THE HILBERT SPACE $L^2$
**Definition:** Let \((\Omega, A, P)\) be a probability space. The set of all random variables \(X: \Omega \rightarrow \mathbb{R}\) satisfying

\[ EX^2 < \infty \]

is denoted as \(L_2\).

**Remark:** \(EX^2 < \infty\) implies that \( E|X| < \infty\) (or equivalently that \(EX \in \mathbb{R}\)), because

\[ |X| \leq X^2 + 1 \Rightarrow E|X| \leq EX^2 + 1. \]

**Proposition:** The set \(L_2\) together with the pointwise scalar multiplication defined for \(X \in L_2\) and \(\lambda \in \mathbb{R}\) by

\[ (\lambda X)(\omega) = \lambda(X(\omega)), \ \omega \in \Omega \]

and the pointwise addition defined for \(X, Y \in L_2\) by

\[ (X+Y)(\omega) = X(\omega) + Y(\omega), \ \omega \in \Omega \]

is a vector space.

**Proof:** (i) The two operations are closed because

\[
X \in L_2, \ \lambda \in \mathbb{R} \Rightarrow EX^2 < \infty \\
\Rightarrow E(\lambda X)^2 = \lambda^2 EX^2 < \infty \\
\Rightarrow \lambda X \in L_2
\]

and

\[
X, Y \in L_2 \Rightarrow EX^2, EY^2 < \infty \\
\Rightarrow E(X+Y)^2 \leq E(2X^2 + 2Y^2) < \infty \\
\Rightarrow X + Y \in L_2.
\]
(ii) The associative, commutative, and distributive properties

\[(X+Y)+Z=X+(Y+Z), \quad (\lambda \mu)X=\lambda(\mu X), \quad X+Y=Y+X,\]

\[\lambda(X+Y)=(\lambda X)+(\lambda Y), \quad (\lambda+\mu)X=(\lambda X)+(\mu X)\]

follow immediately from the pointwise definitions of the two operations. For example, if \(X,Y,Z \in L_2\) then

\[((X+Y)+Z)(\omega)=(X+Y)(\omega)+Z(\omega)\]
\[=X(\omega)+Y(\omega))+Z(\omega)\]
\[=X(\omega)+(Y(\omega)+Z(\omega))\]
\[=X(\omega)+(Y+Z)(\omega)\]
\[=(X+(Y+Z))(\omega), \quad \omega \in \Omega.\]

(iii) The random variable 0 which is identically zero on \(\Omega\) satisfies the property

\[X+0=X \quad \forall X \in L_2\]

of a zero vector.

(iv) For all \(X \in S_2\) there exists an inverse vector \(-X\) defined by

\[(-X)(\omega)=-X(\omega)), \quad \omega \in \Omega,\]

satisfying

\[-X+X=0.\]

(v) \(1X=X\)
Exercise: Show that a function $\langle \cdot, \cdot \rangle : \mathbb{L}_2 \times \mathbb{L}_2 \to \mathbb{R}$ can be defined by

$$\langle X, Y \rangle = \mathbb{E}XY,$$

which satisfies for $X, Y, Z \in \mathbb{L}_2$ and $\lambda \in \mathbb{R}$

$$\langle X+Y, Z \rangle = \langle X, Z \rangle + \langle Y, Z \rangle,$$

$$\langle \lambda X, Y \rangle = \lambda \langle X, Y \rangle,$$

$$\langle X, Y \rangle = \langle Y, X \rangle,$$

$$\langle X, X \rangle \geq 0.$$

Solution: $-\infty \leq \mathbb{E}(-X^2 - Y^2) \leq \mathbb{E}XY \leq \mathbb{E}(X^2 + Y^2) \leq \infty \Rightarrow \mathbb{E}XY \in \mathbb{R},$

$$\langle X+Y, Z \rangle = \mathbb{E}(X+Y)Z = \mathbb{E}XZ + \mathbb{E}YZ = \langle X, Z \rangle + \langle Y, Z \rangle,$$

$$\langle \lambda X, Y \rangle = \mathbb{E}(\lambda X)Y = \lambda \mathbb{E}XY = \lambda \langle X, Y \rangle,$$

