Lecture Notes on Martingales

by Russell Lyons
Extended from the book by Patrick Billingsley

Suppose we gamble on a sequence of fair games. Let X_n be our fortune after n plays and $\mathscr{F}_n := \sigma(X_0, X_1, \ldots, X_n)$ or more; e.g., we might use other randomness. Fairness is $X_n = E[X_{n+1} \mid \mathscr{F}_n]$. Unfavorable gambling has $\ge$ and favorable has $\le$. Note $\mathscr{F}_n \subseteq \mathscr{F}_{n+1}$. One way to be favorable is to get handed some money before each fair game, where the amount has conditional expectation ≥ 0 and the future games remain fair.

Definition. Let $X_n \in L^1(\Omega, \mathscr{F}, P)$ and $\mathscr{F}_n \subseteq \mathscr{F}$ be σ-fields. We call $\langle \mathscr{F}_n \rangle$ a ***filtration*** if $\mathscr{F}_n \subseteq \mathscr{F}_{n+1}$. We call $\langle X_n \rangle$ ***adapted*** to $\langle \mathscr{F}_n \rangle$ if $X_n \in \mathscr{F}_n$ (i.e., X_n is $\mathscr{F}_n$-measurable). We call $\langle (X_n, \mathscr{F}_n) \rangle$ a ***martingale*** (***submartingale***; ***supermartingale***) if $\langle X_n \rangle$ is adapted to the filtration $\langle \mathscr{F}_n \rangle$ and

$$\forall n \qquad X_n = E[X_{n+1} \mid \mathscr{F}_n] \text{ a.s. } (\le\ ;\ \ge)$$

Clearly, $\langle (X_n, \mathscr{F}_n) \rangle$ is a submartingale iff $\langle (-X_n, \mathscr{F}_n) \rangle$ is a supermartingale. We'll later see the reason for the names of the prefixes; it is related to subharmonic functions (convex functions in the case of $\mathbb{R}$).

We call $\langle X_n \rangle$ a ***martingale*** if there is some filtration with respect to which it is a martingale. In such a case, $\mathscr{F}_n := \sigma(X_1, \ldots, X_n)$ always works. Similar definitions and observations hold for submartingales and supermartingales. Note that if $\langle (X_n, \mathscr{F}_n) \rangle$ is a martingale, then $E[X_m \mid \mathscr{F}_n] = X_{m \wedge n}$ a.s. and $E[X_n]$ is the same for all n. Similarly, if $\langle (X_n, \mathscr{F}_n) \rangle$ is a submartingale, then $E[X_m \mid \mathscr{F}_n] \ge X_{m \wedge n}$ a.s. and $E[X_n]$ is increasing in n.

Example 35.1. Sums of independent random variables.

Example 35.5. If $Z \in L^1(\Omega, \mathscr{F}, P)$ and $\langle \mathscr{F}_n \rangle$ is a filtration, then $\langle (E[Z \mid \mathscr{F}_n], \mathscr{F}_n) \rangle$ is a martingale.

Extra Credit Homework XC1: Show that $\langle n\mathbf{1}_{[0,1/n]}\ ;\ n \ge 1 \rangle$ is a martingale on $[0,1]$ with respect to Lebesgue measure.

Homework HW1: 35.2.

Homework HW2: 35.6 (correction: $X_0 = E[X_1]$). Also, show that $\langle Z_n \rangle$ is ***predictable***, i.e., for all n, Z_n is $\mathscr{F}_{n-1}$-measurable.

Example 35.2. On $(\Omega, \mathscr{F})$, let P be a probability measure and ν be a finite, signed measure. Let $\langle \mathscr{F}_n \rangle$ be a filtration. Suppose that $\nu{\restriction}\mathscr{F}_n \ll P{\restriction}\mathscr{F}_n$ for all n. Let $X_n := \mathrm{d}(\nu{\restriction}\mathscr{F}_n)/\mathrm{d}(P{\restriction}\mathscr{F}_n)$. Then $\langle (X_n, \mathscr{F}_n) \rangle$ is a martingale:

$$\begin{aligned} E[X_{n+1} \mid \mathscr{F}_n](P{\restriction}\mathscr{F}_n) &= (X_{n+1}P){\restriction}\mathscr{F}_n = (X_{n+1}P{\restriction}\mathscr{F}_{n+1}){\restriction}\mathscr{F}_n \\ &= (\nu{\restriction}\mathscr{F}_{n+1}){\restriction}\mathscr{F}_n = \nu{\restriction}\mathscr{F}_n = X_n(P{\restriction}\mathscr{F}_n). \end{aligned}$$

Example 35.3. Let P be Lebesgue measure on $(0,1]$ and $\mathscr{F}_n := \sigma(I_k^{(n)}\,;\ 0 \le k < 2^n)$, where $I_k^{(n)} := (k2^{-n}, (k+1)2^{-n}]$. Since $(A \in \mathscr{F}_n, P(A) = 0) \Rightarrow A = \varnothing$, we have for all ν that $\nu \restriction \mathscr{F}_n \ll P \restriction \mathscr{F}_n$. In this case,

$$X_n = \sum_{k=0}^{2^n-1} \mathbf{1}_{I_k^{(n)}} \nu(I_k^{(n)})/P(I_k^{(n)}) = \sum_{k=0}^{2^n-1} \mathbf{1}_{I_k^{(n)}} \nu(I_k^{(n)}) 2^n.$$

Example 35.8. If $\langle (X_n, \mathscr{F}_n) \rangle$ is a martingale, then $\langle (|X_n|, \mathscr{F}_n) \rangle$ is a submartingale.

This is a special case of:

Theorem 35.1. *Let $\langle X_n \rangle$ be integrable random variables adapted to a filtration $\langle \mathscr{F}_n \rangle$, φ be convex on an interval containing all the ranges of X_n, and $\varphi \circ X_n \in L^1(P)$. Then $\langle (\varphi \circ X_n, \mathscr{F}_n) \rangle$ is a submartingale if either*

(i) *$\langle (X_n, \mathscr{F}_n) \rangle$ is a martingale;*
(ii) *$\langle (X_n, \mathscr{F}_n) \rangle$ is a submartingale and φ is increasing; or*
(iii) *$\langle (X_n, \mathscr{F}_n) \rangle$ is a supermartingale and φ is decreasing.*

Proof. In all cases, we have $\varphi(X_n) \le \varphi\big(E[X_{n+1} \mid \mathscr{F}_n]\big)$. By Jensen's inequality, this is $\le E\big[\varphi(X_{n+1}) \mid \mathscr{F}_n\big]$. ∎

We now depart from the book.

Wald's Equation. *Let $Z_n \in L^1(P)$ for $n \ge 1$, τ be an $\mathbb{N}$-valued random variable, and $\mu \in \mathbb{R}$. Suppose that*

(i) *for each $n \ge 1$, we have $P[\tau \ge n] > 0 \Rightarrow E[Z_n \mid \tau \ge n] = \mu$ ($\le$; $\ge$)*

and

(ii) *one of the following holds:*
 (a) *$\forall n\ \ Z_n \ge 0$;*
 (b) *$\sup_{n\,;\,P[\tau \ge n]>0} E\big[|Z_n| \bigm| \tau \ge n\big] < \infty$ and $E[\tau] < \infty$;*
 (c) *$E\big[|\sum_{n=1}^{\tau} Z_n|\big] < \infty$ and $\lim_{n\to\infty} E\big[\sum_{k=1}^{n} Z_k \mathbf{1}_{[\tau > n]}\big] = 0$.*

Then

$$E\bigg[\sum_{n=1}^{\tau} Z_n\bigg] = \mu\, E[\tau], \quad (\le\ ;\ \ge)$$

where $0 \cdot \infty := 0$.

