Resumé of the lectures

N.J. Nielsen

May 21, 2014

Introduction

These notes contain the theorems thast were proved in a different way than the corresponding theorems in the book or were proved in more details in an essential way.

1 Conditional expectations

Let us start with the following two results on Hilbert spaces:

Theorem 1.1 *Let $M \subseteq H$ be a closed subspace. Then there exists an orthogonal projection P with $P(H) = M$.*

Proof: The projection theorem gives that $H = M \oplus M^\perp$. Hence if $z \in H$, we can in a unique way write $z = x + y$ with $x \in M$ and $y \in M^\perp$ (the uniqueness follows from Theorem 4.4 (ii) of the notes). If we put $Pz = x$, P is well defined and the uniqueness of the decomposition of a vector as a sum of something from M and $M^\perp$ also gives that P is linear. It is now easy to see that P is an orthogonal projection with $P(H) = M$ and $P^{-1}(0) = M^\perp$. □

Theorem 1.2 *A linear projection $P : H \to H$ is an orthogonal projection if and only if $(Px, y) = (x, Py)$ for all $x, y \in H$.*

Proof: Let us first assume that P is an orthogonal projection. If $x, y \in H$, then we get

$$(Px, y) = (Px, Py + (y - Py)) = (Px, Py)$$

and

$$(x, Py) = (Px + (x - Px), Py) = (Px, Py).$$

Assume next that $(Px, y) = (x, Py)$ for all $x, y \in H$. If $x \in P(H)$ and $y \in P^{-1}(0)$ are arbitrary, $Px = x$ and we therefore get that

$$(x, y) = (Px, y) = (x, Py) = (x, 0) = 0,$$

so that x and y are orthogonal $\square$

In the sequal we let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathcal{G}$ be a sub–σ–algebra of $\mathcal{A}$. We have the following definition:

Definition 1.3 *Let $Y \in L_1(P)$. A stochastic variable $Z \in L_1(\Omega, \mathcal{G}, P)$ is called a conditional expectation of Y given $\mathcal{G}$ if*

$$\int_A Y dP = \int_A Z DP \quad \textit{for all } A \in \mathcal{G} \tag{1.1}$$

We shall write $Z = E(Y \mid \mathcal{G})$.

We shall later see that $E(Y \mid \mathcal{G})$ exists for all $Y \in L_1(P)$. Her we note that

Proposition 1.4 *Let $Y \in L_1(P)$. If $Z_1, Z_2 \in L_1(\Omega, \mathcal{G}, P)$ both satisfy (1.1), then $Z_1 = Z_2$ a.s. In other words, if $E(Y \mid \mathcal{G})$ exists, it is uniquely determined a.s.*

Proof: Since Z_1 and Z_2 are $\mathcal{G}$–measurable, the set $A = (Z_1 > Z_2)$ is $\mathcal{G}$–measurable and therefore it follows from (1.1) that.

$$\int_A (Z_1 - Z_2) dP = 0$$

and since the integrand is stricly possitive on A, this can only happen if $P(A) = 0$. In a similar manner we can get that $P(Z_2 > Z_1) = 0$ so that $Z_1 = Z_2$ a.s $\square$

In the following we consider the the real Hilbert spaces $L_2(\Omega, \mathcal{F}, P)$ and $L_2(\Omega, \mathcal{G}, P)$. W recall that the inner product in $L_2(\Omega, \mathcal{F}, P)$ is given by

$$(f, g) = \int_\Omega f g dP \quad \text{for all } f, g \in L_2(\Omega, \mathcal{F}, P).$$

We also observe that

$$L_2(\Omega, \mathcal{G}, P) = \{f \in L_2(\Omega, \mathcal{F}, P) \mid f \text{ is } \mathcal{G}\text{–measurable}\}.$$

We shall need the following lemma:

Lemma 1.5 *$L_2(\Omega, \mathcal{G}, P)$ is a closed subspace of $L_2(\Omega, \mathcal{F}, P)$.*

Proof: It is clear from the above that $L_2(\Omega, \mathcal{G}, P)$ is a subspace so we only need to prove that it is closed. Hence let $f \in L_2(\Omega, \mathcal{F}, P)$ and let $(f_n) \subseteq L_2(\Omega, \mathcal{G}, P)$ with $f_n \to f$ in $L_2(\Omega, \mathcal{F}, P)$. Fom measure theory it follows that there is a subsequence (f_{n_k}) so that $f_{n_k} \to f$ a.e. Since all the f_{n_k}'s are $\mathcal{G}$–measurable, it follows that also f is $\mathcal{G}$–measurable and hence $f \in L_2(\Omega, \mathcal{G}, P)$ $\square$

We are now able to prove the existence of conditional expectations and we start with the L_2–case.

Theorem 1.6 *If $Y \in L_2(\Omega, \mathcal{F}, P)$, then the condtional expectation $E(Y \mid \mathcal{G})$ exists. Further $E(Y \mid \mathcal{G}) \in L_2(\Omega, \mathcal{G}, P)$ and*

$$\int_\Omega YXdP = \int_\Omega E(Y \mid \mathcal{G})XdP$$

for all $X \in L_2(\Omega, \mathcal{G}, P)$.

Proof: Let $\mathcal{P}$ be the orthogonal projection of $L_2(\Omega, \mathcal{F}, P)$ onto $L_2(\Omega, \mathcal{G}, P)$ and put sæt $Z = \mathcal{P}(Y)$. If $A \in \mathcal{G}$, then $1_A \in L_2(\Omega, \mathcal{G}, P)$ and hence $\mathcal{P}(1_A) = 1_A$. Comparing this with Theorem 1.2 we get:

$$\int_A ZdP = (\mathcal{P}(Y), 1_A) = (Y, \mathcal{P}(1_A)) = (Y, 1_A) = \int_A YdP,$$

which shows that Z is a version of the conditional expectation $E(Y \mid \mathcal{G})$. If $X \in L_2(\Omega, \mathcal{G}, P)$, then $\mathcal{P}(X) = X$ and again it follows from Theorem 1.2 that

$$\int_\Omega ZXdP = (\mathcal{P}(Y), X) = (Y, \mathcal{P}(X)) = (Y, X) = \int_\Omega YXdP.$$

□

Theorem 1.7 *If $Y \geq 0$ is an stochastic variable, then $E(Y \mid \mathcal{G})$ exists as a $\mathcal{G}$–measurable stochastic variable with values in $[0, \infty]$.*

If $Y \in L_1(\Omega, \mathcal{F}, P)$, then $E(Y \mid \mathcal{G})$ exists as an element of $L_1(\Omega, \mathcal{G}, P)$ and if $X \in L_\infty(\Omega, \mathcal{G}, P)$, then

$$\int_\Omega E(Y \mid \mathcal{G})XdP = \int_\Omega YXdP. \tag{1.2}$$

Proof: Let $Y \geq 0$ be $\mathcal{F}$–measurable. There exists an increasing sequence (Y_n) of simple functions with $Y_n \geq 0$ for all $n \in \mathbb{N}$ so that $Y_n \uparrow Y$. Da $Y_n \in L_2(\Omega, \mathcal{F}, P)$ for alle n, $Z_n = E(Y_n \mid \mathcal{G})$ exists for all n and since $Y_{n+1} - Y_n \geq 0$ for all n, it easily follows that $Z_{n+1} - Z_n \geq 0$ a.s. for all n. Hence $Z_n \leq Z_{n+1}$ n.s. We can therefore define

$$Z(\omega) = \lim_n Z_n(\omega) \quad \text{for almost all } \omega \in \Omega.$$

If $A \in \mathcal{G}$ is arbitrary, the monotone convergence theorem gives

$$\int_A ZdP = \lim_n \int_A Z_n dP = \int_A Y_n dP = \int_A YdP,$$

which shows that Z is a version of$E(Y \mid \mathcal{G})$.

Hvis $Y \in L_1(\Omega, \mathcal{F}, P)$ is arbitrary, we write $Y = Y^+ - Y^-$ and observe that since $|Y| = Y^+ + Y^-$, we have

$$\int_\Omega (E(Y^+ \mid \mathcal{G}) + E(Y^- \mid \mathcal{G})dP = \int_\Omega |Y|dP < \infty$$

which implies that $E(Y^+ \mid \mathcal{G}) < \infty$ a.s. and $E(Y^- \mid \mathcal{G}), \infty$ a.s. It now easily follows that if we put $Z = E(Y^+ \mid \mathcal{G}) - E(Y^- \mid \mathcal{G})$, then Z is a version of the conditional expectation $E(Y \mid \mathcal{G})$.

Let finally $Y \in L_1(\Omega, \mathcal{F}, P)$ and $X \in L_\infty(\Omega, \mathcal{G}, P)$. It is clearly enough to prove (1.2) when $Y \geq 0$ so let us assume that. As above we choose a sequence (Y_n) f simple functions so that $Y_n \uparrow Y$ and conclude as before that $E(Y_n \mid \mathcal{G}) \uparrow E(Y \mid \mathcal{G})$ a.s. Since $Y_n \in L_2(\Omega, \mathcal{F}, P)$ for all $n \in \mathbb{N}$ and $X \in L_2(\Omega, \mathcal{G}, P)$, we get from Theorem 1.6 that

$$\int_\Omega Y_n X dP = \int_\Omega E(Y_n \mid \mathcal{G}) X dP \quad \text{for all } n \in \mathbb{N},$$

and we also have the inequalities:

$$|E(Y_n \mid \mathcal{G})X| \leq E(Y \mid \mathcal{G})|X| \in L_1(\Omega, \mathcal{G}, P)$$

and

$$|Y_n X| \leq Y|X|$$

so that applying the majorized convergence theorem twice we get

$$\begin{aligned} \int_\Omega E(Y \mid \mathcal{G}) X dP &= \lim_n \int_\Omega E(Y_n \mid \mathcal{G}) X dP = \\ \lim_n \int_\Omega Y_n X dP &= \int_\Omega Y X dP. \end{aligned}$$

This proves equation (1.2) □

We now wish to prove some comvergence theorems for conditional expectatations similar to those for usual expectations. We start with

Theorem 1.8 *(***Monotone convergence***) Let (X_n) be a sequence of stochastic variables so that $0 \leq X_n \leq X_{n+1}$ a.s. for all $n \in \mathbb{N}$ and put $X = \lim_n X_n$. Then $E(X \mid \mathcal{G}) = \lim_n E(X_n \mid \mathcal{G})$.*

Proof: You will note that the proof is implicitely given in the beginning of the previous theorem.

