3. Orthogonal projection. Conditional expectation in the wide sense

Let $(X_n)_{n\geq 1}$ be a sequence of random variables with

$$\boldsymbol{E}X_n^2 = \sigma_n^2 \quad \text{and} \quad \boldsymbol{E}X_n \equiv 0.$$

If

$$\boldsymbol{E}X_kX_j = \begin{cases} \sigma_k^2, & k=j \\ 0, & \text{otherwise,} \end{cases}$$

$(X_n)_{n\geq 1}$ is the sequence with orthogonal elements. For $\sigma_k^2 \equiv 1$, this sequence forms *white noise*.

If random variables with finite second moments are not orthogonal, they might be orthogonalize by a specific orthogonalization procedure. We start with an example.

Example 1. Let Y and Z be random variables with $\boldsymbol{E}Y = 0$, $\boldsymbol{E}Z = 0$ and $\boldsymbol{E}Y^2 < \infty$, $\boldsymbol{E}Z^2 < \infty$. We introduce a projection of X on a pair $(1, Z)$ by taking a linear combination of 1 and Z: $\widehat{Y} = c_1 + c_2 Z$. If constants c_1, c_2 can be chosen such that $\boldsymbol{E}\{Y - \widehat{Y}\}Z = 0$ the projection $\widehat{Y}$ is named the orthogonal projection. Under $\boldsymbol{E}Z^2 > 0$, a direct verification shows that the choice $c_1 = 0$ and $c_2 = \frac{\boldsymbol{E}YZ}{\boldsymbol{E}Z^2}$ gives the orthogonal projection

$$\widehat{Y} = \frac{\boldsymbol{E}YZ}{\boldsymbol{E}Z^2}Z. \tag{3.1}$$

Consider now the next setting. Let $Y, Z_1, \dots, Z_n$ be random variables with finite second moments: $\boldsymbol{E}Y^2 < \infty$ and $\boldsymbol{E}Z_k^2 < \infty$, $k = 1, ..., n$ and introduce a linear space generated by $1, Z_1, \cdots, Z_n$:

$$\mathcal{M} = \Big\{\eta : \eta\text{=linear combination of } 1, Z_1, \cdots, Z_n\Big\};$$

if $1, Z_1, \cdots, Z_n$ is replaced by $1, Z_1, \cdots, Z_n, \dots$, $\mathcal{M}$ contains not only linear combination of $1, Z_1, \cdots, Z_n, \dots$ but also their limits in L^2-norm.

The orthogonal projection $\widehat{Y}$ of Y on $\mathcal{M}$ is an element of $\mathcal{M}$ such that $\boldsymbol{E}\eta\big(Y - \widehat{Y}\big) = 0$ for any $\eta \in \mathcal{M}$; symbolically

$$Y - \widehat{Y} \perp \mathcal{M}.$$

3.1. Computation of the orthogonal projection.

Set

- $m = \boldsymbol{E}Y$;
- $m_k^z = \boldsymbol{E}Z_k,\ k = 1, \cdots, n$;
- $\mathcal{Z}$ vector(column) with entries $Z_1, \cdots, Z_n$
- m^Z vector(column) with entries $m_1^z, \cdots, m_n^z$
- $\mathsf{var}(Y) = \boldsymbol{E}(Y - m^Y)^2$;
- $\mathsf{cov}(\mathcal{Z}, \mathcal{Z}) = \boldsymbol{E}(\mathcal{Z} - m^{\mathcal{Z}})(\mathcal{Z} - m^{\mathcal{Z}})^T$ (T is the transposition symbol);
- $\mathsf{cov}(Y, \mathcal{Z}) = \boldsymbol{E}(Y - m^y)(\mathcal{Z} - m^{\mathcal{Z}})^T$.

Theorem 3.1. *Under nonsingular matrix* $\mathsf{cov}(\mathcal{Z},\mathcal{Z})$*, the orthogonal projection of* Y *on* $\mathcal{M}$ *exists and is defined as follows:*

$$\widehat{Y} = m + \mathsf{cov}(Y,\mathcal{Z})\mathsf{cov}^{-1}(\mathcal{Z},\mathcal{Z}))(\mathcal{Z} - m^Z). \tag{3.2}$$

Moreover, the variance of the perpendicular $Y - \widehat{Y}$ *is:*

$$\boldsymbol{E}(Y - \widehat{Y})^2 = \mathsf{var}(Y) - \mathsf{cov}(Y,\mathcal{Z})\mathsf{cov}^{-1}(\mathcal{Z},\mathcal{Z})\mathsf{cov}^T(Y,\mathcal{Z}). \tag{3.3}$$