Proof. In case (ii)(a), we have by Tonelli's theorem (or the MCT) that

$$\begin{aligned} E\bigg[\sum_{n=1}^{\tau} Z_n\bigg] &= E\bigg[\sum_{n=1}^{\infty} Z_n \mathbf{1}_{[n \le \tau]}\bigg] = \sum_{n=1}^{\infty} E[Z_n \mathbf{1}_{[\tau \ge n]}] \\ &= \sum_{n=1}^{\infty} E[Z_n \mid \tau \ge n] P[\tau \ge n] = \sum_{n=1}^{\infty} \mu\, P[\tau \ge n] = \mu\, E[\tau]. \end{aligned} \tag{N1}$$

In case (ii)(b), we have

$$\begin{aligned} E\bigg[\sum_{n=1}^{\infty} |Z_n| \mathbf{1}_{[n \le \tau]}\bigg] &= \sum_{n=1}^{\infty} E\big[|Z_n| \bigm| \tau \ge n\big] P[\tau \ge n] \\ &\le E[\tau] \sup_n E\big[|Z_n| \bigm| \tau \ge n\big] < \infty, \end{aligned}$$

whence we may apply Fubini's theorem as in (N1).

Finally, in case (ii)(c), the LDCT gives

$$E\Bigg[\sum_{k=1}^{\tau} Z_k\Bigg] = \lim_{n\to\infty} E\Bigg[\sum_{k=1}^{\tau} Z_k \mathbf{1}_{[\tau\le n]}\Bigg] = \lim_{n\to\infty} E\Bigg[\sum_{k=1}^{n} Z_k \mathbf{1}_{[k\le\tau\le n]}\Bigg]$$

$$= \lim_{n\to\infty} E\Bigg[\sum_{k=1}^{n} Z_k \big(\mathbf{1}_{[\tau\ge k]} - \mathbf{1}_{[\tau>n]}\big)\Bigg] = \lim_{n\to\infty} \sum_{k=1}^{n} \mu\, P[\tau \ge k] = \mu\, E[\tau].$$

(Note the cases $\mu = 0$ and $E[\tau] = \infty$ or $\mu = \infty$ and $E[\tau] = 0$.) ∎

Definition. Given a filtration $\langle \mathscr{F}_n \rangle$, a random variable $\tau \in \mathbb{N} \cup \{\infty\}$ is called a ***stopping time*** with respect to $\langle \mathscr{F}_n \rangle$ if $\forall n \in \mathbb{N}\ \ [\tau = n] \in \mathscr{F}_n$. Equivalently, $\forall n \in \mathbb{N}\ \ [\tau \le n] \in \mathscr{F}_n$.

We will write X_τ for the random variable $X_\tau(\omega) := X_{\tau(\omega)}(\omega)$ when $\tau < \infty$ a.s. This is measurable because $X_\tau = \sum_{n=0}^{\infty} X_n \mathbf{1}_{[\tau=n]}$.

Homework HW3: Show that if τ_1 and τ_2 are stopping times with respect to $\langle \mathscr{F}_n \rangle$, then so are $\tau_1 \wedge \tau_2$, $\tau_1 \vee \tau_2$, and $\tau_1 + \tau_2$.

Corollary N1. *Let $\langle (X_n, \mathscr{F}_n)\,;\ n \ge 0 \rangle$ be a martingale (submartingale; supermartingale) and τ be an a.s. finite stopping time with respect to $\langle \mathscr{F}_n \rangle$. If*

(a) $\exists c \le 0\ \ \forall n\ \ X_{n+1} - X_n \ge c$ *a.s. and* $E[\tau] < \infty$

or

(b) $E\big[|X_\tau|\big] < \infty$ *and* $\lim_n E[X_n \mathbf{1}_{[\tau>n]}] = 0$,

then

$$E[X_\tau] = E[X_0]. \quad (\ge\ ;\ \le)$$

Proof. In case (a), apply Wald's equation with $Z_n := X_n - X_{n-1} - c$ for $n \ge 1$ and $\mu := -c$. In case (b), apply Wald's equation with $Z_n := X_n - X_{n-1}$ for $n \ge 1$ and $\mu := 0$. Note that $[\tau \ge n] = [\tau \le n-1]^{\mathrm{c}} \in \mathscr{F}_{n-1}$. ∎

Example N2. It is not immediately clear when the hypotheses (b) hold of Corollary N1. One easy case is when $\sup \tau < \infty$. More generally, if $\{X_{\tau\wedge n}\,;\ n \ge 0\}$ is uniformly integrable (UI), then the hypotheses (b) hold, because $E\big[|X_\tau|\big] = \lim_n E\big[|X_{\tau\wedge n}|\big]$ and $E[X_n \mathbf{1}_{[\tau>n]}] = E[X_{\tau\wedge n} \mathbf{1}_{[\tau>n]}]$, with $\{X_{\tau\wedge n} \mathbf{1}_{[\tau>n]}\,;\ n \ge 0\}$ uniformly integrable.

Here are some nice applications.

Example N3. Let $\langle X_n \rangle$ be simple random walk on $\mathbb{Z}$ starting from $X_0 = 0$. Let $\tau_a := \inf\{n\,;\ X_n = a\}$. Then $E[\tau_a] = \infty$ for $a \ne 0$ since the conclusion of Corollary N1 does not hold. But if $\tau := \tau_a \wedge \tau_{-b}$ for $a, b > 0$, then $|X_{\tau\wedge n}|$ are bounded by $a \vee b$, so uniformly integrable, whence

$$0 = E[X_\tau] = a\, P[\tau_a < \tau_{-b}] - b\, P[\tau_{-b} < \tau_a],$$

giving $P[\tau_a < \tau_{-b}] = b/(a+b)$.

Here, we assumed that $\tau < \infty$ a.s. This is implied by a martingale convergence theorem, Theorem 35.5, but we can also prove it as follows at the same time as we calculate $E[\tau]$.

Problem 35.2 shows that $\big\langle\big(X_n^2 - n, \sigma(X_1, \ldots, X_n)\big)\big\rangle$ is a martingale. Fix $N \in \mathbb{N}$. Then $\tau \wedge N$ is a bounded stopping time, so we may apply Corollary N1 to get

$$0 = E[X_0^2 - 0] = E[X_{\tau\wedge N}^2 - \tau \wedge N] = E[X_{\tau\wedge N}^2] - E[\tau \wedge N],$$

i.e., $E[\tau \wedge N] = E[X_{\tau\wedge N}^2]$. Now let $N \to \infty$. The right-hand side is bounded, whence so is the left-hand side; this shows that $E[\tau] < \infty$, whence $\tau < \infty$ a.s. Furthermore, the MCT and the BCT give

$$E[\tau] = E[X_\tau^2] = a^2 \frac{b}{a+b} + b^2 \frac{a}{a+b} = ab.$$

An alternative calculation of $E[\tau]$ that does not use the probabilities $P[\tau_a < \tau_{-b}]$ and $P[\tau_a > \tau_{-b}]$ uses the martingale $\big\langle(a - X_n)(X_n + b) + n\big\rangle$: We have

$$ab = (a - X_0)(X_0 + b) + 0 = E\big[(a - X_{\tau\wedge N})(X_{\tau\wedge N} + b) + \tau \wedge N\big].$$

Since $X_{\tau\wedge N}$ is bounded, it follows that so is $E[\tau \wedge N]$, whence $E[\tau] < \infty$. Letting $N \to \infty$ gives the result.

Extra Credit Homework XC2: Consider three players, A, B, and C, who play the following game: They begin with fortunes $a, b, c \in \mathbb{N}^+$, respectively. At each time, a pair of the players whose fortunes are strictly positive is chosen at random and a random one of that pair gives one unit to the other. Let X_n, Y_n, Z_n be their respective fortunes after n plays.

(a) Show that $\langle X_n Y_n Z_n + n(X_n + Y_n + Z_n)/3\rangle$ is a martingale.
(b) Show that the expected time until some player's fortune is zero equals $3abc/(a + b + c)$.
(c) Show that $\langle X_n Y_n + Y_n Z_n + Z_n X_n + n\rangle$ is a martingale.
(d) Show that the expected time until two players' fortunes are zero equals $ab + bc + ca$.

Definition. Let $p(\cdot, \cdot)$ be a transition probability on a state space, S. A function $f\colon S \to \mathbb{R}$ is called ***harmonic*** (***subharmonic***; ***superharmonic***) if

$$\forall x \in S \;\; f(x) = \sum_{y\in S} p(x, y) f(y). \quad (\le\,;\; \ge)$$

If $\langle X_n\rangle$ is a Markov chain on S with these transition probabilities and f is harmonic (subharmonic; superharmonic), then $\langle f(X_n)\rangle$ is a martingale (submartingale; supermartingale).