Since $0 \leq X_n \leq X_{n+1}$ a.s., we get that $0 \leq E(X_n \mid \mathcal{G}) \leq E(X_{n+1} \mid \mathcal{G})$ a.s. for all $n \in \mathbb{N}$ so let $Z = \lim_n E(X_n \mid \mathcal{G})$ which is clearly $\mathcal{G}$–measurable. To finish the proof we have to show that $Z = E(X \mid \mathcal{G})$ a.s. Hence let $A \in \mathcal{G}$ be arbitrary. By the monotone convergence theorem for integrals and the definition of conditional expectations we get that

$$\begin{aligned} \int_A X dP &= \lim_n \int_A X_n dP = \\ \lim_n \int_A E(X_n \mid \mathcal{G}) dP &= \int_A Z dP \end{aligned}$$

which shows that $Z = E(X \mid \mathcal{G})$ a.s. □

The next result corresponds to Fatous Lemma.

Theorem 1.9 *(**Fatous Lemma**) Let (X_n) be a sequence of s.v's so that $X_n \geq 0$ a.s. Then*

$$E(\liminf_n X_n \mid \mathcal{G}) \leq \liminf_n E(X_n \mid \mathcal{G}).$$

Proof: For every $n \in \mathbb{N}$ we put

$$Y_n = \inf\{X_m \mid n \leq m\}$$

and note that $0 \leq Y_n \leq Y_{n+1}$ and $Y_n \leq X_n$ for all $n \in \mathbb{N}$. By definition of $\liminf$ we get that $\liminf_n X_n = \lim_n Y_n$ a.s. and therefore by the Monotone Convergence Theorem we get

$$\begin{aligned} E(\liminf_n X_n \mid \mathcal{G}) &= \lim E(Y_n \mid \mathcal{G}) \leq \\ \liminf_n E(X_n \mid \mathcal{G}) \end{aligned}$$

which gives the result. □

We shall need the following inequality which is actually a special case of Jensen's inequality below.

Lemma 1.10 *If $X \in L_1(P)$, then $|E(X \mid \mathcal{G})| \leq E(|X| \mid \mathcal{G})$ a.s.*

Proof: Since $X \leq |X|$ and $-X \leq |X|$ a.s. we get that $E(X \mid \mathcal{G}) \leq E(|X| \mid \mathcal{G})$ and $-E(X \mid \mathcal{G}) \leq E(|X| \mid \mathcal{G})$ a.s. Hence $|E(X \mid \mathcal{G})| \leq E(|X| \mid \mathcal{G})$ a.s. □

The next theorem corresponds to the Dominated Convergence Theorem.

Theorem 1.11 *(**The dominated Convergence Theorem**) Let $(X_n) \subseteq L_1(P)$ and $Y \in L_1(P)$ so that $|X_n| \leq Y$ a.s. If X is an s.v. so that $X_n \to X$ a.s., then $E(X_n \mid \mathcal{G}) \to E(X \mid \mathcal{G})$ a.s. and in $L_1(P)$.*

Proof: We first note that the usual dominated convergence theorem for integrals gives that $X \in L_1(P)$ and that $X_n \to X$ in $L_1(P)$. Noting that the triangle inequality gives that
$0 \leq 2Y - |X - X_n|$ a.s and that

$$\liminf(2Y - |X - X_n|) = \lim(2Y - |X - X_n|) = 2Y \quad \text{a.s.},$$

an application of Fatou's Lemma gives

$$\begin{aligned} E(2Y \mid \mathcal{G})) &\leq \liminf E(2Y - |X - X_n| \mid \mathcal{G}) = \\ E(2Y \mid \mathcal{G}) + \liminf(-E(|X - X_n| \mid \mathcal{G})) &= E(2Y \mid \mathcal{G}) - \limsup E(|X - X_n| \mid \mathcal{G}). \end{aligned}$$

Deducting $2E(Y \mid \mathcal{G})$ on both sides we get that $\limsup E(|X - X_n| \mid \mathcal{G}) \leq 0$, but then
$0 \leq \liminf E(|X - X_n| \mid \mathcal{G}) \leq \limsup E(|X - X_n| \mid \mathcal{G}) \leq 0$ and hence

$$|E(X - X_n) \mid \mathcal{G})| \leq E(|X - X_n| \mid \mathcal{G}) \to 0 \quad \text{a.s.}.$$

Since $\Omega \in \mathcal{G}$ we get from the above

$$\begin{aligned}\int_\Omega |E(X \mid \mathcal{G}) - E(X_n \mid \mathcal{G})| dP &\leq \int_\Omega E(|X - X_n| \mid \mathcal{G}) dP = \\ \int_\Omega |X - X_n| dP &\to 0,\end{aligned}$$

which shows that $E(X_n \mid \mathcal{G}) \to E(X \mid \mathcal{G})$ in $L_1(P)$. $\square$

We recall that a function $\phi : \mathbb{R} \to \mathbb{R}$ is called convex if $\phi((1-t)x+ty) \leq (1-t)\phi(x)+t\phi(y)$ for all $x, y \in \mathbb{R}$ and all $t \in [0,1]$. Geometrically this means that that the point $((1-t)x+ty, \phi((1-t)x+ty)) \in \mathbb{R}^2$ lies below the line segment between the points $(x, \phi(x))$ and $(y, \phi(y))$; in other words, the set $\{(x,z) \in \mathbb{R}^2 \mid \phi(x) \leq z\}$ is a convex subset of the plane. It follows easily from this description that if $u, v, w \in \mathbb{R}$ with $u < v < w$, then

$$\frac{\phi(v) - \phi(u)}{v - u} \leq \frac{\phi(w) - \phi(v)}{w - v} \tag{1.3}$$

and hence the left hand term increases with u. We put

$$D_-(v) = \lim_{u \uparrow v} \frac{\phi(v) - \phi(u)}{v - u} \leq \frac{\phi(w) - \phi(v)}{w - v}. \tag{1.4}$$

In particular the limit $D_-(v)$ is finite and since $v - u \to 0$ for $u \uparrow v$, this implies that $\phi(v) - \phi(u) \to 0$. Hence ϕ is continuous from the left. A similar argument letting $w \downarrow v$ shows that ϕ is continuous from the right. A geometrical proof of the continuity of ϕ can be found in W. Rudin, Real and complex analysis. Since it is enough to take the limit in (1.4) along a sequence (u_n) with $u_n \uparrow v$, we see that D_- is the pointwise limit of a sequence of continuous functions so that D_- is Borel measurable We can now prove Jensen's inequality:

Theorem 1.12 *Let $\phi : \mathbb{R} \to \mathbb{R}$ be a convex function and let $X \in L_1(P)$. If $\phi(X) \in L_1(P)$, then*

$$\phi(E(X \mid \mathcal{G})) \leq E(\phi(X) \mid \mathcal{G})$$

Proof: The inequality

$$\phi(x) - \phi(v) \geq D_-(v)(x - v) \quad \text{for alle } x, v \in \mathbb{R} \tag{1.5}$$

can be seen as follows: If $v < x$, then we put $w = x$ in (1.4) and if $x < v$, we put $u = x$ in (1.4). If we put

$$\beta = D_-(E(X \mid \mathcal{G})),$$

then β is$\mathcal{G}$–measurable and from (1.5) we get

$$\phi(X) - \phi(E(X \mid \mathcal{G}) \geq \beta(X - E(X \mid \mathcal{G})). \tag{1.6}$$

If β is bounded and $\phi(E(X \mid \mathcal{G})) \in L_1(P)$, we can take the conditional expectation in (1.6) and get:

$$E(\phi(X) \mid \mathcal{G}) - \phi(E(X \mid \mathcal{G})) \geq \beta(E(X \mid \mathcal{G}) - E(X \mid \mathcal{G})) = 0, \tag{1.7}$$

which gives the result in this special case. However, since β need not be bounded and $\phi(E(X \mid \mathcal{G}))$ need not be in $L_1(P)$, we must continue. For every $n \in \mathbb{N}$ we put vi

$$D_n = \{\omega \in \Omega \mid |E(X \mid \mathcal{G})| \leq n\},$$

and note that $D_n \in \mathcal{G}$. Since ϕ is continuous, it is bounded on the interval $[-n, n]$ and since $D_{-\phi}$ is non–decreasing, it is also bounded on the interval $[-n, n]$. Hence both $1_{D_n}\beta$ and $1_{D_n}\phi(E(X \mid \mathcal{G}))$ are bounded. If we multiply (1.6) with 1_{D_n} and take conditional expectation, we get:

$$1_{D_n}(E(\phi(X) \mid \mathcal{G}) - \phi(E(X \mid \mathcal{G}))) \geq 1_{D_n}\beta(E(X \mid \mathcal{G}) - E(X \mid \mathcal{G})) = 0. \tag{1.8}$$

We observe that $D_n \uparrow \Omega$ for $n \to \infty$ and hence if we let $n \to \infty$ i (1.8) we get:

$$E(\phi(X) \mid \mathcal{G}) - \phi(E(X \mid \mathcal{G})) \geq 0$$

which finishes the proof. □

The following very useful result is known as the Doob–Dynkin Lemma.

Theorem 1.13 *Let X and Y be s.v.'s. Y is $\sigma(X)$–measurable if and only if there exists a Borel function $g : \mathbb{R} \to \mathbb{R}$ so that $Y = g(X)$.*

Proof: It is clear that if $g : \mathbb{R} \to \mathbb{R}$ is a Borel function, then $g(X)$ is $\sigma(X)$–measurable so it is the other direction which is the important one.

Hence assume that Y is $\sigma(X)$–measurable. We first assume that $Y = 1_A$ with $A \in \sigma(X)$. Then there is a Borel set $B \subseteq \mathbb{R}$ with $A = X^{-1}(B)$ and it is then clear that $Y = 1_A = 1_{X^{-1}(B)} = 1_B(X)$. This shows that we can choose $g = 1_B$ in this case.