Proof. With an arbitrary vector(raw) C, let us introduce a random variable

$$\xi = \big(Y - m\big) + C\big(\mathcal{Z} - m^Z\big). \tag{3.4}$$

We intend to choose C such that ξ is orthogonal to $\mathcal{M}$. If it is possible, (3.4) provides the decomposition for Y (here $\mathcal{M}^\perp$ is the orthogonal complement to $\mathcal{M}$):

$$Y = \underbrace{m - C\big(\mathcal{Z} - m^Z\big)}_{\in\mathcal{M}} + \underbrace{\xi}_{\in\mathcal{M}^\perp}. \tag{3.5}$$

Hence, if C is found, we have $\widehat{Y} = m - C\big(\mathcal{Z} - m^Z\big)$. Now, owing $\mathcal{Z} - m^Z \in \mathcal{M}$ and $\xi \in \mathcal{M}^\perp$ with $\boldsymbol{E}\big(\mathcal{Z} - m^Z\big) = 0$ and $\boldsymbol{E}\xi = 0$, we find $\boldsymbol{E}\xi\big(\mathcal{Z} - m^Z\big) = 0$. The latter and (3.4) provide

$$\mathsf{cov}(Y,\mathcal{Z}) + C\mathsf{cov}(\mathcal{Z},\mathcal{Z}) = 0 \tag{3.6}$$

and, whereas $\mathbf{cov}(\mathcal{Z},\mathcal{Z})$ is assumed to be nonsingular matrix, the vector $C^\circ = -\mathsf{cov}(Y,\mathcal{Z})\mathsf{cov}^{-1}(\mathcal{Z},\mathcal{Z})$ solves (3.6).

Hence, the first statement holds true.

Notice now that $\xi = Y - \widehat{Y}$ and

$$\begin{aligned}\boldsymbol{E}\xi^2 &= \mathsf{var}(Y) - 2\mathsf{cov}(Y,\mathcal{Z})\mathsf{cov}^{-1}(\mathcal{Z},\mathcal{Z})\mathsf{cov}^T(Y,\mathcal{Z}) \\ &\quad + \mathsf{cov}(Y,\mathcal{Z})\mathsf{cov}^{-1}(\mathcal{Z},\mathcal{Z})\mathsf{cov}^T(Y,\mathcal{Z}) \\ &= \mathsf{var}(Y) - \mathsf{cov}(Y,\mathcal{Z})\mathsf{cov}^{-1}(\mathcal{Z},\mathcal{Z})\mathsf{cov}^T(\alpha,\mathcal{Z}).\end{aligned}$$

□

Remark 3.2. If the matrix $\mathsf{cov}(\mathcal{Z},\mathcal{Z})$ is singular, the statement of the theorem remains valid with $\mathsf{cov}^{-1}(\mathcal{Z},\mathcal{Z})$ replaced by the Moore-Penrose pseudo-inverse matrix $\mathsf{cov}^{+}(\mathcal{Z},\mathcal{Z}) = T^T(TT^T)^{-2}T$, where

$$T^TT = \mathsf{cov}(\mathcal{Z},\mathcal{Z})$$

with T the rectangular matrix of the full rank so that TT^T a quadratic nonsingular matrix.

3.2. **The conditional expectation in the wide sense.**

The orthogonal projection $\widehat{Y}$ of Y on $\mathcal{M}$ is referred as *conditional expectation in the wide sense given* $\mathcal{M}$, $\widehat{Y} = \widehat{\boldsymbol{E}}(Y|\mathcal{M})$, and plays an important role in the optimal in the mean square sense linear estimation.

We establish below the main properties of $\widehat{\boldsymbol{E}}(Y|\mathcal{M})$.

1.
$$\begin{aligned}
&\boldsymbol{E}\widehat{\boldsymbol{E}}(Y|\mathcal{M}) = \boldsymbol{E}Y, \\
&\widehat{\boldsymbol{E}}(Y|\mathcal{M}) = Y, \qquad \text{if } Y \in \mathcal{M} \\
&\widehat{\boldsymbol{E}}(Y|\mathcal{M}) = 0, \qquad \text{if } Y \in \mathcal{M}^{\perp}.
\end{aligned}$$

Proof. The random variable $Y - \widehat{Y}$ is orthogonal to $\mathcal{M}$, particularly, orthogonal to 1, that is $\boldsymbol{E}(Y - \widehat{Y}) = 0$.
If $\alpha \perp \mathcal{M}$, then by the definition $\widehat{\boldsymbol{E}}(\alpha|\mathcal{M}) = 0$.

