Introduction to Martingales

Robert L. Wolpert
Department of Statistical Science
Duke University, Durham, NC, USA

Informally a *martingale* is simply a family of random variables (or a *stochastic process*) $\{M_t\}$ defined on some probability space $(\Omega, \mathcal{F}, \mathsf{P})$ and indexed by some ordered set $\mathcal{T}$ that is "conditionally constant," *i.e.*, whose predicted value at any future time $s > t$ is the same as its present value at the time t of prediction. The set $\mathcal{T}$ of possible indices $t \in \mathcal{T}$ is usually taken to be the nonnegative integers $\mathbb{N}_0$ or the nonnegative reals $\mathbb{R}_+$, although sometimes $\mathbb{Z}$ or $\mathbb{R}$ or other ordered sets arise. Formally we represent what is known at time t in the form of an increasing family of σ-algebras (or a *filtration*) $\{\mathcal{F}_t\} \subset \mathcal{F}$, possibly generated by some process $\{X_s : s \le t\}$ or even by the martingale itself, $\mathcal{F}_t^M = \sigma\{M_s : s \le t\}$ (this one is called the *natural filtration*). We require that $\mathsf{E}|M_t| < \infty$ for each t (so the conditional expectation below is well-defined) and that

$$M_t = \mathsf{E}[M_s \mid \mathcal{F}_t], \quad t < s.$$

It follows that $\{M_t\}$ is *adapted* to $\{\mathcal{F}_t\}$, *i.e.*, M_t is $\mathcal{F}_t$-measurable for each t. For integer-time processes, like functions of random walks or Markov chains, it is only necessary (by the tower property) to take $s = t + 1$. Usually we take $\mathcal{F}_t = \sigma[X_i : i \le t]$ for some process of interest X_t (perhaps M_t itself, although in general $\mathcal{F}_t$ can be bigger than that) and write

$$M_t = \mathsf{E}[M_{t+1} \mid X_0, ..., X_t].$$

There are several "big theorems" about martingales that make them useful in statistics and probability theory. Most of them are simple to prove for discrete time $\mathcal{T} = \mathbb{N}_0$, and true but more challenging for continuous time $\mathcal{T} = \mathbb{R}_+$, so our text (Resnick, 1998, chap. 10) covers only integer-time martingales.

1 Optional Stopping Theorem

A random "time" $\tau : \Omega \to \mathcal{T}$ is an $\mathcal{F}_t$-*stopping time* or a *Markov time* if for each $t \in \mathcal{T}$ the event $[\tau \le t]$ is in $\mathcal{F}_t$; informally, τ "doesn't depend on the future." For discrete time sets $\mathcal{T}$, τ is a stopping time if and only if $[\tau = t] \in \mathcal{F}_t$ for each $t \in \mathcal{T}$ (can you prove that?).

If τ is a stopping time and if M_t is a martingale, then $M_{t\wedge\tau}$ is a martingale too. The proof is easy for integer-time martingales:

$$\begin{aligned}
\mathsf{E}[M_{(t+1)\wedge\tau} \mid \mathcal{F}_t] &= \mathsf{E}[M_\tau \mathbf{1}_{[\tau\le t]} + M_{t+1}\mathbf{1}_{[\tau>t]} \mid \mathcal{F}_t] \\
&= M_\tau \mathbf{1}_{[\tau\le t]} + \mathbf{1}_{[\tau>t]}\mathsf{E}[M_{t+1} \mid \mathcal{F}_t] \\
&= M_\tau \mathbf{1}_{[\tau\le t]} + \mathbf{1}_{[\tau>t]} M_t \\
&= M_{t\wedge\tau}.
\end{aligned}$$