If Y is a simple function, say $Y = \sum_{k=1}^n a_k 1_{A_k}$ where $a_k \in \mathbb{R}$ for all $1 \leq k \leq n$ and $A_k \in \sigma(X)$ for all $1 \leq k \leq n$ with $A_k \cap A_j = \emptyset$ for $k \neq j$, then we can find Borel sets $B_k \subseteq \mathbb{R}$ so that $A_k = X^{-1}(B_k)$ for all $1 \leq k \leq n$. It is now clear that if we put $g = \sum_{k=1}^n a_k 1_{B_k}$, then $Y = g(X)$.

Next we let $Y \geq 0$. We can then find a sequence (Y_k) of simple $\sigma(X)$– measurable functions with $0 \leq Y_k \leq Y_{k+1}$ for all $k \in \mathbb{N}$ so that $Y = \lim_k Y_k$. By the above we can to each k find a Borel function g_k so that $Y_k = g_k(X)$ for all $k \in \mathbb{N}$. For each $n \in \mathbb{N}$ we put $h_n = \max\{g_k \mid 1 \leq k \leq n\}$. Since the Y_n's are increasing we get that for all $\omega \in \Omega$ we have $g_k(X(\omega)) = Y_k(\omega) \leq Y_n(\omega) = g_n(X(\omega)$ for all $1 \leq k \leq n$ so that $h_n(X(\omega)) = g_n(X(\omega))$. Since by definition the h_n's are increasing we can put $g = \lim_n h_n$. If now $\omega \in \Omega$, then

$$Y(\omega) = \lim_n g_n(X(\omega)) = \lim_n h_n(X(\omega)) = g(X(\omega)).$$

If Y is arbitrary, we write $Y = Y^+ - Y^-$ and apply the above on Y^+ and Y^- to get the result. □

2 Martingales

As before we have a fixed probability space $(\Omega, \mathcal{F}, P)$. To ease our notation in the future we make the following two definitions.

Definition 2.1 *Let$(\mathcal{F}_n)_{n\geq 0}$ be a sequence of sub–σ–algebras of $\mathcal{F}$. If $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$ for all $n \geq 0$, then we call $(\mathcal{F}_n)$ a filtration of $\mathcal{F}$.*

Definition 2.2 *Let $(\mathcal{F}_n)$ be a filtration of $\mathcal{F}$. A sequence (X_n) of stochastic variables is called an $(\mathcal{F}_n)$–adapted stochastic process if X_n is $\mathcal{F}_n$–measurable for all $n \geq 0$.*

If it is clear which filtration is used in the definition, we shall just talk about a stochastic process.

If (X_n) is an arbitrary sequence of s.v.'s, we can consider the filtration $(\sigma(X_0, X_1, \cdots, X_n))$ adapted to which (X_n) becomes a stochastic process. This is the filtration which is mostly used in the book. It is called (X_n)'s own filtration.

Let in the following $(\mathcal{F}_n)$ be a fixed filtration of $\mathcal{F}$.

In the book martingales, submartingales, and supermartingales adapted to $(\mathcal{F}_n)$ are defined and please note that if (X_n) is an $(\mathcal{F}_n)$–martingale, then it is also a martingale in its own filtration. Similar results hold for submartingales and supermartingales.

A motivation for the theory of martingales.

Assume we go down to the casino in Odense and play a game. If we invest 1 kr and win, we get our stake back and win 1 kr. If we loose, we have lost our stake, that is we have lost 1 kr. The probability to win is p where $0 \leq p \leq 1$. The individual games are independent of each other. Hence we get a sequence (X_n) of independent stochastic variables with $P(X_n = 1) = p$ and $P(X_n = -1) = 1 - p$ for all $n \in \mathbb{N}$. It is readily verified that $E(X_n) = 2p - 1$. For every $n \in \mathbb{N}$ we let $S_n = \sum_{k=1}^{n} X_k$ and $\mathcal{F}_n = \sigma(X_1, X_2, \ldots, X_n)$. Since X_{n+1} is independent of $\mathcal{F}_n$ for all n, $E(X_{n+1} \mid \mathcal{F}_n) = E(X_{n+1}) = 2p - 1$ and hence we get

$$E(S_{n+1} \mid \mathcal{F}_n) = S_n + E(X_{n+1}) = S_n + 2p - 1. \tag{2.1}$$

This shows that (S_n) is a submartingale if $p > \frac{1}{2}$, a martingale if $p = \frac{1}{2}$, and a supermartingale if $p < \frac{1}{2}$.

Let us now consider the possibility to improve our result by making a strategy by looking on the results of the first $n-1$ games and then decide what stakes to make in the n'th game. To formalize this we assume that we when we start, we have X_0 kr to our disposal, and X_0 is constant. We further assume that for every $n \in \mathbb{N}$ we have an $\mathcal{F}_{n-1}$–measurable function $g_n : \Omega \to [0, \infty[$ (we put $\mathcal{F}_0 = \{\emptyset, \Omega\}$), that is g_n is the stake we want to do in the n'th game, based on our knowledge of the first $n - 1$ games. It follows from an n–dimensional version of Theorem 1.13 that for every $n \in \mathbb{N}$ there is a Borel measurable function $\phi_n :]0, \infty[\times \{-1, 1\}^{n-1} \to [0, \infty[$ so that $g_n = \phi_n(X_0, X_1, X_2, \cdots, X_{n-1})$. Note that we also allow $g_n(\omega) = 0$ which means that we do not take part in the n'th game. Since everyone has limited ressources for disposal, it is reasonable to assume that all the g_n's are bounded. We now put

$$U_n = X_0 + \sum_{k=1}^{n} g_k X_k \quad \text{for alle } n \geq 0$$

which gives the result of the first n games

Since g_n is bounded for every n, $g_n X_n \in L_1(P)$ and since g_n er $\mathcal{F}_{n-1}$–measurable we get $E(g_n X_n \mid \mathcal{F}_{n-1}) = g_n E(X_n \mid \mathcal{F}_{n-1}) = (2p-1)g_n$ for all $n \in \mathbb{N}$ and hence

$$E(U_n \mid \mathcal{F}_{n-1}) = U_{n-1} + (2p-1)g_n.$$

This shows that (U_n) is a submartingale if $p > \frac{1}{2}$, a martingale if $p = \frac{1}{2}$, and a supermartingale if $p < \frac{1}{2}$. Note that in the case of a submartingale or supermartingale we have a strict inequality for those ω's for which $g_n(\omega) > 0$.

This means that we cannot change the result using a strategy as above! Note also that that the result does not depend on the upper bounds of the (g_n)'s. The boundedness of the g_n's was only used to conclude that $g_n X_n$ is integrable.

It also follows from the results in the book that a strategy using stopping times does not help. □

The next proposition provides an important example of a martingale and should be compared to the motivation just given.

Proposition 2.3 *Let $(X_n)_{n\geq 0} \subseteq L_1(P)$ be a sequence of independent stochastic variables with $E(X_k) = 0$ for all $k \geq 0$, put $S_n = \sum_{k=0}^{n} X_k$, and $\mathcal{F}_n = \sigma(X_0, X_1, \cdots, X_n)$ for all $n \geq 0$. Then (S_n) is a martingale.*

Proof: Let $n \geq 0$ be arbitrary and write $S_{n+1} = S_n + X_{n+1}$. We then get:

$$E(S_{n+1} \mid \mathcal{F}_n) = S_n + E(X_{n+1} \mid \mathcal{F}_n) = S_n,$$

where we in the last equality have used that since X_{n+1} is independent of $\mathcal{F}_n$,
$E(X_{n+1} \mid \mathcal{F}_n) = E(X_{n+1}) = 0$ □
We also have:

Proposition 2.4 *Let $X \in L_1(P)$ and define $X_n = E(X \mid \mathcal{F}_n)$ for all $n \geq 0$. Then (X_n) is a martingale.*

Proof: If $n \geq 0$ is given, then

$$\begin{aligned} E(X_{n+1} \mid \mathcal{F}_n) &= E(E(X \mid \mathcal{F}_{n+1}) \mid \mathcal{F}_n) \\ E(X \mid \mathcal{F}_n) &= X_n \end{aligned}$$

□

The next definition concerns stopping times. Please note the slight difference between our definition and the one in the book.

Definition 2.5 *A function $\tau : \Omega \to \mathbb{N} \cup \{0\} \cup \{\infty\}$ is called a stopping time (adapted to $(\mathcal{F}_n)$) if $(\tau = n) \in \mathcal{F}_n$ for all $0 \leq n < \infty$.*

In the sequel we shall adopt the convention that $\inf \emptyset = \infty$.

The next propostion provides an important example of a stopping time.

Proposition 2.6 *Let (X_n) be a stochastic process (adapted to $(\mathcal{F}_n)$) and let $A \subseteq \mathbb{R}$ be a Borel set. If we define*

$$\tau(\omega) = \inf\{n \geq 0 \mid X_n \notin A\} \quad \textit{for all } \omega \in \Omega,$$

then τ is a stopping time.

Proof: Note that by the above convention $\tau(\omega) = \infty$ if $X_n(\omega) \in A$ for all $0 \leq n < \infty$!

Let now $0 \leq n < \infty$ be arbitrary. We have to show that $(\tau = n) \in \mathcal{F}_n$. Noting that if $\omega \in \Omega$, then $\tau(\omega) = n$ if and only if $X_k(\omega) \in A$ for all $k < n$ and $X_n(\omega) \notin A$ we immediately get

$$(\tau = n) = \cap_{k=0}^{n-1} X^{-1}(A) \cap (\Omega \backslash X_n^{-1}(A)),$$

from where it follows that $(\tau = n) \in \mathcal{F}_n$ $\square$

If $X = (X_n)$ is a stochastic process and τ is a finite stopping time, then as in the book we define the stochastic variable X_τ by

$$X_\tau(\omega) = X_{\tau(\omega)}(\omega) \quad \text{for all } \omega \in \Omega$$

In the book it is not shown that X_τ is measurable so this we do here.

Lemma 2.7 *Let $X = (X_n)$ be a stochastic process and let τ be a finite stopping time. Then X_τ is measurable, i.e a stochastic variable.*

Proof: Let $B \in \mathbb{R}$ be an arbitrary Borel set. Since τ is finite, we get that $\Omega = \cup_{n=0}^{\infty} (\tau = n)$ and hence

$$X_\tau^{-1}(B) = \cup_{n=0}^{\infty}((\tau = n) \cap X_\tau^{-1}(B)) = \cup_{n=1}^{\infty}((\tau = n) \cap X_n^{-1}(B) \in \mathcal{F}$$

which shows that X_τ is measurable. $\square$

As in the book we now want to study upcrossings in order to prove Doob's upcrossing inequality which is Theorem 26.4 in the book. This result is the main tool to prove the martingale convergence theorem. As a motivation we start by defining upcrossings for a sequence of real numbers.