□

2. If c is a constant, then
$$\widehat{\boldsymbol{E}}(cY|\mathcal{M}) = c\widehat{\boldsymbol{E}}(Z|\mathcal{M}).$$

Proof. It holds true since the operator of orthogonal projection is linear.

□

3. If c_1, c_2 are constants and Y_1, Y_2 are random variables, possessing second moments, then
$$\widehat{\boldsymbol{E}}(c_1Y_1 + c_2Y_2|\mathcal{M}) = c_1\widehat{\boldsymbol{E}}(Y_1|\mathcal{M}) + c_2\widehat{\boldsymbol{E}}(Y_2|\mathcal{M}).$$

Proof. It holds true since the operator of orthogonal projection is linear.

□

4. If $\mathcal{M}'$ ia a linear subspace of $\mathcal{M}$: $\mathcal{M}' \subseteq \mathcal{M}$, then
$$\widehat{\boldsymbol{E}}\Big(\widehat{\boldsymbol{E}}(Y|\mathcal{M})|\mathcal{M}'\Big) = \widehat{\boldsymbol{E}}\big(Y|\mathcal{M}'\big).$$

Proof. Since both $\widehat{\boldsymbol{E}}\Big(\widehat{\boldsymbol{E}}(Y|\mathcal{M})|\mathcal{M}'\Big)$ and $\widehat{\boldsymbol{E}}\big(Y|\mathcal{M}'\big)$ are from $\mathcal{M}'$, it suffices to show that $\phi = \widehat{\boldsymbol{E}}\Big(\widehat{\boldsymbol{E}}(\alpha|\mathcal{M})|\mathcal{M}'\Big) - \widehat{\boldsymbol{E}}\big(\alpha|\mathcal{M}'\big)$ is orthogonal to $\mathcal{M}'$. To this purpose, the decomposition
$$Y = \widehat{\boldsymbol{E}}\big(Y|\mathcal{M}'\big) + \big(Y - \widehat{\boldsymbol{E}}\big(Y|\mathcal{M}'\big)\big)$$
is used. Then, whereas $\big(Y - \widehat{\boldsymbol{E}}\big(Y|\mathcal{M}'\big)\big) \perp \mathcal{M}'$, for any $\psi \in \mathcal{M}'$ we have $\boldsymbol{E}\psi Y = \boldsymbol{E}\psi\widehat{\boldsymbol{E}}\big(Y|\mathcal{M}'\big)$. On the other side, since $\mathcal{M}' \subseteq \mathcal{M}$, similarly we get
$$\boldsymbol{E}\psi\widehat{\boldsymbol{E}}\Big(\widehat{\boldsymbol{E}}\big(Y|\mathcal{M}\big)|\mathcal{M}'\Big) = \boldsymbol{E}\psi\widehat{\boldsymbol{E}}\Big(Y|\mathcal{M}\Big) = \boldsymbol{E}\psi Y.$$
Thus, $\boldsymbol{E}\psi\phi = 0$. □

5. Let $\mathcal{M}'$ and $\mathcal{M}''$ are orthogonal linear spaces: $\mathcal{M}' \perp \mathcal{M}''$. Set $\mathcal{M} = \mathcal{M}' \bigoplus \mathcal{M}''$. Then for Y, with $\boldsymbol{E}Y = 0$,
$$\widehat{\boldsymbol{E}}(Y|\mathcal{M}) = \widehat{\boldsymbol{E}}(Y|\mathcal{M}') + \widehat{\boldsymbol{E}}(Y|\mathcal{M}'').$$

Proof. Notice that $\phi = Y - \widehat{\boldsymbol{E}}(Y|\mathcal{M}') + \widehat{\boldsymbol{E}}(Y|\mathcal{M}'')$ is orthogonal to $\mathcal{M}'$ and $\mathcal{M}''$, that is orthogonal to $\mathcal{M}' \bigoplus \mathcal{M}'' := \mathcal{M}$ and the result holds true.