1.1 Application: Simple Random Walks

Fix $0 < p < 1$ and let $\{\xi_j\}$ be iid ±1-valued random variables with $\mathsf{P}[\xi_j = 1] = p$ and $\mathsf{P}[\xi_j = -1] = q := (1-p)$ (hence $\mathsf{E}\xi_j = p - q$ and $\mathsf{V}\xi_j = 4pq$). Set $\mathcal{F}_n := \sigma\{\xi_j : \; j \le n\}$, let $x \in \mathbb{Z}$, and set:

$$X_n := x + \sum_{j\le n} \xi_j, \tag{1}$$

a random walk that is either *symmetric* (if $p = \frac12$) or not (if $p \ne \frac12$). Set $\mu := (p-q)$ and consider for $n \in \mathbb{N}_0 = \{0, 1, \dots\}$ the three processes

$$M_n^{(1)} = X_n - \mu n \tag{2a}$$
$$M_n^{(2)} = (X_n - \mu n)^2 - 4pq\, n \tag{2b}$$
$$M_n^{(3)} = (q/p)^{X_n} \tag{2c}$$

Verify that each of these is a martingale by computing $\mathsf{E}[M_{n+1}^{(i)} \mid \mathcal{F}_n] = M_n^{(i)}$ and applying the tower property and induction. For integers $a \le x$ and $b \ge x$, verify that $\tau := \inf\big\{t \ge 0 : \; X_t \notin (a,b)\big\}$ is a stopping time, finite *a.s.* by Borel-Cantelli.

Gambler's Ruin

Starting with a fortune of \$$x$ and repeatedly betting \$1 at even odds at a game where the probabilities of winning and losing are p and $q := (1-p)$, what is the probability of "winning" by reaching a specified goal $b > x$ before losing by falling to a specified limit $a < x$?

Let $W := [\tau < \infty] \cap [X_\tau = b]$ be the event that X_t exits (a,b) to the right, *i.e.*, that $X_t \ge b$ before $X_t \le a$. If $p = \frac12 = q$ (the symmetric case) then $\mu = 0$ and by DCT

$$\begin{aligned}
x = \mathsf{E}[M_0^{(1)}] &= \lim_{t\to\infty} \mathsf{E}[M_{t\wedge\tau}^{(1)}] \\
&= \mathsf{E}[M_\tau^{(1)}] = b\mathsf{P}[W] + a\mathsf{P}[W^c] \\
&= (b-a)\mathsf{P}[W] + a,
\end{aligned}$$

so the probability of winning is

$$\mathsf{P}[W] = \frac{x-a}{b-a}. \tag{3}$$

Thus in a "fair" game the odds of reaching b before falling to a, starting at $x \in (a, b)$, increases linearly from zero at a to one at b. For an un-fair game, *i.e.*, if $p \neq q$, then $(p/q)^b \neq (p/q)^a$ and again by DCT,

$$\begin{aligned}
(q/p)^x = \mathsf{E}[M_0^{(3)}] &= \lim_{t\to\infty} \mathsf{E}[M_{t\wedge\tau}^{(3)}] = \mathsf{E}[M_\tau^{(3)}] \\
&= (q/p)^b \mathsf{P}[W] + (q/p)^a \mathsf{P}[W^c] \\
&= \big[(q/p)^b - (q/p)^a\big]\mathsf{P}[W] + (q/p)^a, \text{ so} \\
\mathsf{P}[W] &= \frac{(q/p)^x - (q/p)^a}{(q/p)^b - (q/p)^a} \\
&= \frac{(p/q)^{b-x} - (p/q)^{b-a}}{1 - (p/q)^{b-a}} \qquad (4) \\
&\approx (p/q)^{b-x} \text{ if } b \gg a \text{ and } p < \tfrac{1}{2} < q.
\end{aligned}$$

For example, for 1:1 bets in US roulette which win with probability $p = 9/19$ and lose with probability $q = 10/19$, the chance of winning by reaching $b = \$100$ before falling to $a = \$0$ with one-dollar bets beginning at $x = \$90$ is $\mathsf{P}[W] = (0.9^{10} - 0.9^{100})/(1 - 0.9^{100}) = 0.34866$, and the chance of reaching \$100 before \$0 starting at $x = \$50$ is $\mathsf{P}[W] = (0.9^{50} - 0.9^{100})/(1 - 0.9^{100}) = 0.00513$, while these would be 90% and 50% in a fair game. It's surprising to most of us what a dramatic difference the seemingly small departure of $p \approx 0.474$ and $q \approx 0.526$ from 0.500 makes.