$$\langle X, Y \rangle = \mathbb{E}XY = \mathbb{E}YX = \langle Y, X \rangle,$$

$$\langle X, X \rangle = \mathbb{E}XX = \mathbb{E}X^2 \geq 0.$$
The function \( < > \) satisfies all the properties of an inner product except for
\[
<X, X> = 0 \iff X = 0,
\]
because \( EX^2 = 0 \) implies only that \( P(X=0) = 1 \), but not that 
\( X(\omega) = 0 \) for all \( \omega \in \Omega \). Analogously, the function \( \| \| \) satisfies all the properties of a norm except for
\[
\|X\| = 0 \iff X = 0.
\]
To circumvent this problem we identify two random variables if they are equal almost surely, i.e., we switch from the individual random variables \( X \in L_2 \) to equivalence classes
\[
[X] = \{ Y \in L_2 : P(Y = X) = 1 \}
\]
of random variables which agree almost everywhere.

**Definition:** Defining for equivalence classes \([X], [Y]\) of almost surely equal elements of \( L_2 \) and \( \lambda \in \mathbb{R} \)
\[
[X] + [Y] = [X + Y], \ \lambda[X] = [\lambda X], \ <[X], [Y]> = <X, Y>
\]
we obtain an inner product space, which is denoted by \( L^2 \).
**Proposition:** The inner product space $L^2$ of equivalence classes of almost surely equal random variables with finite variances is complete, i.e.,

$$X_n \in L^2 \text{ for all } n, \quad \|X_m - X_n\| \to 0 \Rightarrow \exists X \in L^2 : \|X_n - X\| \to 0.$$ 

Thus $L^2$ is a Hilbert space.

**Remark:** Norm convergence

$$\|X_n - X\| \to 0$$

is equivalent to mean square convergence

$$\|X_n - X\|^2 = \text{E}(X_n - X)^2 \to 0.$$ 

**Exercise:** Show that the relation ~ defined by

$$X \sim Y \iff \text{P}(X = Y) = 1$$

is indeed an equivalence relation by verifying the reflexive, symmetric, and transitive properties

$$X \sim X, \quad X \sim Y \Rightarrow Y \sim X, \quad X \sim Y, Y \sim Z \Rightarrow X \sim Z \quad \forall X, Y, Z \in L_2.$$
Solution: The transitive property is satisfied, because

\[
\{ \omega: X(\omega) = Z(\omega) \} \supseteq \{ \omega: X(\omega) = Y(\omega) = Z(\omega) \}
\]

\[
\Rightarrow \{ \omega: X(\omega) = Z(\omega) \}^C \subseteq \{ \omega: X(\omega) = Y(\omega) = Z(\omega) \}^C
\]

\[
= (\{ \omega: X(\omega) = Y(\omega) \} \cap \{ \omega: Y(\omega) = Z(\omega) \})^C
\]

\[
= \{ \omega: X(\omega) = Y(\omega) \}^C \cup \{ \omega: Y(\omega) = Z(\omega) \}^C
\]

\[
\Rightarrow P(\{ \omega: X(\omega) = Z(\omega) \}^C) \leq P(\{ \omega: X(\omega) = Y(\omega) \}^C)
\]

\[
+ P(\{ \omega: Y(\omega) = Z(\omega) \}^C).
\]

Proposition: If \( E(X_n - X)^2 \rightarrow 0 \) and \( E(Y_n - Y)^2 \rightarrow 0 \), then

(i) \( EX_n \rightarrow EX \),

(ii) \( EX_n Y_n \rightarrow EXY \),

(iii) \( \text{Cov}(X_n, Y_n) \rightarrow \text{Cov}(X, Y) \),

(iv) \( \text{Var}(X_n) \rightarrow \text{Var}(X) \).

Proof:

(i) \( EX_n = EX_n \cdot 1 = \langle X_n, 1 \rangle \rightarrow \langle X, 1 \rangle = EX \cdot 1 = EX \)

(ii) \( EX_n Y_n = \langle X_n, Y_n \rangle \rightarrow \langle X, Y \rangle = EXY \)

(iii) \( \text{Cov}(X_n, Y_n) = EX_n Y_n - EX_n EY_n \rightarrow EXY - EXEY = \text{Cov}(X, Y) \)

(iv) \( \text{Var}(X_n) = \text{Cov}(X_n, X_n) \rightarrow \text{Cov}(X, X) = \text{Var}(X) \)
**Definition:** The conditional expectation of $X \in L^2$ given a closed subspace $S \subseteq L^2$, which contains the constant function 1, is defined to be the projection of $X$ onto $S$, i.e.,

$$E(X|S) = P_S(X).$$

**Remark:** The conditional expectation satisfies

$$\|X - E(X|S)\|^2 < \|X - Y\|^2$$

for all other elements of $S$.

