_j> = \sum_{i=1}^{m} c_i <Y_i, Y_j> = c_j$$

Then

$$\boxed{\Pi(X|\mathcal{M}) = \sum_{i=1}^{m} <X, Y_i> Y_i} \qquad (P1)$$

If the base $\{Y_1, \dots, Y_m\}$ is not orthonormal:

$$[\langle X, Y_1\rangle \cdots\cdots \langle X, Y_m\rangle] =$$

$$[c_1 \ldots c_m] \underbrace{\begin{bmatrix} \langle Y_1, Y_1\rangle & \cdots\cdots & \langle Y_1, Y_m\rangle \\ \vdots & & \\ \langle Y_m, Y_1\rangle & \cdots\cdots & \langle Y_m, Y_m\rangle \end{bmatrix}}_{\triangleq \text{ matrix } S \text{ (it is non-singular!)}}$$

from which we obtain $c_1, \ldots, c_m$:

$$[c_1 \ldots c_m] = [\langle X, Y_1\rangle \ldots \langle X, Y_m\rangle] \cdot S^{-1}$$

so that

$$\boxed{\Pi(X|\mathcal{Y}) = [\langle X, Y_1\rangle \ldots \langle X, Y_m\rangle]\, S^{-1} \begin{bmatrix} Y_1 \\ \vdots \\ Y_m \end{bmatrix}} \quad (P2)$$

For spaces of random vectors $\mathcal{L}_2^n(\Omega, \mathcal{F}, P)$ with $n$ components

$$\boxed{\begin{gathered} \langle X, Y\rangle = E\{X^T Y\} \\ X(\omega), Y(\omega) \in \mathbb{R}^n \end{gathered}}$$

It is possible to define orthogonality also between two random vectors with dimensions $n \neq m$:

$X(\omega) \in \mathbb{R}^n$, $Y(\omega) \in \mathbb{R}^m$

$$\boxed{X \perp Y \iff E\{X_i Y_j\} = 0 \quad \forall\, i = 1, .., n, \; j = 1, .., m.}$$

Moreover, X and Y are uncorrelated if and only if

$$X - E\{X\} \perp Y - E\{Y\}$$

if and only if

$$\Psi_{XY} = 0 \quad .$$

Notice that (P1) and (P2) can be interpreted as orthogonal projections of $X$, $X(\omega) \in \mathbb{R}$, on the space $\mathcal{M}$ of the linear transformations of $Y = (Y_1 \dots Y_m)^T$.

The formula (P2) can be rewritten as:

$$\Pi(X|\mathcal{M}) = E\{XY^T\} \cdot E\{YY^T\}^{-1} Y.$$

If $X = (X_1 \dots X_n)^T$ it is possible to have the orthogonal projection of $X_i$ on $\mathcal{M}$ as

$$\Pi(X_i|\mathcal{M}) = E\{X_i Y^T\} E\{Y Y^T\}^{-1} Y, \; i = 1, \dots, n,$$

and if we put all these together

$$\boxed{\Pi(X|\mathcal{M}) = E\{XY^T\} E\{YY^T\}^{-1} Y},$$

$X(\omega) \in \mathbb{R}^n, \; Y(\omega) \in \mathbb{R}^m.$