Example N4. Consider the asymmetric random walk $\langle X_n\rangle$ on $\mathbb{Z}$ starting from 0 with $p(n, n + 1) = p$ and $p(n, n - 1) = q := 1 - p$. Then $f(n) := (q/p)^n$ is harmonic, whence $\langle (q/p)^{X_n}\rangle$ is a martingale. If $\tau := \tau_a \wedge \tau_{-b}$, then this martingale is bounded up to time τ, whence Corollary N1 applies:

$$1 = E\big[(q/p)^{X_0}\big] = E\big[(q/p)^{X_\tau}\big] = (q/p)^a\, P[\tau_a < \tau_{-b}] + (q/p)^{-b}\, P[\tau_{-b} < \tau_a].$$

We can solve this to get another derivation of the probabilities in gambler's ruin when $p \ne q$. Again, the martingale convergence theorem will imply that $\tau < \infty$ a.s., but it is also a consequence of the following homework.

Homework HW4: Let $\langle X_n\rangle$ be asymmetric random walk on $\mathbb{Z}$ with $p > q$, $s := \sqrt{q/p}$, and $r := 2\sqrt{pq}$. Show that $\langle s^{X_n} r^{-n}\rangle$ is a martingale. Deduce that if $X_0 = 0$, then $P[\tau_a \ge n] \le s^{-a} r^n$ for $a \ge 1$.

Example N5. What is $E[\tau_a]$ when $p > q$, $a \geq 1$, and $X_0 = 0$? We use another martingale: $\langle X_n - (p-q)n \rangle$. Fix $N \in \mathbb{N}$ and define $\tau := \tau_a \wedge N$. Then τ is bounded, so $0 = E[X_0 - (p-q)0] = E[X_\tau - (p-q)\tau]$, i.e., $(p-q)E[\tau_a \wedge N] = E[X_{\tau_a \wedge N}]$. By the MCT, $E[\tau_a \wedge N] \to E[\tau_a]$ as $N \to \infty$. Since $|X_{\tau_a \wedge N}| \leq \tau_a \wedge N \leq \tau_a \in L^1$ by Homework HW4, the LDCT gives $E[X_{\tau_a \wedge N}] \to E[X_{\tau_a}] = a$ as $N \to \infty$. In conclusion,

$$E[\tau_a] = \frac{a}{p-q}.$$

Homework HW5: For $\langle X_n \rangle$ be asymmetric random walk on $\mathbb{Z}$ starting from 0 with $p > q$, calculate $E[\tau_a \wedge \tau_{-b}]$ for $a, b > 0$.

Homework HW6: For $\langle X_n \rangle$ be asymmetric random walk on $\mathbb{Z}$ starting from 0 with $p > q$, calculate $\mathrm{Var}(\tau_a)$ for $a > 0$. *Hint:* Use problem 35.2.

Example N6. Consider an ordinary deck of 26 red and 26 black cards, well shuffled. The cards are turned over one by one. At any time, and once only, we may bet that the next card is red. What's the best strategy to maximize our chance to be correct? It cannot hurt to bet, so we assume that we do bet once.

Let A_k be the event that the kth card is red. If τ is the time we bet, then $[\tau = k] \in \sigma(A_1, \ldots, A_{k-1}) =: \mathscr{F}_k$, i.e., τ is a stopping time. At time k, the conditional chance of winning would be $X_k := P(A_k \mid \mathscr{F}_k)$, so the chance of winning is $E[X_\tau]$. (Alternatively, the event of winning is $A_\tau := \bigcup_{k=1}^{52} A_k \cap [\tau = k]$ and $P(A_\tau) = E[X_\tau]$.) Now for $k \leq 51$,

$$E[X_{k+1} \mid \mathscr{F}_k] = E\big[P(A_{k+1} \mid \mathscr{F}_{k+1}) \bigm| \mathscr{F}_k\big] = P(A_{k+1} \mid \mathscr{F}_k) = X_k,$$

because all cards we have not seen are in random order. Thus, $\langle X_k\,;\ 1 \leq k \leq 52 \rangle$ is a martingale, so $E[X_\tau] = E[X_1] = 1/2$. All strategies are equally effective.

Having learned about this martingale, we realize that $\forall k\ \ X_k = E[X_{52} \mid \mathscr{F}_k]$. But $X_{52} = P(A_{52} \mid \mathscr{F}_{52}) = \mathbf{1}_{A_{52}}$, so $X_k = P(A_{52} \mid \mathscr{F}_k)$. In other words, X_k is the same as the chance to win if, at time k, we bet that the last card will be red. This shows more directly that any strategy is equivalent to betting on the last card. That's why it does not matter. In fact, we can see this without any use of martingales: betting on the next card has the same chance of winning as betting on the last card, no matter what we have seen, because all cards we have not seen are in random order.

We can learn still more. Suppose we used the strategy, for some $\alpha > 1/2$, to bet at the first k (if any), that $X_k \geq \alpha$. More precisely, let $\tau := \max\{k \leq 52\,;\ X_k \geq \alpha \text{ or } k = 52\}$. Since this strategy wins with probability 1/2, it follows that $1/2 = E[X_\tau] \geq \alpha\, P[\max_{k \leq 52} X_k \geq \alpha]$, whence $P[\max_{k \leq 52} X_k \geq \alpha] \leq 1/(2\alpha)$. Later (Theorem 35.3), we will see a more general, but similar, maximal inequality.

Recall problem 16.9: $\mathcal{X}$ is uniformly integrable iff $\sup\big\{E\big[|X|\big]\,;\ X \in \mathcal{X}\big\} < \infty$ and $\forall \epsilon > 0\ \exists \delta > 0\ \ P(A) < \delta \Rightarrow \sup\big\{E\big[|X|\,;\ A\big]\,;\ X \in \mathcal{X}\big\} < \epsilon$.

Proposition N7. *If $\mathcal{X}$ is UI and $\mathcal{G}$ is a collection of sub-σ-fields, then $\{E[X \mid \mathscr{G}]\,;\ X \in \mathcal{X},\ \mathscr{G} \in \mathcal{G}\}$ is UI.*

Example N8. $\mathcal{X} = \{Z\}$ for some $Z \in L^1(P)$ and $\mathcal{G}$ is a filtration.

Proof of Proposition N7. Let $M := \sup\{E\big[|X|\big]\,;\ X \in \mathcal{X}\}$. Given $\epsilon > 0$, choose α such that

$$\sup\{E\big[|X|\,;\ A\big]\,;\ X \in \mathcal{X},\ P(A) \le M/\alpha\} < \epsilon.$$

Now for $X \in \mathcal{X}$ and $\mathscr{G} \in \mathcal{G}$, we have

$$P\big[|E[X \mid \mathscr{G}]| \ge \alpha\big] \le \alpha^{-1}\, E\big[|E[X \mid \mathscr{G}]|\big] \le \alpha^{-1}\, E\big[E\big[|X| \bigm| \mathscr{G}\big]\big] = \alpha^{-1}\, E\big[|X|\big] \le M/\alpha.$$

Hence, with $A := \{|E[X \mid \mathscr{G}]| \ge \alpha\}$, we have

$$E\big[|E[X \mid \mathscr{G}]|\,;\ A\big] \le E\big[E\big[|X| \bigm| \mathscr{G}\big]\mathbf{1}_A\big] = E\big[E\big[|X|\mathbf{1}_A \bigm| \mathscr{G}\big]\big] = E\big[|X|\mathbf{1}_A\big] < \epsilon. \qquad \blacksquare$$

Let $\langle \mathscr{F}_n \rangle$ be a filtration and τ a stopping time. What is the information up to time τ, $\mathscr{F}_\tau$? On the part of Ω where $\tau = n$, it is $\mathscr{F}_n$. Thus, we define

$$\mathscr{F}_\tau := \{A \in \mathscr{F}\,;\ \forall n \in \mathbb{N}\ \ A \cap [\tau = n] \in \mathscr{F}_n\}.$$

This is a σ-field and $\tau \in \mathscr{F}_\tau$.

Homework HW7: Show that if $\langle X_n \rangle$ is adapted to $\langle \mathscr{F}_n \rangle$ and τ is a finite stopping time, then $X_\tau \in \mathscr{F}_\tau$.