□

6. For every $\eta \in \mathcal{M}$

$$\boldsymbol{E}(Y-\eta)^2 \geq \boldsymbol{E}(Y-\widehat{Y})^2 = \boldsymbol{E}Y^2 - \boldsymbol{E}\widehat{Y}^2.$$

Proof. Write

$$\begin{aligned}
\boldsymbol{E}(Y-\eta)^2 &= = \boldsymbol{E}\Big(Y-\widehat{Y}-\big(\eta-\widehat{Y}\big)\Big)^2 \\
&= \boldsymbol{E}(Y-\widehat{Y})^2 + \boldsymbol{E}(\eta-\widehat{Y})^2 - 2\boldsymbol{E}(Y-\widehat{Y})(\eta-\widehat{Y}) \\
&= \boldsymbol{E}(Y-\widehat{Y})^2 + \boldsymbol{E}(\eta-\widehat{Y})^2 \\
&\geq \boldsymbol{E}(Y-\widehat{Y})^2.
\end{aligned}$$

□

7. The Cauchy-Schwarz inequality

$$\boldsymbol{E}\big(\widehat{\boldsymbol{E}}(Y|\mathcal{M})\big)^2 \leq \boldsymbol{E}Y^2.$$

Proof. Write

$$0 \leq \boldsymbol{E}\big(Y - \widehat{\boldsymbol{E}}(Y|\mathcal{M})\big)^2 = \boldsymbol{E}Y^2 - \boldsymbol{E}\big(\widehat{\boldsymbol{E}}(Y|\mathcal{M})\big)^2.$$

□

3.3. **Law of large numbers for sequence of orthogonal random variables.**

Let $X_1, X_2, \ldots$ be a sequence of orthogonal random variables with $\boldsymbol{E}X_n \equiv 0$ and $\boldsymbol{E}X_n^2 \equiv \sigma_n^2$. Denote by

$$a_n = \frac{1}{n}\sum_{k=1}^{n} X_k.$$

Proposition 3.3. *Under* $\sum\limits_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty$, *it holds* $\lim\limits_{n\to\infty} \boldsymbol{E}a_n^2 = 0$.

Proof. The result is implied by $\boldsymbol{E}a_n^2 = \frac{1}{n^2}\sum_{k=1}^n \sigma_k^2$ and the Kronecker lemma, which being adapted to the case considered, states:

$$\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty \Longrightarrow \lim_{n\to\infty} \frac{1}{n^2}\sum_{k=1}^{n} \sigma_k^2 = 0. \tag{3.7}$$

We give below a sketch of the proof for (3.7). Set $V_n = \sum\limits_{k=1}^{n} \frac{\sigma_k^2}{k^2}$ and notice that $\sum\limits_{k=1}^{n} \sigma_k^2 = \sum\limits_{k=1}^{n} (V_k - V_{k-1})k^2$. Now, summing by parts we find $\sum_{k=1}^n \sigma_k^2 = V_n n^2 - \sum_{k=1}^n V_{k-1}\big(k^2-(k-1)^2\big)$, so that

$$\frac{1}{n^2}\sum_{k=1}^{n} \sigma_k^2 = \Big(V_n - \frac{1}{n^2}\sum_{k=1}^{n} V_{k-1}\big(k^2-(k-1)^2\big)\Big).$$

The use of a telescopic sum $n^2 = \sum_{k-1}^{n} \big(k^2 - (k-1)^2\big)$ allows to transform the above equality to the following one

$$\frac{1}{n^2}\sum_{k=1}^{n}\sigma_k^2 = \frac{1}{n^2}\sum_{k=1}^{n}\big(V_n - V_{k-1}\big)\big(k^2-(k-1)^2\big)$$

and evaluate the right hand side of that equality. For $n > N$, write

$$\frac{1}{n^2}\sum_{k=1}^{n}\big(V_n - V_{k-1}\big)\big(k^2-(k-1)^3\big) = \frac{1}{n^2}\sum_{k=1}^{N-1}\big(V_n - V_{k-1}\big)\big(k^2-(k-1)^3\big) + \frac{1}{n^2}\sum_{k=N}^{n}\big(V_n - V_{k-1}\big)\big(k^2-(k-1)^3\big)$$

and notice that the first summand is evaluated above by

$$\frac{N^2}{n^2}V_\infty = \frac{N^2}{n^2}\sum_{k=1}^{\infty}\frac{\sigma_k^2}{k^2} \to 0, \ n\to\infty$$

while the second by $V_\infty - V_{N-1} = \sum_{k=N-1}^{\infty}\frac{\sigma_k^2}{k^2} \to 0, \ N\to\infty.$ □

3.4. **The martingale in the wide sense.**

Let $X_1, X_2, ...$ be a sequence of random variables, with $\boldsymbol{E}X_k^2 < \infty$, $k \ge 1$, and $\mathcal{M}$ be a linear space generated by $1, X_1, X_2, \dots$ and $\mathcal{M}_n$ be a linear space generated by $1, X_1, X_2, \dots, X_n$. For a random variable ξ, with $E\xi^2 < \infty$, set $Y_n = \widehat{\boldsymbol{E}}(\xi|\mathcal{M}_n)$, $n \ge 1$. By property 4. of the conditional expectation in the wide sense, for $n > m$ provides

$$\widehat{\boldsymbol{E}}(Y_n|\mathcal{M}_m) = Y_m.$$

The random sequence with this property, created by the real conditional expectation, is named martingale. In our setting, the conditional expectation in the wide sense is used and so a notion of martingale in the wide sense is appropriate.