We recall that if $A \subseteq \mathbb{N}$ is non–empty, then A has a first element $\min A$. If $A \subseteq \mathbb{N}$, we therefore define $\min A$ as usual if $A \neq \emptyset$ and put $\min A = \infty$ if $A = \emptyset$. Let $(x_n)_{n=0}^{\infty} \subseteq \mathbb{R}$ be a sequence and $a, b \in \mathbb{R}$ with $a < b$. Inductively we define tallene:

$$s_1 = \min\{n \geq 0 \mid x_n < a\} \qquad t_1 = \min\{n > s_1 \mid x_n > b\}$$

and for $k \geq 2$

$$s_k = \min\{n > t_{k-1} \mid x_n < a\} \qquad t_k = \min\{n > s_k \mid x_n > b\}.$$

Definition 2.8 *The number of upcrossings from a to b of the sequence (x_n) is defined to be ∞ if $t_k < \infty$ for all $k \in \mathbb{N}$ og to be k, if $t_k < \infty$ and $t_{k+1} = \infty$.*

Try in a coordinate system to put the n's on the x–axis, place the interval $[a, b]$ on the y–axis, and from each n you go x_n up the y–axis.

We have the following:

Lemma 2.9 *The sequence (x_n) is convergent in $[-\infty, \infty]$ if and only if the number of upcrossings from a to b af (x_n) is finite for all rational numbers a, b with $a < b$*

Proof: Let us first assume that $x_n \to x \in [-\infty, \infty]$ and let $a, b \in \mathbb{Q}$ with $a < b$. Then either $x > a$ or $x < b$. In the first case there is an n_0 so that $x_n > a$ for all $n \geq n_0$. If we choose k with $t_k \geq n_0$, then $s_{k+1} = \infty$ and hence the number of upcrossings will be less than or equal to k. The case where $x < b$ can be treated in a similar manner.

Let us now assume that (x_n) divergent. Then $-\infty \leq \liminf x_n < \limsup x_n \leq \infty$ and since $\mathbb{Q}$ is dense in $\mathbb{R}$, we can find $a, b \in \mathbb{Q}$ with $\liminf x_n < a < b < \limsup x_n$. From the definition of $\liminf$ and $\limsup$ we get the following inequalities for all $k \in \mathbb{N}$

$$\inf\{x_n \mid n \geq k\} < a$$

$$\sup\{x_n \mid n \geq k\} > b.$$

From these it follows immediately that $x_n > b$ for for infinitely many n and that $x_n < a$ for infinitely many n and this of course gives that the number of upcrossings from a to b is infinite. □

Let now $X = (X_n)_{n \geq 0}$ be an $(\mathcal{F}_n)$–adapted stochastic process and $a, b \in \mathbb{R}$ with $a < b$ be given. For a given $\omega \in \Omega$ we wish to estimate the number of upcrossings from a to b of $(X_n(\omega))$. In analogy with the above we define $T_0 = 0$ and :

$$\begin{aligned} S_1(\omega) &= \min\{n \geq 0 \mid X_n(\omega) < a\} \\ T_1(\omega) &= \min\{n > S_1(\omega) \mid X_n(\omega) > b\}, \end{aligned}$$

and inductively for $k \geq 2$:

$$\begin{aligned} S_k(\omega) &= \min\{n > T_{k-1}(\omega) \mid X_n(\omega) < a\} \\ T_k(\omega) &= \min\{n > S_k(\omega) \mid X_n(\omega) > b\}. \end{aligned}$$

The next lemma shows that we really have stopping times.

Lemma 2.10 *S_k og T_k are stopping times for all $k \in \mathbb{N}$.*

Proof: The proof is by induction. Let $n \in \mathbb{N}$. For $k = 1$ we find:

$$\begin{aligned} (S_1 = 0) &= (X_0 < a) \in \mathcal{F}_0 \\ (S_1 = n) &= \cap_{m=0}^{n-1}(X_m \geq a) \cap (X_n < a) \in \mathcal{F}_n. \end{aligned}$$

Note that (besides T_0) S_1 is the only one which can take the value 0. Further we get

$$(T_1 = n) = \cup_{m=0}^{n-1}(S_1 = m, X_{m+1}, \cdots, X_{n-1} \leq b, X_n > b) \in \mathcal{F}_n.$$

Let now $k \geq 2$ and assumed that we have proved that $(S_j)_{j=1}^{k-1}$ and $(T_j)_{j=1}^{k-1}$ are stopping times. We then get:

$$(S_k = n) = \cup_{m=1}^{n-1}(T_{k-1} = m, X_{m+1}, X_{m+2}, \cdots, X_{n-1} \geq a, X_n < a) \in \mathcal{F}_n,$$

so that S_k is a stopping time and we continue with:

$$(T_k = n) = \cup_{m=1}^{n-1}(S_k = m, X_{m+1}, \cdots, X_{n-1} \leq b, X_n > b) \in \mathcal{F}_n$$

which shows that T_k is a stopping time. □

We have the following

Definition 2.11 *Let $n \in \mathbb{N}$, og lad $a, b \in \mathbb{R}$ with $a < b$. The number of upcrossings $U_n[a,b](\omega)$ of $(X_n(\omega)$ until the time n is defined to be k if $T_k(\omega) \leq n$ and $T_{k+1}(\omega) > n$.*

Note that $U_n[a,b](\omega) = 0$ if and only if $T_1(\omega) > n$.

The following small result is missing in the book.

Theorem 2.12 *If n, a og b are as in Definition 2.11, then $U_n[a,b]$ is a measurable function.*

Proof: We consider first the case $U_N[a,b] = 0$ and get

$$(U_n[a,b] = 0) = (T_1 > n) \in \mathcal{F}_n \subseteq \mathcal{F}$$

since T_1 is a stopping time. For $k \geq 1$ we get:

$$(U_n[a,b] = k) = (T_k \leq n) \cap (T_{k+1} > n) \in \mathcal{F}_n \subseteq \mathcal{F}$$

since T_k and T_{k+1} are stopping times. This shows that $U_n[a,b]$ is measurable. □

We can now formulate and prove the book's Theorem 26.4. Our proof is roughly the same as in the book, but we do it in more detail.

Theorem 2.13 *Lad n, a og b be as before and let (X_n) be a submartingale (with respect to $(\mathcal{F}_n)$). Then*

$$E(U_n[a,b]) \leq (b-a)^{-1}E[(X_n - a)^+] \tag{2.2}$$

Proof: Put $Y_n = (X_n - a)^+$. Since the function $\phi(x) = (x-a)^+$ for all $x \in \mathbb{R}$ is convex and increasing, we know from Jensen's inequality (Exercise 6) that (Y_n) is a submartingale. Since by definition $S_{n+1} > n$, we get that

$$Y_n = Y_{S_1 \wedge n} + \sum_{i=1}^{n} (Y_{S_{i+1} \wedge n} - Y_{S_i \wedge n}) = \tag{2.3}$$

$$Y_{S_1 \wedge n} + \sum_{i=1}^{n} (Y_{T_i \wedge n} - Y_{S_i \wedge n}) + \sum_{i=1}^{n} (Y_{S_{i+1} \wedge n} - Y_{T_i \wedge n}). \tag{2.4}$$

Let now $\omega \in \Omega$ so that $U_n[a,b](\omega) = k \in \mathbb{N}$. Then $T_k(\omega) \le n$ and $T_{k+1}(\omega) > n$ and therefore we have:

$$\sum_{i=1}^{n} (Y_{T_i \wedge n} - Y_{S_i \wedge n})(\omega) = \sum_{i=1}^{k} (Y_{T_i} - Y_{S_i})(\omega) + (Y_n - Y_{S_{k+1} \wedge n})(\omega) \ge \tag{2.5}$$

$$k(b-a) + (Y_n - Y_{S_{k+1} \wedge n}(\omega)) \tag{2.6}$$

where whe have used that $Y_{T_i}(\omega) - Y_{S_i}(\omega) \ge (b-a)$ for all $1 \le i \le k$. If $S_{k+1}(\omega) \ge n$, then the last term in the last inequality of (2.5) is 0, and if $S_{k+1}(\omega) < n$, then $Y_{S_{k+1} \wedge n}(\omega) = (X_{S_{k+1}}(\omega) - a)^+ = 0$ in which case the term becomes $Y_n(\omega) \ge 0$. Hence we can remove the last term in the equation and get that the left hand side is greater than or equal to $(b-a)k$. Since $Y_{S_1 \wedge n} \ge 0$, we get all in all

$$(b-a)k \le Y_n - \sum_{i=1}^{n} (Y_{S_{i+1} \wedge n} - Y_{T_i \wedge n}) \tag{2.7}$$

or written in another way

$$(b-a)U_n[a,b](\omega) \le Y_n(\omega) - \sum_{i=1}^{n} (Y_{S_{i+1} \wedge n} - Y_{T_i \wedge n})(\omega) \tag{2.8}$$

We also have to verify (2.8) in case $U_n[a,b](\omega) = 0$, but in that case the left hand side reduces to 0 while the left hand side reduces to $Y_n(\omega) \ge 0$, because already $T_1(\omega) > n$. Hence (2.8) holds for all $\omega \in \Omega$. Since it follows from the book's Theorem 24.6 that $E(Y_{S_{i+1} \wedge n} - Y_{T_i \wedge n}) \ge 0$ for all $n \in \mathbb{N}$, we get by taking expectation in (2.8) that

$$(b-a)E(U_n[a,b]) \le E(Y_n)$$

which was what we wanted. □

We can now show the important

Theorem 2.14 *Let* (X_n) *be a submartingale so that* $\sup_n E(|X_n|) < \infty$. *Then there is an* $X \in L_1(P)$ *so that* $X_n \to X$ *n.s.*

Proof: Put $K = \sup_n E(|X_n|)$. If $a, b \in \mathbb{R}$ with $a < b$, it is clear that the sequence $(U_n[a,b])$ is increasing so we put

$$U_\infty[a,b] = \lim_n U_n[a,b].$$

Remembering how upcrossings are counted, it also follows that $U_\infty[a,b]$ is the number of upcrossings from a to b. By Theorem 2.13 and the monotone convergence theorem it follows that

$$E(U_\infty[a,b]) = \lim_n E(U_n[a,b]) \leq (b-a)^{-1}(K + |a|),$$

and hence $U_\infty[a,b] < \infty$ a.s. If we put

$$A = \cap(U_\infty[a,b] < \infty \mid a, b \in \mathbb{Q}, a < b),$$

then $P(A) = 1$. If $\omega \in A$, it follows that the number of upcrossings from a to b of the sequence $(X_n(\omega))$ is finite for all $a, b \in \mathbb{Q}$, $a < b$, and therefore that sequence is convergent. We define

$$X(\omega) = \lim_n X_n(\omega) \quad \text{for all } \omega \in A,$$

and hence X is a random variable defined almost everywhere and with values in $[-\infty, \infty]$. However., the Fatous lemma gives us that

$$E(|X|) \leq \liminf E(|X_n|) \leq K,$$

which shows that $|X| < \infty$ a.s. and that $X \in L_1(P)$. □

One should note that it does not follow that (X_n) converges to X in $L_1(P)$. We shall in the next sections discuss what conditions should be put on the X_n's in order to achieve that.