Homework HW8: Let $\langle Y_k \rangle$ be a sequence of random variables with $E\big[\sup_k |Y_k|\big] < \infty$, $\langle \mathscr{F}_k \rangle$ be a filtration, τ be a finite stopping time, and $X_k := E[Y_k \mid \mathscr{F}_k]$. Show that Y_τ is integrable and $X_\tau = E[Y_\tau \mid \mathscr{F}_\tau]$.

For instance, in Example N6, if $Y_k := \mathbf{1}_{A_k}$, then $Y_\tau = \mathbf{1}_{A_\tau}$ and $X_\tau = P(A_\tau \mid \mathscr{F}_\tau)$.

Homework HW9: Show that if τ_1 and τ_2 are stopping times, then $\mathscr{F}_{\tau_1 \wedge \tau_2} = \mathscr{F}_{\tau_1} \cap \mathscr{F}_{\tau_2}$. *Hint:* First show that $\tau_1 \le \tau_2$ implies $\mathscr{F}_{\tau_1} \subseteq \mathscr{F}_{\tau_2}$.

Proposition N9. *Let $\langle (X_n, \mathscr{F}_n) \rangle$ be a martingale (submartingale; supermartingale) and τ be a finite stopping time. Then for all n,*

$$E[X_n \mid \mathscr{F}_\tau] = X_{\tau \wedge n} \ \text{a.s.}\ (\ge\ ;\ \le)$$

Proof. It's sufficient to do the submartingale case. Note that $X_{\tau \wedge n} \in \mathscr{F}_\tau$. Since both sides of the desired inequality are in $\mathscr{F}_\tau$, it suffices to show that for all $A \in \mathscr{F}_\tau$,

$$E[X_n\,;\ A] \ge E[X_{\tau \wedge n}\,;\ A]. \tag{N2}$$

This holds by the following calculation:

$$\begin{aligned} E[X_n \mathbf{1}_A] &= \sum_k E[X_n \underbrace{\mathbf{1}_A \mathbf{1}_{[\tau = k]}}_{\in \mathscr{F}_k}] \ge \sum_k E[X_{n \wedge k} \mathbf{1}_A \mathbf{1}_{[\tau = k]}] \\ &= \sum_k E[X_{n \wedge \tau} \mathbf{1}_A \mathbf{1}_{[\tau = k]}] = E[X_{n \wedge \tau} \mathbf{1}_A]. \qquad \blacksquare \end{aligned}$$

Homework HW10: Show that if $\langle (X_n, \mathscr{F}_n) \rangle$ is a submartingale and τ is a stopping time, then $\langle (X_{\tau \wedge n}, \mathscr{F}_n) \rangle$ is a submartingale.

Corollary N10. *If $\langle X_n \rangle$ is a UI martingale and τ is a finite stopping time, then $\{X_{\tau\wedge n}\}$ is UI.*

Proof. Apply Propositions N7 and N9. ∎

We'll now usually write our results just for the submartingale case.

Corollary N11. *If $\langle X_n \rangle$ is a UI submartingale and τ is a finite stopping time, then $\{X_{\tau\wedge n}\}$ is UI.*

Proof. Since $|X_{\tau\wedge n}| \le |X_\tau| + |X_n|$, it suffices to show that $E\big[|X_\tau|\big] < \infty$. Now $\langle X_n^+ \rangle$ is a submartingale by Theorem 35.1, so by Proposition N9, $E[X_{\tau\wedge n}^+] \le E[X_n^+]$. Also, $E[X_0] \le E[X_{\tau\wedge n}]$ by Corollary N1 to Wald's equation. Hence,

$$E\big[|X_{\tau\wedge n}|\big] = 2\,E[X_{\tau\wedge n}^+] - E[X_{\tau\wedge n}] \le 2\,E[X_n^+] - E[X_0] \le 2\,E\big[|X_n|\big] - E[X_0].$$

Fatou's lemma gives the result desired. ∎

Homework HW11: 35.7.

Proposition N12. *If $\langle X_n \rangle$ is a submartingale and $\tau_1 \le \tau_2 \le j \in \mathbb{N}$ are bounded stopping times, then $E[X_{\tau_1}] \le E[X_{\tau_2}]$.*

Proof. Write $Z_n := X_n - X_{n-1}$. Then $X_n = X_0 + \sum_{k=1}^n Z_k$, so

$$E[X_{\tau_2} - X_{\tau_1}] = E\Bigg[\sum_{k=\tau_1+1}^{\tau_2} Z_k\Bigg] = E\Bigg[\sum_{k=1}^{j} Z_k \mathbf{1}_{[\tau_1 < k \le \tau_2]}\Bigg] \ge 0$$

since $[\tau_1 < k \le \tau_2] \in \mathscr{F}_{k-1}$. ∎

The following extends Propositions N9 and N12:

The Optional Sampling Theorem. *If $\tau_1 \le \tau_2$ are finite stopping times and $\langle (X_{\tau_2\wedge n}, \mathscr{F}_n) \rangle$ is a UI submartingale, then*

$$X_{\tau_1} \le E[X_{\tau_2} \mid \mathscr{F}_{\tau_1}] \ \textit{a.s.}$$

Note that by Homework HW10 and Corollary N11, $\langle (X_{\tau_2\wedge n}, \mathscr{F}_n) \rangle$ is a UI submartingale if $\langle (X_n, \mathscr{F}_n) \rangle$ is a UI submartingale.

Proof. First, we show that $E[X_{\tau_1}] \le E[X_{\tau_2}]$. Fix $j \in \mathbb{N}$. Then $\tau_1 \wedge j \le \tau_2 \wedge j \le j$ are bounded stopping times, whence Proposition N12 gives $E[X_{\tau_1\wedge j}] \le E[X_{\tau_2\wedge j}]$. Now $E[X_{\tau_2\wedge j}] \to E[X_{\tau_2}]$ since $\{X_{\tau_2\wedge j}\}$ is UI. Furthermore, by Corollary N11, $\{X_{\tau_1\wedge j}\} = \{X_{\tau_1\wedge(\tau_2\wedge j)}\}$ is UI, whence $E[X_{\tau_1\wedge j}] \to E[X_{\tau_1}]$. Therefore $E[X_{\tau_1}] \le E[X_{\tau_2}]$. Note, too, that by UI, X_{τ_1} and X_{τ_2} are integrable.

To show the finer conclusion desired, it suffices to show that $\forall A \in \mathscr{F}_{\tau_1}\ E[X_{\tau_1}\,;\ A] \le E[X_{\tau_2}\,;\ A]$. By Homework HW9, $\mathscr{F}_{\tau_1} = \mathscr{F}_{\tau_1\wedge\tau_2} \subseteq \mathscr{F}_{\tau_2}$. Given $A \in \mathscr{F}_{\tau_1}$, set $\tau := \tau_1 \mathbf{1}_A + \tau_2 \mathbf{1}_{A^c}$. This is a stopping time, because for all n,

$$[\tau \le n] = \underbrace{\big([\tau_1 \le n] \cap A\big)}_{\in \mathscr{F}_n} \cup \underbrace{\big([\tau_2 \le n] \cap A^c\big)}_{\in \mathscr{F}_n}.$$

Also, $\tau \le \tau_2$, so by the result proved in the first paragraph, we have

$$0 \le E[X_{\tau_2} - X_\tau] = E[X_{\tau_2} - X_{\tau_1}\mathbf{1}_A - X_{\tau_2}\mathbf{1}_{A^c}] = E[X_{\tau_2}\mathbf{1}_A - X_{\tau_1}\mathbf{1}_A]. \qquad \blacksquare$$

When does a submartingale converge a.s.? Note that the heuristic that it tends to increase suggests that it always has an extended-real limit. However, this is not so, even if it is a.s. bounded:

Example N13. Let $\langle x_n \rangle$ be any sequence in $\mathbb{R}$. Then there is a martingale $\langle X_n \rangle$ that behaves like $\langle x_n \rangle$, i.e., $X_n - x_n$ is eventually constant a.s. (stabilizes a.s.). To see this, let $z_n := x_n - x_{n-1}$ and $\langle Z_n \rangle$ be independent, mean-0 random variables with $\sum_n P[Z_n \ne z_n] < \infty$. Set $X_n := \sum_{k=1}^n Z_k$. Then $\langle X_n \rangle$ is a martingale and by the Borel–Cantelli lemma, $X_n - x_n$ stabilizes a.s.