Martingale $M_t^{(2)}$ can help us find the expected *duration* of a fair game. For $p = \frac{1}{2} = q$, $\mu = 0$ and $4pq = 1$, so

$$\begin{aligned}
x^2 = M_0^{(2)} &= \lim_{t\to\infty} \mathsf{E}[M_{t\wedge\tau}^{(2)}] = \mathsf{E}[M_\tau^{(2)}] \\
&= \mathsf{E}[{X_\tau}^2 - \tau] \\
&= b^2\mathsf{P}[W] + a^2\mathsf{P}[W^c] - \mathsf{E}[\tau] \\
&= \frac{b^2(x-a) + a^2(b-x)}{b-a} - \mathsf{E}[\tau] \\
&= (a+b)x - ab - \mathsf{E}[\tau] \text{ so} \\
\mathsf{E}[\tau] &= (a+b)x - ab - x^2 = (b-x)(x-a). \qquad (5)
\end{aligned}$$

The expected time until $X_t = 100$ or $X_t = 0$ starting at $x = 90$ is 900 turns and starting at $x = 50$ is 2500 turns, or 30 and 83 hours respectively at a typical rate of two turns per minute. For unfair games we can find $\mathsf{E}\tau$ from $M_\tau^{(1)}$:

$$\begin{aligned}
x = M_0^{(1)} &= \lim_{t\to\infty} \mathsf{E}[M_{t\wedge\tau}^{(1)}] = \mathsf{E}[M_\tau^{(1)}] \\
&= \mathsf{E}[X_\tau - \mu\tau] \\
&= \frac{b[(q/p)^x - (q/p)^a] + a[(q/p)^b - (q/p)^x]}{(q/p)^b - (q/p)^a} - \mu\mathsf{E}[\tau], \text{ so} \\
\mathsf{E}\tau &= \frac{(b-x)[(q/p)^x - (q/p)^a] + (a-x)[(q/p)^b - (q/p)^x]}{\mu[(q/p)^b - (q/p)^a]} \\
&= \frac{(b-x)[(p/q)^{b-x} - (p/q)^{b-a}] - (x-a)[1-(p/q)^{b-x}]}{(p-q)[1-(p/q)^{b-a}]} \qquad (6)
\end{aligned}$$

or approximately $\mathsf{E}\tau \approx (x-a)/(q-p)$ for $a \ll b$ and $p < q$. For US roulette, $\mathsf{E}\tau = 1047.5$ for $x = 90$ (with a slim 35% chance of winning) and $\mathsf{E}\tau = 940.258$ for $x = 50$ (with about a 1/200 chance). Larger bets make the game go quicker and improve the chance of winning; for \$10 bets, set $a = 0$, $b = 10$ and try $x = 5$, $x = 9$ to see the probability of winning increase to $\mathsf{P}[W] = 37\%$ or 87% with $\mathsf{E}[\tau] = 24.46$ or 10.17, respectively, much closer to the values 50%, 90% for $\mathsf{P}[W]$ and 25, 10 for $\mathsf{E}\tau$ in a fair game. Even faster (and more favorable) is the optimal strategy of *bold play*, betting $x \wedge (b-x)$ each time; for $x = 50$ this amounts to betting all \$50 at once ($\mathsf{E}[W] = 9/19 = 47.37\%$, $\mathsf{E}\tau = 1$) while for $x = \$90$, $\mathsf{E}[W] = 87.94\%$. Upon taking the limit as $a \to -\infty$ in Eqns (3, 4) we find that $\mathsf{P}[X_t \ge b$ for *any* $t < \infty]$ has probability one if $p \ge \frac{1}{2}$, but for $p < \frac{1}{2}$ the probability is $(p/q)^{b-x} < 1$; thus even an infinitely-rich patron has only a $0.9^{10} = 34.8678\%$ chance of winning \$10 in US roulette with successive \$1 bets. The expected time to reach $b > x$ is infinite for $p \le \frac{1}{2}$, but for $p > \frac{1}{2}$ the expected time is finite, $\mathsf{E}[\tau] = (b-x)/(p-q) < \infty$.