Let us end this section with Doob's decomposition theorem.

Theorem 2.15 *Let (X_n) be an $(\mathcal{F}_n)$–adapted process. Then (X_n) has a Doob decomposition*

$$X_n = X_0 + M_n + A_n \quad \textit{for all } n \geq 0,$$

where (M_n) is a martingale with $M_0 = 0$ and (A_n) with $A_0 = 0$ is a predictable process, which means that A_n is $\mathcal{F}_{n-1}$–measurable for all $n \geq 1$. (M_n) and (A_N) are uniquely determined up to "almost surely".

(X_n) is a submartingale if and only if $A_n \leq A_{n+1}$ a.s. for all $n \geq 0$.

Proof: Assume (M_n) and (A_n) satisfy the assumptions in the theorem. Then for all $n \geq 1$ we have

$$\begin{aligned} E(X_n - X_{n-1} \mid \mathcal{F}_{n-1}) &= E(M_n - M_{n-1} \mid \mathcal{F}_{n-1}) + \\ E(A_n - A_{n-1} \mid \mathcal{F}_{n-1}) &= A_n - A_{n-1} \end{aligned}$$

which shows that

$$A_n = \sum_{k=1}^{n} E(X_k - X_{k-1} \mid \mathcal{F}_{k-1}) \tag{2.9}$$

and of course

$$M_n = X_n - X_0 - A_n. \tag{2.10}$$

This shows that (A_n) and (M_n) are uniquely determined. To prove the existence we define A_n as in (2.9) and put $A_0 = 0$ and M_n as in (2.10). Clearly $X_n = X_0 + M_n + A_n$ for all $n \geq 0$. It follows directly from (2.9) that A_n is $\mathcal{F}_{n-1}$–measurable for all $n \geq 1$ so we need to show that (M_n) defined by (2.10) is a martingale. For $n \geq 0$ we get

$$\begin{aligned} E(M_n \mid \mathcal{F}_{n-1}) &= E(X_n \mid \mathcal{F}_{n-1}) - X_0 - E(A_n \mid \mathcal{F}_{n-1}) = \\ E(X_n \mid \mathcal{F}_{n-1}) + X_0 + A_n &= E(X_n \mid \mathcal{F}_{n-1}) - X_0 - \sum_{k=1}^{n} E(X_k - X_{k-1} \mid \mathcal{F}_{n-1}) = \\ X_{n-1} - X_0 - A_{n-1} &= M_{n-1}, \end{aligned}$$

which shows that (M_n) is a martingale.

If (X_n) is a submartingale, then each term in the definition of (A_n) is non–negative and therefore (A_n) is almost surely increasing. On the other hand, if (A_n) is increasing, then for all $n \geq 1$ we get

$$E(X \mid \mathcal{F}_n) = X_0 + M_{n-1} + A_n \geq X_0 + M_{n-1} + A_{n-1} = X_{n-1},$$

which shows that (X_n) is a submartingale. □

3 L_2–martingales

We start with a simple result from general Hilbert space theory. We recall that if H is a Hilbert space with inner product $(\cdot, \cdot)$, then a sequence $(x_n) \subseteq H$ is called an orthogonal sequence if $(x_n, x_m) = 0$ for all $n \neq m$. We have

Proposition 3.1 *Let H be a Hilbert space and $(x_n) \subseteq H$ be an orthogonal sequence. Then $\sum_{k=1}^{\infty} x_k$ converges in H if and only if $\sum_{k=1}^{\infty} \|x_k\|^2 < \infty$. If this is the case, then $\| \sum_{k=1}^{\infty} x_k\|^2 = \sum_{k=1}^{\infty} \|x_k\|^2$.*

Proof: For every $n \in \mathbb{N}$ we put $s_n = \sum_{k=1}^{n} x_k$ and $u_n = \sum_{k=1}^{n} \|x_k\|^2$. By the Pythagoras Theorem we get for all $n < m$:

$$\|s_m - s_n\|^2 = \| \sum_{k=n+1}^{m} x_k\|^2 = \sum_{k=1}^{n} \|x_k\|^2 = u_m - u_n,$$

which shows that (s_n) is a Cauchy sequence in H if and only if (u_n) is a Cauchy sequence in $\mathbb{R}$. Since H is complete, we get the result.

If we know the convergence, then for all ≥ 1 we get

$$\|s_n\|^2 = \sum_{k=1}^{n} \|x_k\|^2$$

and if we let $n \to \infty$, we get the desired formula. □

We recall that the real $L_2(P)$ is a Hilbert space with the inner product

$$(f, g) = \int_\Omega f g dP \quad \text{for all } f, g \in L_2(P).$$

We also recall that from our construction of conditional expectations it follows that $E(\cdot \mid \mathcal{F}_n)$ is the orthogonal projection of $L_2(P)$ onto the subspace $L_2(\Omega, \mathcal{F}_n, P)$ for all $n \geq 0$.

If $(M_n) \subseteq L_2(P)$ is a martingale, then for every $n \geq 1$ we get that $E(M_n - M_{n-1} \mid \mathcal{F}_{n-1}) = 0$, and this means that $M_n - M_{n-1}$ is orthogonal to $L_2(\Omega, \mathcal{F}_{n-1}, P)$. Hence it follows that the sequence $(M_n - M_{n-1})$ is an orthogonal sequence in $L_2(P)$. Therefore we get for all $n \geq 1$

$$\|M_n\|_2^2 = \|M_0\|_2^2 + \sum_{k=1}^{n} \|M_k - M_{k-1}\|_2^2. \tag{3.1}$$

This observation gives rise to the following convergence theorem for L_2–martingales.

Theorem 3.2 *Let $(M_n) \subseteq L_2(P)$ be a martingale. (M_n) is bounded in $L_2(P)$ if and only $\sum_{k=1}^{\infty} \|M_k - M_{k-1}\|_2^2 < \infty$.*

When this is the case, there is an $M_\infty \in L_2(P)$ so that $M_n \to M_\infty$ a.s. and in $L_2(P)$. Moreover, for all $n \geq 0$ we have that $E(M_\infty \mid \mathcal{F}_n) = M_n$.

Proof: It is clear from (3.1) that (M_n) is bounded in $L_2(P)$ if and only if $\sum_{k=1}^{\infty} \|M_k - M_{k-1}\|_2^2 < \infty$.

If this series converges, we can write $M_n = M_0 + \sum_{k=1}^{n}(M_k - M_{k-1})$ and use Proposition 3.1 to get that (M_n) converges in $L_2(P)$. We put

$$M_\infty = lim_n M_n \quad \text{in } L_2(P)$$

Since $E(|M_n|) = \|M_n\|_1 \leq \|M_n\|_2$ for all n, we have that (M_n) is also bounded in $L_1(P)$ and hence the martingale convergence theorem gives, that there is a $Y \in L_1(P)$ so that $M_n \to Y$ a.s., but then of course $M_\infty = Y$ a.s.

Exercise 5 gives that for every $k \in \mathbb{N}$ $E(\cdot \mid \mathcal{F}_k)$ is a continuous operation on $L_2(P)$, and since $M_n \to M_\infty$ in $L_2(P)$, we get that $E(M_n \mid \mathcal{F}_k) \to E(M_\infty \mid \mathcal{F}_k)$, but for $n \geq k$ we have $E(M_n \mid \mathcal{F}_k) = M_k$ and hence $E(M_\infty \mid \mathcal{F}_k) = M_k$. □

Let us give an application of this convergence theorem.

Theorem 3.3 *Let $(X_k) \subseteq L_2(P)$ be a sequence of independent random variables so that $E(X_k) = 0$ for all $k \in \mathbb{N}$. Put $\sigma_k^2 = E(X_K^2)$ for all $k \in \mathbb{N}$.*

(i) $\sum_{k=1}^\infty \sigma_k^2 < \infty$ if and only if $\sum_{k=1}^\infty X_k$ converges in $L_2(P)$. In that case the latter sum also converges almost surely.

(ii) Assume that there exists a constant $K > 0$ so that $|X_k| \le K$ a.s. If $\sum_{k=1}^\infty X_k$ converges a.s., then $\sum_{k=1}^\infty \sigma_k^2 < \infty$.

Proof: For every $n \in \mathbb{N}$ we put $M_n = \sum_{k=1}^n X_k$ and $A_n = \sum_{k=1}^n \sigma_k^2$. Further we let $\mathcal{F}_n = \sigma(X_1, X_2, \cdots, X_n)$ and put for convenience $M_0 = 0$, $A_0 = 0$, and $\mathcal{F}_0 = \{\Omega, \emptyset\}$.

We know from earlier results and exercises that (M_n) is a martingale.

Since $E((M_n - M_{n-1})^2) = \sigma_n^2$ (i) follows directly from Theorem 3.2.