Yet L^1-boundedness assures convergence:

Theorem 35.5 (Doob's Martingale Convergence Theorem). *Let $\langle X_n \rangle$ be a submartingale (or supermartingale) with $K := \sup_n E\big[|X_n|\big] < \infty$. Then $X := \lim_{n\to\infty} X_n$ exists a.s. and $E\big[|X|\big] \le K$.*

To prove this, we will bound the number of upcrossings of intervals $[\alpha, \beta]$: Given $\langle X_1, \ldots, X_n \rangle$, set $\sigma_1 := \min\{j \in [1, n]\,;\ X_j \le \alpha \text{ or } j = n\}$, $\tau_1 := \min\{j \in (\sigma_1, n]\,;\ X_j \ge \beta \text{ or } j = n\}$, $\sigma_2 := \min\{j \in (\tau_1, n]\,;\ X_j \le \alpha \text{ or } j = n\}$, $\tau_2 := \min\{j \in (\sigma_2, n]\,;\ X_j \ge \beta \text{ or } j = n\}$, etc. These are clearly stopping times. The upcrossings are the intervals $[\sigma_i, \tau_i]$ for which $X_{\sigma_i} \le \alpha$ and $X_{\tau_i} \ge \beta$. Note that each upcrossing is responsible for an increase of $\ge \beta - \alpha$.

Theorem 35.4 (Doob's Upcrossing Inequality). *If $\langle X_1, \ldots, X_n \rangle$ is a submartingale and $\alpha < \beta$, then the number U of upcrossings of $[\alpha, \beta]$ satisfies*

$$E[U] \le \frac{E\big[(X_n - \alpha)^+\big]}{\beta - \alpha}.$$

Proof. Define $\langle Y_j \rangle := \langle (X_j - \alpha)^+ \rangle$. This is a submartingale, with U being the number of upcrossings by $\langle Y_j \rangle$ of $[0, \gamma]$, where $\gamma := \beta - \alpha$. Because all $Y_j \ge 0$, we have

$$Y_n = Y_{\tau_n} \ge Y_{\tau_n} - Y_{\sigma_1} = \sum_{k=1}^n \big(Y_{\tau_k} - Y_{\sigma_k}\big) + \sum_{k=2}^n \big(Y_{\sigma_k} - Y_{\tau_{k-1}}\big).$$

Since $n \ge \sigma_k \ge \tau_{k-1}$, Proposition N12 gives $E[Y_{\sigma_k} - Y_{\tau_{k-1}}] \ge 0$. Also, $Y_{\tau_k} - Y_{\sigma_k} \ge \gamma$ for each of the U upcrossings and is ≥ 0 otherwise. Hence $\sum_{k=1}^n \big(Y_{\tau_k} - Y_{\sigma_k}\big) \ge \gamma U$, so $E[Y_n] \ge E[\gamma U]$, i.e., $E[U] \le E[Y_n]/\gamma$. $\blacksquare$

Extra Credit Homework XC3: Show that if $\langle X_1, \ldots, X_n \rangle$ is a supermartingale and $\alpha < \beta$, then the number U of upcrossings of $[\alpha, \beta]$ satisfies

$$E[U] \le \frac{E\big[(X_n - \alpha)^-\big]}{\beta - \alpha}.$$

Hint: $(\beta - \alpha)U \le \sum_{k=1}^n \big(X_{\tau_k} - X_{\sigma_k}\big) + (X_n - \alpha)^-$.

Proof of Theorem 35.5. To show first that $\lim X_n$ exists a.s. in $[-\infty, \infty]$, it suffices to show that the number of upcrossings of each $[\alpha, \beta]$ is finite a.s. for $\alpha, \beta \in \mathbb{Q}$ with $\alpha < \beta$. Let U_n be the number of upcrossings of $[\alpha, \beta]$ by time n. Then

$$E[U_n] \le \frac{E\big[(X_n - \alpha)^+\big]}{\beta - \alpha} \le \frac{E\big[|X_n|\big] + |\alpha|}{\beta - \alpha} \le \frac{K + |\alpha|}{\beta - \alpha}.$$

Since $U_n \uparrow$, the MCT shows that $\lim U_n < \infty$ a.s., as desired.

Let $X := \lim X_n$. Then $|X| = \lim |X_n|$ and Fatou's lemma gives $E\big[|X|\big] \le K$, so $|X| < \infty$ a.s. ∎

Corollary N14. *If $\langle X_n \rangle$ is a nonnegative supermartingale, then $X := \lim_{n\to\infty} X_n$ exists a.s. and $E[X] \le E[X_1]$.* ∎

Example N15. Consider the asymmetric random walk $\langle X_n \rangle$ on $\mathbb{Z}$ starting from 0 with $p(n, n+1) = p \ge 1/2$ and $p(n, n-1) = 1 - p$. Let $\tau_a := \inf\{n\,;\ X_n = a\}$ with $a > 0$. Then $\langle a - X_{n \wedge \tau_a} \rangle$ is a nonnegative supermartingale. By Corollary N14, it converges a.s., whence $\tau_a < \infty$ a.s. It follows that if $p = 1/2$, the case of simple random walk, $\langle X_n \rangle$ is recurrent.

Extra Credit Homework XC4: Let $\langle Y_n \rangle$ be IID with law $(\delta_1 + \delta_{-1})/2$.

(a) Find $B_n \in \sigma(Y_1, \ldots, Y_n)$ such that $P(B_n) \to 0$ and $P(B_n \text{ i.o.}) = 1$.

(b) Define $X_0 := 0$, $X_{n+1} := X_n(1 + Y_{n+1}) + \mathbf{1}_{B_n} Y_{n+1}$ for $n \ge 1$. Show that $\langle X_n \rangle$ is a martingale.

(c) Show that $X_n \to 0$ in probability.

(d) Show that $P[X_n \text{ converges}] = 0$.

It is very useful to know when $E[\lim X_n] = \lim E[X_n]$. Recall from the corollary to Theorem 16.14 that if $X_n \to X$ a.s., then $\{X_n\}$ is UI iff $X_n \to X$ in L^1 (in which case $E[X_n] \to E[X]$). Thus:

Proposition N16. *If $\langle X_n \rangle$ is a submartingale, then the following are equivalent:*

(i) *$\{X_n\}$ is UI.*

(ii) *$\lim X_n$ exists a.s. and in L^1.*

(iii) *$\lim X_n$ exists in L^1.*

Proof. (ii) $\Rightarrow$ (iii) is trivial. (ii) $\Rightarrow$ (i) by the general corollary and (i) $\Rightarrow$ (ii), (iii) $\Rightarrow$ (ii) by Theorem 35.5. ∎

However, there is a martingale $\langle X_n \rangle$ that is bounded in L^1 with $E[X_n] = 0 = E[\lim X_n]$, yet $E\big[|X_n|\big] \not\to 0$. For example, let $X_1 = X_2 = X_3 = \cdots \sim (\delta_1 + \delta_{-1})/2$. It is also possible to have an example with $X_n \to 0$ a.s.; see Extra Credit Homework XC6.

We get an additional equivalence for martingales:

Theorem N17. *If $\langle (X_n, \mathscr{F}_n) \rangle$ is a martingale, then $X_n \to X$ in L^1 iff $X_n = E[X \mid \mathscr{F}_n]$ a.s. and X has a version in $\sigma\big(\bigcup_n \mathscr{F}_n\big)$.*

Write $\mathscr{F}_\infty := \sigma\big(\bigcup_n \mathscr{F}_n\big)$; we also write this as $\mathscr{F}_n \uparrow \mathscr{F}_\infty$ and as $\mathscr{F}_\infty = \bigvee_n \mathscr{F}_n$. We say that $\langle (X_n, \mathscr{F}_n) \rangle$ is ***closed*** by $(X, \mathscr{F}_\infty)$ in the situation of the theorem.

Proof. $\Rightarrow$: Because $X_n \xrightarrow{L^1} X$, we have $X_n \mathbf{1}_A \xrightarrow{L^1} X\mathbf{1}_A$ for all A. Take $A \in \mathscr{F}_m$. Then

$$E[X\,;\,A] = \lim E[X_n\,;\,A] = E[X_m\,;\,A].$$

That is, X_m is a version of $E[X \mid \mathscr{F}_m]$. Since all $X_n \in \mathscr{F}_\infty$, also X has a version in $\mathscr{F}_\infty$.