**Definition:** The conditional expectation of $X \in L^2$ given $X_1, \ldots, X_n \in L^2$ is defined to be the projection of $X$ onto the closed subspace $M(X_1, \ldots, X_n)$ spanned by all random variables of the form $g(X_1, \ldots, X_n)$, where $g$ is some measurable function $g: \mathbb{R}^n \to \mathbb{R}$, i.e.,

$$E(X|X_1, \ldots, X_n) = P_{M(X_1, \ldots, X_n)}(X).$$

**Remarks:** (i) It follows from

$$\text{span}(1, X_1, \ldots, X_n) \subseteq M(X_1, \ldots, X_n)$$

that

$$\|X - E(X|X_1, \ldots, X_n)\|^2 \leq \|X - E(X|\text{span}(1, X_1, \ldots, X_n))\|^2.$$

(ii) For elements of $L^2$ the definition of $E(X|X_1, \ldots, X_n)$ above coincides with the more general definition of conditional expectation as the mean of the conditional distribution.
**Exercise:** Show that the bivariate normal density

\[ f(x) = f(x_1, x_2) = \frac{1}{\sqrt{(2\pi)^2 \det \Sigma}} \exp\left(-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu)\right) \]

with mean vector \( \mu = (\mu_1, \mu_2)^T \) and covariance matrix

\[ \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix} \]

factors into two univariate normal densities, the marginal density \( f_1 \) with mean \( \mu_1 \) and variance \( \sigma_1^2 \) and the conditional density \( f_{2|1} \) with mean \( \mu_2 + \rho \sigma_2 \frac{x_1 - \mu_1}{\sigma_1} \) and variance \( (1-\rho^2) \sigma_2^2 \).

**Solution:** Putting \( z_1 = \frac{x_1 - \mu_1}{\sigma_1}, z_2 = \frac{x_2 - \mu_2}{\sigma_2} \) and completing squares we obtain

\[ (x-\mu)^T \Sigma^{-1} (x-\mu) = \frac{\begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}^T \begin{pmatrix} \sigma_1^2 & -\rho \sigma_1 \sigma_2 \\ -\rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}}{\sigma_1^2 \sigma_2^2 (1-\rho^2)} \]

\[ = \frac{\sigma_1^2 (x_1 - \mu_1)^2 - 2\rho \sigma_1 \sigma_2 (x_1 - \mu_1)(x_2 - \mu_2) + \sigma_2^2 (x_2 - \mu_2)^2}{\sigma_1^2 \sigma_2^2 (1-\rho^2)} \]

\[ = \frac{z_1^2 - 2\rho z_1 z_2 + z_2^2}{1-\rho^2} = \frac{z_1^2 - \rho^2 z_1^2}{1-\rho^2} + \frac{\rho^2 z_1^2 - 2\rho z_1 z_2 + z_2^2}{1-\rho^2} = z_1^2 + \frac{(z_2 - \rho z_1)^2}{1-\rho^2}. \]

Thus,

\[ f(x_1, x_2) = \frac{1}{\sqrt{2\pi \sigma_1^2}} \exp\left(-\frac{1}{2} z_1^2\right) \frac{1}{\sqrt{2\pi(1-\rho^2) \sigma_2^2}} \exp\left(-\frac{1}{2} \frac{(z_2 - \rho z_1)^2}{1-\rho^2}\right). \]
Remark: The last exercise shows that in the case of a bivariate normal random vector \((X_1, X_2)\) the mean of the conditional distribution of \(X_2\) given \(X_1\) is a linear function of 1 and \(X_1\).

More generally, if \((X, X_1, \ldots, X_n)^T\) has a multivariate normal distribution, then

\[
E(X|X_1,\ldots,X_n) = E(X|\text{span}(1, X_1,\ldots,X_n)).
\]