(ii) For every $n \ge 0$ we put $N_n = M_n^2 - A_n$ and wish to prove that (N_n) is a martingale. The argument for this is similar to the one given in Exercise 7. Since X_k is independent of $\mathcal{F}_{k-1}$ we get

$$E((M_k - M_{k-1})^2 \mid \mathcal{F}_{k-1}) = E(X_k^2 \mid \mathcal{F}_{k-1}) = E(X_k^2) = \sigma_k^2,$$

and therefore

$$\sigma_k^2 = E(M_k^2 \mid \mathcal{F}_{k-1}) + M_{k-1}^2 + 2E(M_k M_{k-1} \mid \mathcal{F}_{k-1}) = E(M_k^2 \mid \mathcal{F}_{k-1}) - M_{k-1}^2.$$

If we add A_{k-1} on both side of this equation and reorganize the terms, we get $E(N_k \mid \mathcal{F}_{k-1}) = N_{k-1}$ which shows that (N_n) is a martingale. If $c \in \mathbb{N}$ is arbitrary and we define $\tau = \inf\{n \mid |M_n| > c\}$, then τ is a stopping time by Proposition 2.6 and it follows from the book's Theorem 24.6 or Exercise 18, that $(N_{\tau\wedge n})$ is a martingale. In particular

$$E(M_{\tau\wedge n}^2) - E(A_{\tau\wedge n}) = E(N_0) = 0. \tag{3.2}$$

Since $((\tau \wedge n) - 1 < \tau$, $|M_{(\tau\wedge n)-1}| \le c$ and therefore

$$|M_{\tau\wedge n}| \le |X_{\tau\wedge n}| + |M_{(\tau\wedge n-1}| \le K + c,$$

whence from (3.2) we obtain

$$E(A_{\tau\wedge n}) = E(M_{\tau\wedge n}^2) \le (K + c)^2 \quad \text{for all } n \in \mathbb{N}. \tag{3.3}$$

Since $\sum_{k=1}^\infty X_k$ converges almost surely, we get that the sequence $(M_n(\omega))$ is bounded for almost all ω; in order words that

$$P(\cup_{c=1}^\infty \cap_{n=1}^\infty (|M_n| \le c)) = 1.$$

This implies that there exists a c so that

$$P(\tau = \infty) = P(\cap_{n=1}^\infty |M_n| \le c) > 0.$$

Actually we can get that probability so close to 1 as we wish, by choosing c big enough. (3.3) now gives:

$$P(\tau = \infty)A_n \leq E(A_{\tau \wedge n}) \leq (K + c)^2 \quad \text{for all } n \in \mathbb{N}$$

and hence

$$A_n \leq P(\tau = \infty)^{-1}(K + c)^2 \quad \text{for all } n \in \mathbb{N},$$

but then $\sum_{k=1}^{\infty} \sigma_k^2 \leq P(\tau = \infty)^{-1}(K + c)^2 < \infty$. □

4 Uniformly integrable martingales

In this section we shall see what is needed for a martingale to converge in $L_1(P)$.

We recall the following definition (See also the book, page 105):

Definition 4.1 *A subset $\mathcal{H} \subseteq L_1(P)$ is called uniformly integrable if*

$$\lim_{x \to \infty} (\sup_{X \in \mathcal{H}} \int_{(|X| \geq x)} |X| dP) = 0. \tag{4.1}$$

The first proposition is rather obvious.

Proposition 4.2 *If $\mathcal{H} \subseteq L_1(P)$ is uniformly integrable, then $\mathcal{H}$ is a bounded subset of $L_1(P)$.*

Proof: Since $\mathcal{H}$ is uniformly integrable, we can find an $x_0 > 0$ so that

$$\int_{(|X| \geq x_0)} |X| dP \leq 1 \quad \text{for all } X \in \mathcal{H},$$

but then for all $X \in \mathcal{H}$ we have

$$\begin{aligned} \|X\|_1 = \int_\Omega |X| dP &= \int_{(|X| \geq x_0)} |X| dP + \int_{(|X| < x_0)} |X| dP \leq \\ 1 + x_0 P(|X| < x_0) &\leq 1 + x_0. \end{aligned}$$

This shows that $\mathcal{H}$ is bounded in $L_1(P)$. □

We will now find some criteria for uniform integrability. One could hope that the other direction of Proposition 4.2 is also true but this is clearly false. Indeed e.g. the unit ball of $L_1(0, 1)$ is bounded by definition, but clearly not uniformly integrable. However, we have

Theorem 4.3 *Let $1 < p \leq \infty$. If $\mathcal{H}$ is a bounded subset of $L_p(P)$, then $\mathcal{H}$ is uniformly integrable.*

Proof: Let first $p = \infty$. By definition there is a $K \geq 0$ so that $|X| \leq K$ a.e. for all $X \in \mathcal{H}$, but then $P(|X| \geq x) = 0$ for all $x > K$ and all $X \in \mathcal{H}$. This clearly lead to uniform integrability.

Let now $1 < p < \infty$. By definition there is a $K \geq 0$ so that $\|X\|_p^p \leq K$. Note that if $v, x \in \mathbb{R}$ with $0 < x \leq v$, then $(\frac{v}{x})^{p-1} \geq 1$ and multiplying with v on both sides we get that $v \leq x^{1-p} v^p$. Using this simple inequality we get the following for all $X \in \mathcal{H}$ and all $x > 0$

$$\int_{(|X| \geq x)} |X| dP \leq x^{1-p} \int_{(|X| \geq x)} |X|^p dP \leq x^{1-p} K. \tag{4.2}$$

Let now $\varepsilon > 0$ be arbitrary. Since $x^{1-p} \to 0$ for $x \to \infty$, we can find an $x_0 > 0$ so that $x^{1-p} K \leq \varepsilon$ for all $x \geq x_0$. For such x we get from (4.2) that for all $x \in \mathcal{H}$ we have

$$\int_{(|X| \geq x)} |X| dP \leq \varepsilon$$

which shows the uniform integrability of $\mathcal{H}$. □

We also need

Theorem 4.4 *If $X \in L_1(P)$, then the family $\{E(X \mid \mathcal{G}) \mid \mathcal{G} \quad$ a sub–σ–algebra of $\mathcal{F}\}$ is uniformly integrable.*

Proof: Let $\varepsilon > 0$ and choose $\delta > 0$ so that

$$\forall A \in \mathcal{F} : P(A) < \delta \Rightarrow \int_A |X| dP < \varepsilon \tag{4.3}$$

and let $x > \delta^{-1} E(|X|)$ be arbitrary.

If $\mathcal{G}$ is any subalgebra of $\mathcal{F}$, then Jensen's inequality gives that

$$|E(X \mid \mathcal{G}) \leq E(|X| \mid \mathcal{G}) \tag{4.4}$$

and therefore $E(|E(X \mid \mathcal{G})|) \leq E(|X|)$ and

$$xP(|E(X \mid \mathcal{G})| \geq x) \leq E(|E(X \mid \mathcal{G})|) \leq E(|X|).$$

This inequality and the choice of x implies that $P(|E(X \mid \mathcal{G})| \geq x) < \delta$. Since $(|E(X \mid \mathcal{G})| \geq x) \in \mathcal{G}$, we get from (4.4) that

$$\begin{aligned} \int_{(|E(X|\mathcal{G})| \geq x)} |E(X \mid \mathcal{G})| dP &\leq \int_{(|E(X|\mathcal{G})| \geq x)} E(|X| \mid \mathcal{G}) dP = \\ \int_{|E(X|\mathcal{G})| \geq x)} |X| dP &\leq \varepsilon. \end{aligned}$$

This shows that our family is uniformly integrable. □

Note that Theorem 4.4 implies that if (X_n) is a martingale and there exists an $X \in L_1(P)$ so that $E(X \mid \mathcal{F}_n) = X_n$ for all $n \in \mathbb{N}$, then necessarily (X_n) is uniformly integrable.

Before we can prove our main theorem of this section, we need

Theorem 4.5 *Let $(X_n) \subseteq L_1(P)$ and let $X \in L_1(P)$. If $X_n \to X$ a.s. and (X_n) is uniformly integrable, then $X_n \to X$ in $L_1(P)$.*

Proof: If $K > 0$, we define the function $\phi_K : \mathbb{R} \to \mathbb{R}$ by $\phi_K(x) = x$ for all $-K \leq x \leq K$, $\phi_K(x) = K$ for all $x > K$, and $\phi_K(x) = -K$ for all $x < -K$. It is easy to see that ϕ_K has the following properties (check it!!):

(i) $|\phi_K(x) - x| \leq |x|$ for all $x \in \mathbb{R}$.

(ii) $|\phi_K(x) - \phi_K(y)| \leq |x - y|$ for all $x, y \in \mathbb{R}$

By (i) we get that for all $n \in \mathbb{N}$

$$\int_\Omega |\phi_K(X_n) - X_n| dP \leq \int_{(|X_n| > K)} |X_n| dP$$

and

$$\int_\Omega |\phi_K(X) - X| dP \leq \int_{(|X| > K)} |X| dP$$

Let now $\varepsilon > 0$ be given. Using the uniform integrability of the sequence (X_n) and of $\{X\}$ we get from these integral inequalities that there exists a $K > 0$ so that

(iii) $\|\phi_K(X_n) - X_n\|_1 \leq \frac{\varepsilon}{3}$ for all $n \in \mathbb{N}$.

(iv) $\|\phi_K(X) - X\|_1 \leq \frac{\varepsilon}{3}$.

We now fix such a K. Since $X_n \to X$ a.s., it follows from (ii) that $\phi_K(X_n) \to \phi_K(X)$ a.s. and the dominated convergence theorem therefore gives that $\phi_K(X_n) \to \phi_K(X)$ in $L_1(P)$. Hence we can find an $n_0 \in \mathbb{N}$ so that $\|\phi_K(X_n) - \phi_K(X)\|_1 \leq \frac{\varepsilon}{3}$ for all $n \geq n_0$. The triangle inequality gives that for all $n \geq n_0$ we have

$$\|X - X_n\|_1 \leq \|X - \phi_K(X)\|_1 + \|\phi_K(X) - \phi_K(X_n)\|_1 + \|\phi_K(X_n) - X_n\|_1 \leq \varepsilon.$$

and therefore $X_n \to X$ in $L_1(P)$. $\square$

We are now ready to state and prove our main result.