$\Leftarrow$: We have seen that $\{X_n\}$ is UI, so let $X_\infty := \limsup X_n$. We must show that $X = X_\infty$ a.s. Since $\bigcup \mathscr{F}_n$ is a π-system, it suffices to check that $\forall A \in \bigcup \mathscr{F}_n\ E[X\,;\,A] = E[X_\infty\,;\,A]$. Indeed, if $A \in \mathscr{F}_n$, then by the tower property, $E[X\,;\,A] = E[X_n\,;\,A] = E[X_\infty\,;\,A]$, the second equality following as in the other part of the proof. ∎

Theorem 35.6. *If $\mathscr{F}_n \uparrow \mathscr{F}_\infty$ and $Z \in L^1(P)$, then $E[Z \mid \mathscr{F}_n] \to E[Z \mid \mathscr{F}_\infty]$ a.s.*

Proof. Apply Theorem N17 to $X := E[Z \mid \mathscr{F}_\infty]$. ∎

Corollary (Lévy's 0-1 Law). *If $\mathscr{F}_n \uparrow \mathscr{F}_\infty$ and $A \in \mathscr{F}_\infty$, then $P(A \mid \mathscr{F}_n) \to \mathbf{1}_A$ a.s.* ∎

Corollary (Kolmogorov's 0-1 Law). *If $\{\mathscr{F}_n\}$ are independent, then the tail σ-field of $\{\mathscr{F}_n\}$ is trivial.*

Proof. Let $\mathscr{G}_n := \sigma(\mathscr{F}_1, \ldots, \mathscr{F}_n)$. If A is in the tail σ-field, then $A \in \bigvee_n \mathscr{G}_n$, so $\mathbf{1}_A = \lim P(A \mid \mathscr{G}_n) = \lim P(A)$ a.s. since A is independent of $\mathscr{G}_n$. Thus, $P(A) \in \{0,1\}$. ∎

Homework HW12: 35.4 (In fact, show that if $\exists \epsilon > 0\ \sum_n P[Y_n > 1+\epsilon] = \infty$, then $X = 0$ a.s.)

Homework HW13: Let Z_n be IID, mean-0 random variables. Let $\Theta \in L^1$ and $Y_n := \Theta + Z_n$. Show that $E[\Theta \mid Y_1, \ldots, Y_n] \to \Theta$ a.s.

Homework HW14: (Pólya's urn) An urn contains 1 red ball and 1 green ball. Choose one ball at random and replace it together with another of the same color. Repeat indefinitely. Show that the fraction of red balls in the urn has an a.s. limit.

Extra Credit Homework XC5: Identify the limit distribution in Homework HW14.

Extra Credit Homework XC6: Give a martingale $\langle X_n \rangle$ such that $\sup E\big[|X_n|\big] < \infty$, for all k, $E[X_k] = 0 = \lim X_n$ a.s., but $E\big[|X_n|\big] \not\to 0$.

Recall Example 35.2: On $(\Omega, \mathscr{F})$, let P be a probability measure and ν be a finite, signed measure. Let $\langle \mathscr{F}_n \rangle$ be a filtration. Suppose that $\nu{\restriction}\mathscr{F}_n \ll P{\restriction}\mathscr{F}_n$ for all n. Let $X_n := \mathrm{d}(\nu{\restriction}\mathscr{F}_n)/\mathrm{d}(P{\restriction}\mathscr{F}_n)$. We saw that $\langle (X_n, \mathscr{F}_n) \rangle$ is a martingale. Let $X := \limsup X_n$. Write $\mathscr{F}_\infty := \bigvee_{n \ge 1} \mathscr{F}_n$.

Theorem 35.7. *Adopt the notation of Example 35.2.*
(i) *If $\nu{\restriction}\mathscr{F}_\infty \ll P{\restriction}\mathscr{F}_\infty$, then $X = \mathrm{d}(\nu{\restriction}\mathscr{F}_\infty)/\mathrm{d}(P{\restriction}\mathscr{F}_\infty)$ P-a.s.*
(ii) *If $\nu{\restriction}\mathscr{F}_\infty \perp P{\restriction}\mathscr{F}_\infty$, then $P[X = 0] = 1$.*

Proof. (i) Write $Z := \mathrm{d}(\nu{\restriction}\mathscr{F}_\infty)/\mathrm{d}(P{\restriction}\mathscr{F}_\infty)$. Checking the definition verifies that $X_n = E[Z \mid \mathscr{F}_n]$ a.s., whence $X = Z$ a.s. by Theorem 35.6.

(ii) Consider the probability measure $Q := (|\nu| + P)/(\|\nu\| + 1)$. Then $\nu{\restriction}\mathscr{F}_\infty \ll Q{\restriction}\mathscr{F}_\infty$, whence the result follows from (i). ∎

Also recall Example 35.3, the specialization to P being Lebesgue measure on $(0,1]$, $\mathscr{F}_n := \sigma(I_k^{(n)}\,;\ 0 \le k < 2^n)$, where $I_k^{(n)} := (k2^{-n}, (k+1)2^{-n}]$. We always have that $\nu\restriction\mathscr{F}_n \ll P\restriction\mathscr{F}_n$ and

$$X_n = \sum_{k=0}^{2^n-1} \mathbf{1}_{I_k^{(n)}} \nu(I_k^{(n)}) 2^n.$$

We get that $X_n \to F'$ P-a.s., where $F(\omega) := \nu\big((0,\omega]\big)$ is function of bounded variation: see Example 35.10 on p. 471.

Theorem 35.3 (Doob's Maximal Inequality). *If $\langle (X_1, \mathscr{F}_1), \ldots, (X_n, \mathscr{F}_n)\rangle$ is a submartingale and $\alpha > 0$, then*

$$P[\max_{i\le n} X_i \ge \alpha] \le \alpha^{-1} E[X_n\,;\ \max_{i\le n} X_i \ge \alpha] \le \alpha^{-1} E\big[|X_n|\big].$$

This extends Kolmogorov's inequality (Theorem 22.4), since if S_i are the partial sums of independent, mean-0 random variables, then $\langle S_i^2\rangle$ is a submartingale.

Proof. The idea is similar to the discussion ending Example N6: Suppose that we must choose exactly one X_k via a stopping time, our goal being to maximize its expectation. According to Proposition N9, we cannot do better than choosing X_n. Nevertheless, we consider the strategy to choose X_τ, where $\tau := \min\{i \le n\,;\ X_i \ge \alpha \text{ or } i = n\}$. Write $M_k := \max_{i\le k} X_i$. Then for all k,

$$[M_n \ge \alpha] \cap [\tau \le k] = [M_k \ge \alpha] \in \mathscr{F}_k,$$

i.e., $[M_n \ge \alpha] \in \mathscr{F}_\tau$. Thus, (N2) gives

$$\alpha P[M_n \ge \alpha] \le E[X_\tau\,;\ M_n \ge \alpha] \le E[X_n\,;\ M_n \ge \alpha] \le E\big[|X_n|\big]. \qquad \blacksquare$$

Homework HW15: Show that if τ is a finite stopping time and $\langle (X_{\tau\wedge n}, \mathscr{F}_n)\rangle$ is a UI submartingale, then for all $\alpha > 0$,

$$P\big[\sup_n X_{\tau\wedge n} \ge \alpha\big] \le \alpha^{-1} E[X_\tau\,;\ \sup_n X_{\tau\wedge n} \ge \alpha] \le \alpha^{-1} E\big[|X_\tau|\big].$$

Extra Credit Homework XC7: Show that if $\langle X_1, \ldots, X_n\rangle$ is a supermartingale and $\alpha > 0$, then

$$\alpha P[\max_{i\le n} X_i \ge \alpha] \le E[X_1] + E\big[|X_n|\big].$$

Doob's L^p Inequality. *If $\langle X_1, \ldots, X_n\rangle$ is a nonnegative submartingale and $p > 1$, then*

$$\|\max_{i\le n} X_i\|_p \le \frac{p}{p-1}\|X_n\|_p.$$

Proof. Because of Theorem 35.3, it suffices to show that for any random variables $X, Y \ge 0$, if

$$\forall \alpha > 0 \quad P[Y \ge \alpha] \le \alpha^{-1} E[X\,;\ Y \ge \alpha], \tag{N3}$$

then

$$\forall p > 1 \quad \|Y\|_p \le \frac{p}{p-1}\|X\|_p.$$

Namely, apply this inequality to $Y := \max_{i\le n} X_i$ and $X := X_n$.

