5.2. $\mathcal{L}^2$ **as a Hilbert space.** We shall apply some general Hilbert space arguments to L^2. First, we note *Pythagoras' rule*

$$\|f+g\|_2^2 = \|f\|_2^2 + 2\langle f, g\rangle + \|g\|_2^2$$

and the *parallelogram law*

$$\|f+g\|_2^2 + \|f-g\|_2^2 = 2(\|f\|_2^2 + \|g\|_2^2).$$

If $\langle f, g\rangle = 0$, then we say that f and g are *orthogonal*. For any subset $V \subseteq L^2$, we define

$$V^\perp = \{f \in L^2 : \langle f, v\rangle = 0 \text{ for all } v \in V\}.$$

A subset $V \subseteq L^2$ is *closed* if, for every sequence $(f_n : n \in \mathbb{N})$ in V, with $f_n \to f$ in L^2, we have $f = v$ a.e., for some $v \in V$.

Theorem 5.2.1 (Orthogonal projection)**.** *Let V be a closed subspace of L^2. Then each $f \in L^2$ has a decomposition $f = v + u$, with $v \in V$ and $u \in V^\perp$. Moreover, $\|f - v\|_2 \leq \|f - g\|_2$ for all $g \in V$, with equality only if $g = v$ a.e..*

The function v is called (*a version of*) the *orthogonal projection of f on V*.

Proof. Choose a sequence $g_n \in V$ such that

$$\|f - g_n\|_2 \to d(f, V) = \inf\{\|f - g\|_2 : g \in V\}.$$

By the parallelogram law,

$$\|2(f - (g_n + g_m)/2)\|_2^2 + \|g_n - g_m\|_2^2 = 2(\|f - g_n\|_2^2 + \|f - g_m\|_2^2).$$

But $\|2(f-(g_n+g_m)/2)\|_2^2 \geq 4d(f,V)^2$, so we must have $\|g_n - g_m\|_2 \to 0$ as $n, m \to \infty$. By completeness, $\|g_n - g\|_2 \to 0$, for some $g \in L^2$. By closure, $g = v$ a.e., for some $v \in V$. Hence

$$\|f - v\|_2 = \lim_n \|f - g_n\|_2 = d(f, V).$$

Now, for any $h \in V$ and $t \in \mathbb{R}$, we have

$$d(f,V)^2 \leq \|f - (v + th)\|_2^2 = d(f,V)^2 - 2t\langle f - v, h\rangle + t^2\|h\|_2^2.$$

So we must have $\langle f - v, h\rangle = 0$. Hence $u = f - v \in V^\perp$, as required. $\square$

5.3. **Variance, covariance and conditional expectation.** In this section we look at some L^2 notions relevant to probability. For $X, Y \in L^2(\mathbb{P})$, with means $m_X = \mathbb{E}(X), m_Y = \mathbb{E}(Y)$, we define *variance*, *covariance* and *correlation* by

$$\begin{aligned}\operatorname{var}(X) &= \mathbb{E}[(X - m_X)^2],\\ \operatorname{cov}(X,Y) &= \mathbb{E}[(X - m_X)(Y - m_Y)],\\ \operatorname{corr}(X,Y) &= \operatorname{cov}(X,Y)/\sqrt{\operatorname{var}(X)\operatorname{var}(Y)}.\end{aligned}$$

Note that $\operatorname{var}(X) = 0$ if and only if $X = m_X$ a.s.. Note also that, if X and Y are independent, then $\operatorname{cov}(X,Y) = 0$. The converse is generally false. For a random variable $X = (X_1, \dots, X_n)$ in $\mathbb{R}^n$, we define its *covariance matrix*

$$\operatorname{var}(X) = (\operatorname{cov}(X_i, X_j))_{i,j=1}^n.$$

Proposition 5.3.1. *Every covariance matrix is non-negative definite.*

Suppose now we are given a countable family of disjoint events $(G_i : i \in I)$, whose union is Ω. Set $\mathcal{G} = \sigma(G_i : i \in I)$. Let X be an integrable random variable. The *conditional expectation* of X given $\mathcal{G}$ is given by

$$Y = \sum_i \mathbb{E}(X|G_i)1_{G_i},$$

where we set $\mathbb{E}(X|G_i) = \mathbb{E}(X1_{G_i})/\mathbb{P}(G_i)$ when $\mathbb{P}(G_i) > 0$, and define $\mathbb{E}(X|G_i)$ in some arbitrary way when $\mathbb{P}(G_i) = 0$. Set $V = L^2(\mathcal{G}, \mathbb{P})$ and note that $Y \in V$. Then V is a subspace of $L^2(\mathcal{F}, \mathbb{P})$, and V is complete and therefore closed.

Proposition 5.3.2. *If $X \in L^2$, then Y is a version of the orthogonal projection of X on V.*

6. Convergence in $L^1(\mathbb{P})$

6.1. **Bounded convergence.** We begin with a basic, but easy to use, condition for convergence in $L^1(\mathbb{P})$.

Theorem 6.1.1 (Bounded convergence)**.** *Let $(X_n : n \in \mathbb{N})$ be a sequence of random variables, with $X_n \to X$ in probability and $|X_n| \le C$ for all n, for some constant $C < \infty$. Then $X_n \to X$ in L^1.*

Proof. By Theorem 2.6.1, X is the almost sure limit of a subsequence, so $|X| \le C$ a.s.. For $\varepsilon > 0$, there exists N such that $n \ge N$ implies

$$\mathbb{P}(|X_n - X| > \varepsilon/2) \le \varepsilon/(4C).$$

Then

$$\mathbb{E}|X_n - X| = \mathbb{E}(|X_n - X|1_{|X_n - X| > \varepsilon/2}) + \mathbb{E}(|X_n - X|1_{|X_n - X| \le \varepsilon/2}) \le 2C(\varepsilon/4C) + \varepsilon/2 = \varepsilon.$$

□