Theorem 4.6 *Let $(X_n) \subseteq L_1(P)$ be a martingale. The following statements are equivalent:*

(i) (X_n) is uniformly integrable.

(ii) There is an $X_\infty \in L_1(P)$ so that $X_n \to X_\infty$ in $L_1(P)$.

(iii) There is an $X \in L_1(P)$ so that $E(X \mid \mathcal{F}_n) = X_n$ for all $n \geq 0$.

If (ii) (or one of the equivalent statements) holds, then $X_n \to X_\infty$ a.s. and $E(X_\infty \mid \mathcal{F}_n) = X_n$ for all $n \geq 0$.

Proof: $(i) \Rightarrow (ii)$**:** If (X_n) is uniformly integrable, then it is bounded in $L_1(P)$ by Proposition 4.2 and the martingale convergence theorem therefore gives that there is an $X_\infty \in L_1(P)$ so that $X_n \to X_\infty$ a.s. Theorem 4.5 now gives that gives that $X_n \to X_\infty$ in $L_1(P)$.

$(ii) \Rightarrow (iii)$**:** If (ii) holds then the continuity in $L_1(P)$ of $E(\cdot \mid \mathcal{F}_n)$ ensures that

$$E(X_\infty \mid \mathcal{F}_n) = \lim_m E(X_m \mid \mathcal{F}_n) = X_n \quad \text{in } L_1(P),$$

which proves (iii).

$(iii) \Rightarrow (i)$**:** If (iii) holds, then Theorem 4.4 shows that (X_n) is uniformly integrable.

The implication $(i) \to (ii)$ shows that $X_n \to X_\infty$ a.s. and the implication $(ii) \to (iii)$ shows that $E(X_\infty \mid \mathcal{F}_n) = X_n$ for all $n \geq 0$. □

One can ask what the relation between X_∞ satisfying (ii) and an X satisfying (iii) in Theorem 4.6 is. The answer is given by the following corollary.

Corollary 4.7 *Let $(X_n) \subseteq L_1(P)$ be a uniformly integrable martingale, let X_∞ satisfy (ii) in Theorem 4.6, and let $X \in L_1(P)$ satisfy (iii) in that theorem. Then $E(X \mid \mathcal{F}_\infty) = X_\infty$ where $\mathcal{F}_\infty = \sigma(\mathcal{F}_n \mid n \geq 0)$*

Proof: We note that since $X_n \to X_\infty$ a.s and every X_n is $\mathcal{F}_n$–measurable, X_∞ is $\mathcal{F}_\infty$–measurable.

Let $\mathcal{G}$ be the class of all those $A \in \mathcal{F}$ for which

$$\int_A X dP = \int_A X_\infty dP. \tag{4.5}$$

It is easy to see (and left to the reader) that $\mathcal{G}$ is a σ–algebra. Let now $n \geq 0$ and let $A \in \mathcal{F}_n$ be arbitrary. Since $E(X \mid \mathcal{F}_n) = X_n = E(X_\infty \mid \mathcal{F}_n)$, it follows that $\int_A X dP = \int_A X_\infty dP$ which implies that $\mathcal{F}_n \subseteq \mathcal{G}$. Since this is true for all n and $\mathcal{G}$ is a σ–algebra, it follows that $\mathcal{F}_\infty \subseteq \mathcal{G}$. Hence (4.5) holds for all $A \in \mathcal{F}_\infty$ and since X_∞ is $\mathcal{F}_\infty$–measurable, the defition of the conditional expectation shows that $E(X \mid \mathcal{F}_\infty) = X_\infty$. □

5 Strong Law of Large Numbers

In this section we shall give a proof of the Strong Law of Large numbers based on martingale theory.

We start by defining a backwards martingale

Definition 5.1 *Let $X \in L_1(P)$ and let $\mathcal{G}_{-n}$ be a sequence of sub–σ–algebras of $\mathcal{F}$ so that*

$$\mathcal{G}_{-(n+1)} \subseteq \mathcal{G}_{-n} \quad \textit{for all } n \geq 1.$$

For every $n \in \mathbb{N}$ we put $X_{-n} = E(X \mid \mathcal{G}_{-n})$. (X_{-n}) is called a backwards martingale.

The reason for the name is of course that $X_{-(n+1)} = E(X_{-n} \mid \mathcal{G}_{-(n+1)})$ for all $n \in \mathbb{N}$.

We have the following theorem on backwards martingales.

Theorem 5.2 *Let (X_{-n}) be a backwards martingale as in Definition 5.1 and put $\mathcal{G}_{-\infty} = \cap_{n=1}^{\infty} \mathcal{G}_{-n}$. There exists an $X_{-\infty} \in L_1(P)$ so that $X_{-n} \to X_{-\infty}$ for $n \to \infty$ both a.s. and in $L_1(P)$. Actually $X_{-\infty} = E(X_1 \mid \mathcal{G}_{-\infty}) = E(X \mid \mathcal{G}_{-\infty})$*

Proof: If $n \in \mathbb{N}$ we consider the finite sequence $X_{-n}, X_{-(n-1)}, \cdots, X_{-1}$ which is a finite martingale starting with X_{-n} and ending with X_{-1}. If $a < b$ we let $U_{-n}[a,b]$ denote the number of upcrossings from a to b for that martingale. From the upcrossing Theorem 2.13 we get that

$$E(U_{-n}[a,b]) \leq (b-a)^{-1} E((X_{-1}-a)^+) \leq (b-a)^{-1}(E(|X_{-1}|)+|a|) \leq (b-a)^{-1}(E(|X|)+|a|).$$

The sequence $(U_{-n}[a,b])$ is increasing a.s. and if we put $U_{-\infty}[a,b] = \lim_n U_{-n}[a,b]$, then it is readily verified that this limit is the number of downcrossings from b to a of the sequence (X_{-n}) (we say it that way when we go backwards). The monotone convergence theorem then gives that

$$E(U_{-\infty}[a,b]) = \lim_n E(U_{-n}[a,b]) \leq E(|X|) + |a| < \infty.$$

Therefore $U_{-\infty}[a,b] < \infty$ a.s. Since this holds for all $a < b$, there exists an sv $X_{-\infty} : \Omega \to [-\infty, \infty]$ so that $X_{-\infty} = \lim_n X_{-n}$. The Fatou Lemma now gives us that

$$E(|X_{-\infty}|) \leq \liminf_n E(|X_{-n}|) \leq E(|X_{-1}|) < \infty$$

which implies that $X_{-\infty} \in \mathbb{R}$ a.s. and that $X_{-\infty} \in L_1(P)$.

Since $X_{-n} = E(X \mid \mathcal{G}_{-n})$ for all $n \in \mathbb{N}$, we get from Theorem 4.4 that (X_{-n}) is uniformly integrable and hence Theorem 4.5 implies that $X_{-n} \to X_{-\infty}$ in $L_1(P)$.

To get the last statement we let $A \in \mathcal{G}_{-\infty}$ be arbitrary and observe that by the L_1–convergence we get

$$\int_A X_{-\infty} dP = \lim_n \int_A E(X_{-1} \mid \mathcal{G}_{-n}) dP = \int_A X_{-1} dP,$$

where the last equality holds because $A \in \mathcal{G}_{-n}$ for all n. This shows that $X_{-\infty} = E(X_{-1} \mid \mathcal{G}_{-\infty})$. □

Before we can prove the Strong Law of Large numbers we need to recall a few results from general measure theory and probability theory. We start with:

Definition 5.3 *Let $X : \Omega \to \mathbb{R}$ be a stochastic variable. We define the Borel probability measure $X(P)$ on $\mathbb{R}$ by*

$$X(P)(A) = P(X^{-1}(A)) \quad \textit{for all } \mathcal{B}.$$

$X(P)$ is called the distribution of X, the law of X, or simply the image measure of P by X.

By defintion two sv's X and Y are identically distributed when $X(P) = Y(P)$. We also recall:

Theorem 5.4 *Let X be an sv and let $f : \mathbb{R} \to \mathbb{R}$ be a Borel–measurable function. Then $f \circ X \in L_1(P)$ if and only $f \in L_1(X(P))$ and in that case*

$$\int_\Omega f \circ X dP = \int_{-\infty}^{\infty} f dX(P).$$

Note that if $f \circ X \in L_1(P)$ and A is a Borel set and we use the above formula with $f1_A$ instead of f we get

$$\int_{X^{-1}(A)} f \circ X dP = \int_A f dX(P).$$

We need a generalization of Definition 5.3 and Theorem 5.4 to stochastic variables taking values in $\mathbb{R}^n$. If $X_k : \Omega \to \mathbb{R}$, $1 \le k \le n$ are stochastic variables, we can make the stochastic variable $X : \Omega \to \mathbb{R}^n$ having the X_k's as its coordinates, i.e. $X = (X_1, X_2, \cdots, X_n)$. Similar to Definition 5.3 we have:

Definition 5.5 *Let $n \in \mathbb{N}$ and let $X : \Omega \to \mathbb{R}^n$ be an sv. We define the Borel probability measure $X(P)$ on $\mathbb{R}^n$ by*

$$X(P)(A) = P(X^{-1}(A)) \quad \textit{for all } A \in \mathcal{B}^n.$$

Here $\mathcal{B}^n$ denotes the set of Borel subsets of $\mathbb{R}^n$. $X(P)$ is called the distribution of X, the law of X, or simply the image measure of P by X.

If $X_k : \Omega \to \mathbb{R}$, $1 \le k \le n$, are the coordinates of X, i.e $X = (X_1, X_2, \cdots, X_n)$, $X(P)$ is also called the joint distribution of $X_1, X_2, \cdots, X_n$.

Similar to Theorem 5.4 we have the following theorem, the proof of which is roughly the same.

Theorem 5.6 *Let $n \in \mathbb{N}$, let $X : \Omega \to \mathbb{R}^n$ be an sv, and let $f : \mathbb{R}^n \to \mathbb{R}$ be a Borel–measurable function. Then $f \circ X \in L_1(P)$ if and only $f \in L_1(X(P))$ and in that case*

$$\int_\Omega f \circ X dP = \int_{\mathbb{R}^n} f dX(P)$$

Again we can note that if $f \circ X \in L_1(P)$ and $A \subseteq \mathbb{R}^n$ is a Borel set, we can use the above formula with $f1_A$ instead of f to get

$$\int_{X^{-1}(A)} f \circ X dP = \int_A f dX(P). \tag{5.1}$$

If $X : \Omega \to \mathbb{R}^n$ is an sv, say $X = (X_1, X_2, \cdots, X_n)$, it is in general difficult to express the distribution $X(P)$ in terms of the distributions $X_k(P)$, $1 \le k \le n$. However, if the X_k's are independent then it is easy as the next result shows.