Now

$$\begin{aligned}\|Y\|_p^p = E[Y^p] &= \int_0^\infty P[Y^p > s]\,\mathrm{d}s = \int_0^\infty P[Y > s^{1/p}]\,\mathrm{d}s \\ &= \int_0^\infty pt^{p-1}\,P[Y > t]\,\mathrm{d}t \le \int_0^\infty pt^{p-2}\int_\Omega X\mathbf{1}_{[Y\ge t]}\,\mathrm{d}P\,\mathrm{d}t \\ &= \int_\Omega pX\int_0^\infty t^{p-2}\mathbf{1}_{[Y\ge t]}\,\mathrm{d}t\,\mathrm{d}P = \int_\Omega \frac{p}{p-1}XY^{p-1}\,\mathrm{d}P \\ &\le \frac{p}{p-1}\|X\|_p\|Y^{p-1}\|_q,\end{aligned}$$

where $q^{-1} = 1 - p^{-1} = (p-1)/p$. That is,

$$E[Y^p] \le \frac{p}{p-1}\,E[Y^p]^{1/q}\|X\|_p.$$

If $\|Y\|_p < \infty$, then this gives the desired inequality. If not, then note that we may replace Y by $Y\wedge K$ in the hypothesis: $[Y\wedge K \ge \alpha] = [Y\ge\alpha]$ for $\alpha \le K$ and $[Y\wedge K\ge\alpha] = \varnothing$ for $\alpha > K$. This gives

$$\|Y\wedge K\|_p \le \frac{p}{p-1}\|X\|_p,$$

and now we may let $K\to\infty$. ∎

Doob's L^p Convergence Theorem. *If $\langle X_n\rangle$ is a martingale or nonnegative submartingale, $p > 1$, and $\sup_n E\big[|X_n|^p\big] < \infty$, then $\exists X$ such that $X_n \to X$ a.s. and in L^p.*

Proof. Since $\|X_n\|_1 \le \|X_n\|_p$, Theorem 35.5 gives $X_n \to X$ a.s. Since $|X_n|$ is a submartingale, Doob's L^p inequality and the MCT (or Fatou's lemma) give $\sup|X_n| \in L^p$. Since $|X_n - X|^p \le 2^p\sup|X_n|^p \in L^1$, the LDCT gives $X_n \to X$ in L^p. ∎

Remark N18. Problem 35.4 gave a counterexample for $p = 1$. Also, doubling bets on fair games gives a martingale counterexample.

On the other hand, Doob also showed that if $\langle X_1,\ldots,X_n\rangle$ is a nonnegative submartingale, then

$$E\big[\max_{i\le n} X_i\big] \le \frac{\mathrm{e}}{\mathrm{e}-1}\big(1 + E[X_n\log^+ X_n]\big). \tag{N4}$$

This follows from the following consequence of (N3):

$$E[Y] \le \frac{\mathrm{e}}{\mathrm{e}-1}\big(1 + E[X\log^+ X]\big).$$

By replacing X_i with $X_i/E[X_n]$, we may improve (N4) to

$$E\big[\max_{i\le n} X_i\big] \le \frac{\mathrm{e}}{\mathrm{e}-1}\,E\big[X_n\big(1 + \log^+(X_n/E[X_n])\big)\big]. \tag{N5}$$

(N4) can be further improved by changing $\log^+$ to log, and thus similarly for (N5): see "On pathwise counterparts of Doob's maximal inequalities" by Gushchin (2014). Note that if $\langle X_n\rangle$ is a submartingale and $\sup_n E\big[|X_n|\log^+|X_n|\big] < \infty$, then Doob's martingale convergence theorem and UI already guarantees that $\langle X_n\rangle$ converges a.s. and in L^1.

Homework HW16: Let $p > 1$, A_n be independent with $P(A_n) = \left(\frac{n}{n+1}\right)^p$ for $n \ge 1$, and $X_n := n\mathbf{1}_{\bigcap_{k=1}^{n-1} A_k}$. Show that $\langle X_n \rangle$ is a supermartingale and $E[X_n^p] = 1$, but $X_n \to 0$ a.s. Thus, Doob's L^p convergence theorem does not extend to nonnegative supermartingales.

A sub/super/martingale indexed by the negative integers is called a ***backwards*** or ***reversed*** sub/super/martingale. This adjective refers only to the index set! A filtration can be defined with respect to any poset.

Example N19. Let $\langle Y_n \rangle$ be IID in $L^1(P)$, $S_n := \sum_{k=1}^n Y_k$, $X_{-n} := S_n/n$, and $\mathscr{F}_{-n} := \sigma(S_n, S_{n+1}, \ldots)$. Then for $1 \le k \le n$, $E[Y_1 \mid \mathscr{F}_{-n}] = E[Y_k \mid \mathscr{F}_{-n}]$ a.s. by symmetry.

To prove this rigorously, we note that both sides here depend only on the joint distribution of Y_1 (or Y_k) and $\langle S_n, S_{n+1}, \ldots \rangle$: see Proposition N20 below. So assume that P is a product measure, the law of $\langle Y_n \rangle$ on $\mathbb{R}^\infty$. Then there is a bijection of the probability space $(\mathbb{R}^\infty, \mathcal{R}^\infty, P)$ interchanging Y_1 and Y_k but preserving P and S_m for all $m \ge n$. This proves the equality.

Since $\sum_{k=1}^n E[Y_k \mid \mathscr{F}_{-n}] = E[S_n \mid \mathscr{F}_{-n}] = S_n$ a.s., it follows that $E[Y_1 \mid \mathscr{F}_{-n}] = S_n/n = X_{-n}$ a.s. Thus, $\langle (X_{-n}, \mathscr{F}_{-n}) \rangle$ is a UI reversed martingale.

Proposition N20. *Let $X\colon (\Omega, \mathscr{F}, P) \to (\Omega', \mathscr{F}')$ and $Y\colon (\Omega, \mathscr{F}, P) \to (\Omega'', \mathscr{F}'')$ be random variables and $f\colon (\Omega', \mathscr{F}') \to \mathbb{R}$ be Borel with $f(X) \in L^1(P)$. Then $E[f(X) \mid Y]$ depends only on f and the law of (X, Y). More precisely, for every probability measure μ on $\mathscr{F}' \times \mathscr{F}''$ such that $\int |f(x)| \,\mathrm{d}\mu(x,y) < \infty$, there is a $g_\mu\colon (\Omega'', \mathscr{F}'') \to (\mathbb{R}, \mathcal{R})$ such that $g_{(X,Y)_* P}(Y)$ is a version of $E[f(X) \mid Y]$.*

Proof. Write $\bar{f}(x,y) := f(x)$ and $\bar{\mathscr{F}}'' := \{\Omega' \times A\,;\ A \in \mathscr{F}''\}$. Set g_μ to be the Radon–Nikodym derivative of $(\bar{f}\mu)\restriction\bar{\mathscr{F}}''$ with respect to $\mu\restriction\bar{\mathscr{F}}''$. Since $g_\mu \in \bar{\mathscr{F}}''$, we may also consider g_μ to be an $\mathscr{F}''$-measurable function on Ω''. A check of the definition of conditional expectation shows that this works: writing $\nu := (X,Y)_* P$, we obtain that for all $B \in \mathscr{F}''$,

$$\int_{[Y \in B]} g_\nu(Y) \,\mathrm{d}P = \int_{[y \in B]} g_\nu(y) \,\mathrm{d}\nu(x,y) = \int_{[y \in B]} f(x) \,\mathrm{d}\nu(x,y) = \int_{[Y \in B]} f(X) \,\mathrm{d}P. \quad \blacksquare$$

Note: every reversed martingale is UI. This automatically gives us better properties than ordinary martingales have.

Write $\mathscr{F}_{-\infty} := \bigcap_n \mathscr{F}_{-n}$.