1.1.1 Other Random Walks

More generally we can construct a process X_n as in (1) for any iid $\{\xi_j\} \subset L_2$ and martingales $M_n^{(k)}$ as in (2), with $\mu = \mathsf{E}\xi_j$ in (2a), replacing $4pq$ with $\sigma^2 = \mathsf{V}\xi_j$ in (2b), and replacing (q/p) with e^{t^*} where $t^* \ne 0$ is the solution to $M(t^*) = 1$ for the MGF $M(t)$ of ξ_j ($t^* < 0$ if $\mu > 0$, $t^* > 0$ if $\mu < 0$). Now the probabilities of Eqns (3, 4) and expectations of Eqns (5, 6) will only be approximate, since X_τ won't be *exactly* a or b. Abraham Wald (1945) studied the discrepancy in some detail, motivated by the following important application, the key to modern sequential clinical trials.

1.2 The SPRT Sequential Statistical Test

If iid random variables $\{Y_j\}$ are known to come from one of two possible distributions, with densities (w.r.t. any σ-finite reference measure) f_0 and f_1, the *likelihood ratio* (against the

Null) for the first n observations is

$$\Lambda_n := \prod_{j\le n} \frac{f_1(Y_j)}{f_0(Y_j)}.$$

In Wald's Sequential Probability Ratio Test (SPRT), one observes data sequentially until Λ_n passes an upper boundary $U \in (1,\infty)$ (in which case the null hypothesis $H_0: \ Y_j \overset{\text{iid}}{\sim} f_0(y)\,dy$ is *rejected*) or a lower boundary $L \in (0,1)$ (in which case the test fails to reject H_0). The test has optimality properties (Wald and Wolfowitz, 1948) similar to those of fixed-sample-size likelihood ratio tests (Neyman and Pearson, 1933). The logarithm $X_n = \log\Lambda_n$ is a random walk under both f_0 and f_1, and $\tau := \inf\{n: \ \Lambda_n \notin (L,U)\} = \inf\{n: \ X_n \notin (a := \log L, b := \log U)\}$ is Wald's stopping time, so the results of Section (1.1.1) apply. In addition, Λ_n itself is a martingale under f_0, as is Λ^{-1} under f_1, giving convenient tools for bounding the probability of incorrect hypothesis-test results or the expected duration of a sequential test: the approximate size $\alpha = \mathsf{P}_0[\Lambda_\tau \ge U]$, power $[1-\beta] = \mathsf{P}_1[\Lambda_\tau \ge U]$ are:

$$\alpha \approx (1-L)/(U-L) \qquad\qquad 1-\beta \approx U(1-L)/(U-L)$$

so any desired size and power can be obtained by setting

$$L \approx \beta/(1-\alpha) \qquad\qquad U \approx (1-\beta)/\alpha$$

The approximate expected sample sizes S_0 (under f_0) and S_1 (under f_1) can be found by applying martingale methods to the random walk $X_n := \log\Lambda_n$, whose iid steps have means μ_i (so $(X_n - n\mu_i)$ is a martingale) given by

$$\begin{aligned} \mu_0 &= -\mathcal{K}(f_0 : f_1) & \qquad \mu_1 &= \mathcal{K}(f_1 : f_0), \\ S_0 &\approx \frac{\alpha\log U + (1-\alpha)\log L}{\mu_0} & \qquad S_1 &\approx \frac{(1-\beta)\log U + \beta\log L}{\mu_1} \end{aligned}$$

under distribution $\{Y_j\} \overset{\text{iid}}{\sim} f_i$ for $i = 0,1$ respectively. Here

$$\mathcal{K}(f:g) := \int \log\frac{f(y)}{g(y)}\, f(y)\,dy$$

denotes the Kullback-Leibler divergence from f to g, a measure of the discrepancy between two distributions with pdfs f, g. If f_0 and f_1 are rather similar, then μ_0 and μ_1 will be small and the sample sizes S_0 or S_1 needed to attain small size α and large power $(1-\beta)$ will be large.