Theorem 5.7 *Let $n \in \mathbb{N}$, $X_k : \Omega \to \mathbb{R}$, $1 \le k \le n$, and put $X = (X_1, X_2, \cdots, X_n)$. Then the X_k's are independent if and only if*

$$X(P) = \bigotimes_{k=1}^{n} X_k(P) \tag{5.2}$$

where $\bigotimes_{k=1}^{n} X_k(P)$ denotes the product measure of the $X_k(P)$'s.

Proof: In order to check when (5.2) holds, it is enough to check when the two measures are equal on boxes in $\mathbb{R}^n$. Hence let A_k, $1 \le k \le n$ be Borel subsets of $\mathbb{R}$ and put $A = \prod_{k=1}^{n} A_k$. Note that

$$X^{-1}(A) = \cap_{k=1}^{n} X_k^{-1}(A_k),$$

and therefore

$$X(P)(A) = P(\cap_{k=1}^{n} X_k^{-1}(A_k)) \tag{5.3}$$

On the order hand, by definition of the product measure we have

$$\bigotimes_{k=1}^{n} X_k(P)(A) = \prod_{k=1}^{n} X_k(P)(A_k) = \prod_{k=1}^{n} P(X_k^{-1}(A_k)). \tag{5.4}$$

If the X_k's are independent, then the right hand sides of (5.3) and (5.4) are equal for all choices of the A_k's and therefore the left hands sides are equal too which means that (5.2) holds.

If (5.2) holds, then the left hand sides of (5.3) and (5.4) are equal for all choices of the A_k's and the right hand sides are equal too, which means that the X_k's are independent. □

We are going to discuss some results on measures on $\mathbb{R}^n$ which have to be used in the proof of The Strong Law of Large Numbers.

Let $n \in \mathbb{N}$ be fixed. If π is a permutation of the numbers $\{1, 2, \cdots, n\}$, we can define the map $\Pi : \mathbb{R}^n \to \mathbb{R}^n$ by $\Pi(x_1, x_2, \cdots, x_n) = (x_{\pi(1)}, x_{\pi(2)}, \cdots, x_{\pi(n)})$ for all $x_1, x_2, \cdots, x_n \in \mathbb{R}$. Clearly Π is Borel measurable. Often we shall not distinguiss between Π and π and just call Π a permutation on $\mathbb{R}^n$, meaning that Π is a map on $\mathbb{R}^n$ which permutes the coordinates.

Let μ be a Borel probability measure on $\mathbb{R}$ and let μ^n denote the n–fold product of μ with itself, so μ^n is a Borel probability measure on $\mathbb{R}^n$. Note that if Π is a permutation on $\mathbb{R}^n$, then $\Pi(\mu^n) = \mu^n$. Indeed, if $A = \prod_{k=1}^{n} A_k$ is a box in $\mathbb{R}^n$, then

$$\begin{aligned} \Pi(\mu^n)(A) &= \mu^n(\prod_{k=1}^{n} A_{\pi^{-1}(k)}) = \\ \prod_{k=1}^{n} \mu(A_{\pi^{-1}(k)}) &= \mu^n(A). \end{aligned}$$

If $f \in L_1(\mu^n)$ and $A \in \mathcal{B}^n$, then (5.1) gives

$$\int_{\Pi^{-1}(A)} f \circ \Pi d\mu^n = \int_A f d\mu^n. \tag{5.5}$$

If in addition A is Π–invariant, i.e. $\Pi(A) = A$ (equivalently $\pi^{-1}(A) = A$), then we get

$$\int_A f \circ \Pi d\mu^n = \int_A f d\mu^n. \tag{5.6}$$

The next theorem and its corollary are very useful in the proof of our main theorem of this section.

Theorem 5.8 *Let μ be a Borel probability measure on $\mathbb{R}$, let $f \in L_1(\mu^n)$, and let Π be a permutation on $\mathbb{R}^n$. Further we define $s_n : \mathbb{R}^n \to \mathbb{R}$ by:*

$$s_n(x_1, x_2, \cdots, x_n) = \sum_{k=1}^{n} x_k$$

If $B \subseteq \mathbb{R}$ is a Borel set, then $s_n^{-1}(B)$ is Π–invariant and

$$\int_{s_n^{-1}(B)} f \circ \Pi d\mu^n = \int_{s_n^{-1}(B)} f d\mu^n \tag{5.7}$$

Proof: If $\{x_k \mid 1 \leq k \leq n\} \subseteq \mathbb{R}$, then

$$s_n \circ \Pi(x_1, x_2, \cdots, x_n) = \sum_{k=1}^{n} x_{\pi(k)} = s_n(x_1, x_2, \cdots, x_n)$$

which shows that $s_n \circ \Pi = s_n$ and hence $s_n^{-1}(B) = \Pi^{-1} s_n^{-1}(B)$ so that $s_n^{-1}(B)$ is Π–invariant. The conclusion of the theorem now follows from (5.6). □

Corollary 5.9 *Let μ and s_n be as in Theorem 5.8 and let us for every $1 \leq k \leq n$ define*

$$p_k(x_1, x_2, \cdots, x_n) = x_k \tag{5.8}$$

for all $x_1, x_2, \cdots, x_n \in \mathbb{R}$. If $p_1 \in L_1(\mu^n)$ and $B \subseteq \mathbb{R}$ is a Borel set, then

$$\int_{s_n^{-1}(B)} p_k d\mu^n = \int_{s_n^{-1}(B)} p_1 d\mu^n \quad \textit{for all } 1 \leq k \leq n. \tag{5.9}$$

Proof: Let$1 \leq k \leq n$. If we define the permutation Π_k by

$$\Pi_k(x_1, x_2, \cdots, x_k, \cdots, x_n) = (x_k, x_2, \cdots, x_1, \cdots, x_n),$$

then clearly $p_k = p_1 \circ \Pi_k$ and hence the conclusion of the corollary follows from (5.7) of Theorem 5.8 $\square$

After these measure theoretical considerations we are finally going back to the main subject. Our first result states:

Theorem 5.10 *Let $n \in \mathbb{N}$, let $\{X_k \mid 1 \leq k \leq n\} \subseteq L_1(P)$ be independent and identically distributed, and put $S_n = \sum_{k=1}^n X_k$. Then*

$$E(X_k \mid \sigma(S_n)) = E(X_1 \mid \sigma(S_n)) = \frac{1}{n} S_n \tag{5.10}$$

Proof: Let $\mu = X_1(P)$ and let $X = (X_1, X_2, \cdots, X_n)$. Since the X_k's are identically distributed, $\mu = X_k(P)$ for all $1 \leq k \leq n$ and since they are independent, it follows from Theorem 5.7 that $\mu^n = X(P)$. If we let p_k, $1 \leq k \leq n$ and s_n be as above, we note that $S_n = s_n(X)$ and hence (5.1) of Theorem 5.6 and (5.9) of Corollary 5.9 give for every Borel set $B \subseteq \mathbb{R}$:

$$\begin{aligned} \int_{S_n^{-1}(B)} X_k dP &= \int_{X^{-1} s_n^{-1}(B)} p_k(X) dP = \\ \int_{s_n^{-1}(B)} p_k d\mu^n &= \int_{s_n^{-1}(B)} p_1 d\mu^n = \\ \int_{S_n^{-1}(B)} X_1 dP. \end{aligned} \tag{5.11}$$

Since $\sigma(S_n) = \{S_n^{-1}(B) \mid B \subseteq \mathbb{R} \text{ a Borel set}\}$, we get from (5.11) that $E(X_k \mid \sigma(S_n)) = E(X_1 \mid \sigma(S_n))$, but then

$$\begin{aligned} \frac{1}{n} S_n = \frac{1}{n} E(S_n \mid \sigma(S_n)) &= \\ \frac{1}{n} \sum_{k=1}^n E(X_k \mid \sigma(S_n)) &= E(X_1 \mid \sigma(S_n)). \end{aligned}$$

$\square$

We are now ready to prove:

Theorem 5.11 The Strong Law of Large Numbers. *Let $(X_n) \subseteq L_1(P)$ be a sequence of independent and identically distributed stochastic variables and put $S_n = \sum_{k=1}^n X_k$ for all $n \in \mathbb{N}$. Then $\frac{1}{n} S_n \to E(X_1)$ both a.s and in $L_1(P)$.*

Proof: For every $n \in \mathbb{N}$ we put $\mathcal{F}_{-n} = \sigma(S_k \mid k \geq n) = \sigma(S_n, X_k \mid k \geq n+1)$. Please think about the last equality. If we define $X_{-n} = E(X_1 \mid \mathcal{F}_{-n})$, then (X_n) is a backwards martingale.

Since for every n $\sigma(X_1, S_n)$ is independent of $\sigma(X_k \mid k \geq n+1)$, it follows from Exercise 21 and Theorem 5.10 that

$$X_{-n} = E(X_1 \mid \sigma(S_n)) = \frac{1}{n} S_n \quad \text{for all } n \in \mathbb{N}.$$

Theorem 5.2 now gives that there is an $X_{-\infty} \in L_1(P)$ so that $\frac{1}{n} S_n \to X_{-\infty}$ both a.s. and in $L_1(P)$. The L_1–convergence implies that

$$E(X_1) = E(\lim_n \frac{1}{n} S_n) = E(X_{-\infty}).$$

Note that if $k \in \mathbb{N}$ is fixed, then $\frac{1}{n}\sum_{j=1}^{k} X_j \to 0$ for $n \to \infty$ and hence $\frac{1}{n}\sum_{j=1}^{n} X_{k+j} \to X_{-\infty}$ a.s. This shows that $X_{-\infty}$ is measurable with respect to the tail algebra $\cap_{k=1}^{\infty} \sigma\{X_m \mid m \geq k\}$ of the X_k's. Therefore it is constant a.s. by Kolmogorov's $0-1$ law which implies that $X_{-\infty} = E(X_1)$ a.s. $\square$