Theorem 35.8. *Let $\langle (X_{-n}, \mathscr{F}_{-n}) \rangle$ be a reversed submartingale. Then its limit, $X_{-\infty} := \lim_{n\to\infty} X_{-n}$, exists a.s. in $[-\infty, \infty)$ and $X_{-\infty} \le E[X_{-1} \mid \mathscr{F}_{-\infty}]$ a.s. In case $\langle (X_{-n}, \mathscr{F}_{-n}) \rangle$ is a reversed martingale, then $X_{-\infty} \in (-\infty, \infty)$ a.s. and $X_{-\infty} = E[X_{-1} \mid \mathscr{F}_{-\infty}]$ a.s.*

Proof. Given $[\alpha, \beta]$, let U_n be the number of upcrossings of $[\alpha, \beta]$ by $\langle X_{-n}, X_{-n+1}, \ldots, X_{-1} \rangle$. Then the upcrossing inequality gives

$$E[U_n] \le \frac{E\big[(X_{-1} - \alpha)^+\big]}{\beta - \alpha}.$$

Hence $\sup U_n < \infty$ a.s., so $X_{-\infty} := \lim X_{-n}$ exists a.s. in $[-\infty, \infty]$.

Given $a < 0$, $\langle (X_{-n} \vee a, \mathscr{F}_{-n}) \rangle$ is a reversed submartingale. Thus,

$$\forall A \in \mathscr{F}_{-\infty} \ \ \forall n \geq 1 \ \ E[X_{-n} \vee a \, ; \ A] \leq E[X_{-1} \vee a \, ; \ A].$$

Taking $n \to \infty$ gives $E[X_{-\infty} \, ; \ A] \leq E[X_{-\infty} \vee a \, ; \ A] \leq E[X_{-1} \vee a \, ; \ A]$ by Fatou's lemma. The rest then follows by taking $a \to -\infty$ using integrability of X_{-1}. ∎

Theorem 35.9. *If $Z \in L^1(P)$ and $\langle \mathscr{F}_{-n} \rangle$ is a filtration, then $E[Z \mid \mathscr{F}_{-n}] \to E[Z \mid \mathscr{F}_{-\infty}]$ a.s.*

Proof. $\langle (E[Z \mid \mathscr{F}_{-n}], \mathscr{F}_{-n}) \rangle$ is a reversed martingale, so its limit, $Z_{-\infty}$, satisfies

$$Z_{-\infty} = E\big[E[Z \mid \mathscr{F}_{-1}] \bigm| \mathscr{F}_{-\infty}\big] = E[Z \mid \mathscr{F}_{-\infty}] \text{ a.s.}$$ ∎

Example N21. With Example N19, we have

$$S_n/n \to X_{-\infty} = E[X_{-1} \mid \mathscr{F}_{-\infty}] = E[Y_1 \mid \mathscr{F}_{-\infty}]$$

a.s. and in L^1. On the other hand, by Kolmogorov's 0-1 law, $\lim S_n/n$ is a constant a.s. Hence $X_{-\infty} = E[X_{-\infty}] = E\big[E[Y_1 \mid \mathscr{F}_{-\infty}]\big] = E[Y_1]$ a.s. This gives a new derivation of the SLLN. While it doesn't give Etemadi's version that assumed only pairwise independence, it doesn't need full independence either—exchangeability is enough, though we don't then get that the limit is constant a.s.

Here, we call $\langle X_1, \ldots, X_n \rangle$ ***exchangeable*** if for all permutations π of $\langle 1, 2, \ldots, n \rangle$, $\langle X_{\pi(1)}, \ldots, X_{\pi(n)} \rangle$ has the same distribution. We call $\langle X_i \, ; \ i \geq 1 \rangle$ ***exchangeable*** if for all n, $\langle X_1, \ldots, X_n \rangle$ is exchangeable.

Suppose that all X_i are indicators (i.e., Bernoulli random variables). It is not hard to give examples of exchangeable $\langle X_1, \ldots, X_n \rangle$; for example, pick k uniformly out of n to be 1. In fact, every exchangeable sequence $\langle X_1, \ldots, X_n \rangle$ of indicators is a mixture of these examples, because if $S_n := \sum_{k=1}^n X_k$, then the law of $\langle X_1, \ldots, X_n \rangle$ conditional on S_n is the uniform measure on S_n out of n. The analogue for infinite exchangeable sequences is that every example is a mixture of Bernoulli trials:

Theorem 35.10 (de Finetti's Theorem). *Let $\langle X_i \, ; \ i \geq 1 \rangle$ be an exchangeable sequence of indicators, $S_n := \sum_{k=1}^n X_k$, and $\Theta := \limsup_{n \to \infty} S_n/n$. Then given Θ, a.s. $\langle X_i \rangle$ has the distribution of Bernoulli trials with success probability Θ, i.e., for all m and $u_i \in \{0, 1\}$, $P[X_1 = u_1, \ldots, X_m = u_m \mid \Theta] = \Theta^k (1 - \Theta)^{m-k}$ a.s., where $k := \sum_{i=1}^m u_i$.*

Proof. This is just a limit of the finite case in the same way that sampling with replacement is a limit of sampling without replacement: Define $\mathscr{F}_{-n} := \sigma(S_n, S_{n+1}, \ldots)$. Fix $m \geq 1$, and consider $n \geq m$. Because $\langle X_1, \ldots, X_n \rangle$ is exchangeable, so is its law conditional on $\mathscr{F}_{-n}$ by Proposition N20. Thus, given $\mathscr{F}_{-n}$, all ways to choose S_n of the n indicators to be 1 are equally likely. Given $u_i \in \{0, 1\}$, write $A := [X_1 = u_1, \ldots, X_m = u_m]$. Then with $k := \sum_{i=1}^m u_i$, there are $\binom{n-m}{S_n - k}$ values in $\{0, 1\}^n$ of $\langle X_1, \ldots, X_n \rangle$ where A occurs, so that a.s.

$$\begin{aligned} P(A \mid \mathscr{F}_{-n}) &= \binom{n-m}{S_n - k} \Big/ \binom{n}{S_n} \\ &= \frac{S_n}{n} \cdot \frac{S_n - 1}{n - 1} \cdots \frac{S_n - k + 1}{n - k + 1} \cdot \frac{n - S_n}{n - k} \cdot \frac{n - S_n - 1}{n - k - 1} \cdots \frac{n - S_n - m + k + 1}{n - m + 1} \\ &\to \Theta^k (1 - \Theta)^{m-k} \end{aligned}$$

a.s. as $n \to \infty$ by Example N21. By Theorem 35.9, we obtain $P(A \mid \mathscr{F}_{-\infty}) = \Theta^k(1-\Theta)^{m-k}$ a.s. Because this is Θ-measurable, the result follows by the tower property, Theorem 34.4. ∎

Homework HW17: (Pólya's urn) An urn contains r red balls and g green balls. Choose one ball at random and replace it together with another of the same color. Repeat indefinitely. Let X_n be the indicator that the nth ball is red. Show that $\langle X_n \rangle$ is exchangeable.

Extra Credit Homework XC8: Identify the distribution of the random variable Θ that gives the mixture of de Finetti's theorem for Homework HW17.

Extra Credit Homework XC9: With the notation of the proof of de Finetti's theorem, show that $\mathscr{F}_{-\infty} = \sigma(\Theta)$ up to sets of probability 0.

Homework HW18: Let $\langle X_i\,;\ i \ge 1 \rangle$ be simple random variables, each taking values in a finite set, U.

(a) Suppose that $\langle X_1, \ldots, X_n \rangle$ is exchangeable. Define $Z_n(u) := |\{i \le n\,;\ X_i = u\}|$. Describe the law of $\langle X_1, \ldots, X_n \rangle$ conditional on $Z_n := \langle Z_n(u)\,;\ u \in U \rangle$.

(b) Suppose that $\langle X_i\,;\ i \ge 1 \rangle$ is exchangeable. Show that $Z := \lim_{n\to\infty} Z_n/n$ exists a.s. *Hint:* Let $\mathscr{F}_{-n} := \sigma(Z_n, X_{n+1}, X_{n+2}, \ldots)$.

(c) With the assumptions and notation of part (b), show that given Z, a.s. $\langle X_i \rangle$ is IID, namely, for all m and $u_i \in U$, $P[X_1 = u_1, \ldots, X_m = u_m \mid Z] = \prod_{i=1}^m Z(u_i)$ a.s.

You may omit pp. 474–478.