The aim of this Section is to show that Y_n, $n \to \infty$, converges in L^2-norm to a limit $\widehat{\boldsymbol{E}}(\xi|\mathcal{M})$:

$$\underset{n\to\infty}{\text{l.i.m.}}\, \widehat{\boldsymbol{E}}(\xi|\mathcal{M}_n) = \widehat{\boldsymbol{E}}(\xi|\mathcal{M}). \tag{3.8}$$

Proof of (3.8). Denote $y_1 = Y_1$, $Y_n = Y_n - Y_{n-1}, n \ge 2$ and notice that for $n > m$ the random variable y_n is orthogonal to $\mathcal{M}_m^X$, whereas $\boldsymbol{E}y_n = 0$ and moreover $\widehat{\boldsymbol{E}}(y_n|\mathcal{M}_m) = \widehat{\boldsymbol{E}}(Y_n - Y_{n-1}|\mathcal{M}_m) = Y_m - Y_m = 0$. Hence, $Y_n = \sum\limits_{k=1}^{n} y_k$ is the sum of zero mean orthogonal random variables $x_1, \dots, x_n$ and thanks of that $\boldsymbol{E}Y_n^2 = \sum_{k=1}^{n}\boldsymbol{E}y_k^2$. Then, obviously, $\boldsymbol{E}Y_n^2$ increases in n. On the other hand, since by property 7. for the conditional expectation in the wide sense $\boldsymbol{E}Y_n^2 \le \boldsymbol{E}\xi^2$, we have that for any n

$$\sum_{k=1}^{n} \boldsymbol{E}y_k^2 \le \boldsymbol{E}\xi^2,$$

that is $\sum_{k=1}^{\infty} \boldsymbol{E}y_k^2 \le \boldsymbol{E}\xi^2$. Then, the sequence $Y_n, n \ge 1$ converges in L^2-norm to $Y_\infty = \sum_{k=1}^{\infty} y_k$, since by the Cauchy criteria (Theorem 2.2 for Lect. 2)

$$\boldsymbol{E}(Y_n - Y_m)^2 = \boldsymbol{E}\Big(\sum_{k=m+1}^{n} y_k\Big)^2 = \sum_{k=m+1}^{n} \boldsymbol{E}y_k^2 \to 0,\ n, m \to \infty.$$

3.5. **Markov process in the wide sense.**

Assume

$$\widehat{\boldsymbol{E}}(X_n|\mathcal{M}_{n-1}) = a_n X_{n-1},\ n \ge 2, \tag{3.9}$$

where a_n is a sequence of numbers. Set $\varepsilon_n = X_n - a_n X_{n-1}$ and show that $(\varepsilon_n)_{n\ge 2}$ is a sequence of zero mean orthogonal random variables:

$$\boldsymbol{E}\varepsilon_n \equiv 0$$

$$\boldsymbol{E}\varepsilon_n\varepsilon_m = \begin{cases} 0, & n \ne m, \\ \boldsymbol{E}X_n^2 - a_n^2 \boldsymbol{E}X_{n-1}^2 := b_b^2, & \text{otherwise.} \end{cases} \tag{3.10}$$

In fact, whereas by property 1. of the conditional expectation $\boldsymbol{E}X_n = \boldsymbol{E}\widehat{\boldsymbol{E}}\big(X_n|\mathcal{M}_{n-1}\big)$, the first part from (3.10) is valid. By the definition ε_n is orthogonal to $\mathcal{M}_{n-1}$ and so for $\mathcal{M}_m$ with $m < n$. The last property from (3.10) is provided by $\boldsymbol{E}\big(X_n - a_n X_{n-1}\big)^2 = \boldsymbol{E}X_n^2 - a_n^2 \boldsymbol{E}X_{n-1}^2$.

Thus, the sequence $(X_n)_{n\ge 1}$ is defined by a recursion

$$X_n = a_n X_{n-1} + \varepsilon_n$$

subject by X_1 as the initial value, where $(\varepsilon_n)_{n\ge 2}$ is the sequence of zero mean and orthogonal random variables with $\boldsymbol{E}\varepsilon_n^2 = b_n^2$.

Random sequences of such the type are named *Markov processes in the wide sense.*