A Bayesian with prior $\mathsf{P}[H_0] = \pi_0$ would report posterior probability $\mathsf{P}[H_0 \mid \text{ Data }] = (1 + \frac{\pi_1}{\pi_0}\Lambda_\tau)^{-1}$, or about $\pi_0/(\pi_0 + \pi_1 a)$ if $X_\tau \le a$ and $\pi_0/(\pi_0+\pi_1 b)$ if $X_\tau \ge b$, lending guidance about the selection of a and b. By Doob's maximal inequality, for $0 < \alpha, \beta < 1$ the SPRT with $L = \beta$ and $U = 1/\alpha$ will satisfy $\mathsf{P}[\text{ Reject } H_0 \mid H_0] \le \alpha$ and $\mathsf{P}[\text{ Reject } H_0 \mid H_1] \ge 1-\beta$, the classical Frequentist error bounds.

2 Martingale Path Regularity

If M_t is a martingale and $a < b$ are real numbers, denote by $\nu_{[a,b]}^{(t)}$ the number of "upcrossings" of the interval $[a, b]$ by M_s prior to time t, *i.e.*, the number of times M_s passes from below a to above b at times $0 \le s \le t$. Then:

$$\mathsf{E}\left[\nu_{[a,b]}^{(t)}\right] \le \frac{\mathsf{E}|M_t| + |a|}{b-a}$$

and, in particular, martingale paths don't oscillate infinitely often— they have left and right limits at every point. This is also the key lemma for proving the Martingale Convergence Theorem below. Here's the idea, attributed to both Doob and to Snell:

Set $\beta_0 := 0$ and, for $n \in \mathbb{N}$, define

$$\begin{aligned} \alpha_n &:= \inf\{t > \beta_{n-1} : \ M_t \le a\} \\ \beta_n &:= \inf\{t > \alpha_n : \ M_t \ge b\}, \end{aligned}$$

or infinity if the indicated event never occurs (*i.e.*, we take $\inf\{\emptyset\} = \infty$). Define a process Y_t by

$$Y_t := \sum_{n\in\mathbb{N}} [M_{t\wedge\beta_n} - M_{t\wedge\alpha_n}].$$

Starting with the first time α_1 that $M_t \le a$, Y_t accumulates the increments of M_t until the first time β_1 that $M_t \ge b$; the process continues if the martingale $M_t \le a$ again falls below a (at time α_2), and so forth. All the terms vanish for n large enough that $\alpha_n > t$, so there are at most $1 + \nu_{[a,b]}^{(t)}$ non-zero terms, each at least $[b - a]$ except possibly the last if $\alpha_n < t < \beta_n$ for some n. Then

$$\begin{aligned} Y_t &:= \sum_{n\in\mathbb{N}} [M_{t\wedge\beta_n} - M_{t\wedge\alpha_n}] \\ &\ge (b-a)\nu_{[a,b]}^{(t)} + [M_t - a] \\ \mathsf{E}Y_t &\ge (b-a)\mathsf{E}\nu_{[a,b]}^{(t)} + \mathsf{E}[M_t - a] \\ &\ge (b-a)\mathsf{E}\nu_{[a,b]}^{(t)} - \mathsf{E}(M_t - a)_- \\ &\ge (b-a)\mathsf{E}\nu_{[a,b]}^{(t)} - \mathsf{E}|M_t| - |a|. \end{aligned}$$

By the Optional Stopping Theorem, Y_t is a martingale and hence $\mathsf{E}Y_t = \mathsf{E}Y_0 = 0$; it follows that $\mathsf{E}\nu_{[a,b]}^{(t)} \le \big(\mathsf{E}|M_t| + |a|\big)/(b-a)$.

The important conclusion is that $\mathsf{E}\nu_{[a,b]}^{(t)} < \infty$, so $\nu_{[a,b]}^{(t)}$ is almost-surely finite— leading to:

Theorem 1 (Martingale Path Regularity) *Let M_t^0 be a martingale with index set $\mathcal{T} = \mathbb{R}_+$. Then with probability one, M_t^0 has limits from the left and from the right at every point $t \in \mathcal{T}$, and at each t is almost-surely equal to the right-continuous process $M_t := \lim_{s\searrow t} M_s^0$. If the filtration is right-continuous, $\mathcal{F}_t = \cap_{s>t}\mathcal{F}_s$, then M_t is also a martingale.*

If M_t is *uniformly* bounded in L_1, $\mathsf{E}|M_t| \le c < \infty$ for all $t \in \mathcal{T}$, then by Fatou's lemma we can even take $t \to \infty$ so $\mathsf{E}\nu^{(\infty)}_{[a,b]} \le [c + |a|]/(b-a) < \infty$, and the number of times $\nu^{(\infty)}_{[a,b]}$ that M_t *ever* crosses the interval $[a,b]$ is almost-surely finite. This is the key for proving:

3 Martingale Convergence Theorems

Theorem 2 (Martingale Convergence Theorem) *Let M_t be an L_1-bounded martingale (so for some $c \in \mathbb{R}_+$ it satisfies $\mathsf{E}|M_t| \le c$ for all $t \in \mathcal{T}$). Then there exists a random variable $M_\infty \in L_1$ such that $M_t \to M_\infty$ a.s. as $t \to \infty$. If $\{M_t\}$ is Uniformly Integrable (for example, if $(\forall t \in \mathcal{T})\mathsf{E}|M_t|^p \le c_p$ for some $p > 1$ and $c_p > 0$), then also $M_t \to M_\infty$ in L_1.*

Proof. Define $M_\infty := \liminf_{t\to\infty} M_t$ and $M^\infty := \limsup_{t\to\infty} M_t$. Suppose (for contradiction) that $\mathsf{P}[M_\infty = M^\infty] < 1$. Then there exist numbers $a < b$ for which $\mathsf{P}[M_\infty < a < b < M^\infty] > 0$. But $\nu^{(\infty)}_{[a,b]} = \infty$ on this event, contradicting $\mathsf{E}\nu^{(\infty)}_{[a,b]} \le \big(c + |a|\big)/(b-a) < \infty$. The result about UI martingales now follows by the UI convergence theorem. ∎

Corollary 1 *Let M_t be a martingale and τ a stopping time. Then*

$$\mathsf{E}M_0 = \mathsf{E}M_\tau$$

if *either $\{M_t\}$ is uniformly integrable, or if $\mathsf{E}\tau < \infty$ and $|M_s - M_t| \le c|s-t|$ a.s. for some $c < \infty$.*

Proof. Obviously $M_\tau = \lim_{t\to\infty} M_{t\wedge\tau}$ *a.s*; the family $\{M_{t\wedge\tau}\}$ will be UI under either of the stated conditions. ∎

Note that *some* condition is necessary in the Corollary above. The simple symmetric random walk $S_0 = 0$, $S_{n+1} = S_n \pm 1$ (with probability 1/2 each) is a martingale, and the hitting time $\tau := \inf\{t : S_t = 1\}$ is a stopping time that is almost-surely finite, but

$$\mathsf{E}[S_\tau] = 1 \ne 0 = \mathsf{E}[S_0]$$

so the conclusion of Corollary 1 fails. Note that S_n is not UI here, and $|S_s - S_t| \le |s-t|$ is linearly bounded, but $\mathsf{E}\tau = \infty$. For another example, let $X \sim \mathsf{Ge}(\frac{1}{2})$ be a geometric random variable with $\mathsf{P}[X = x] = 2^{-x-1}$ for $x \in \mathbb{N}_0$, and set $M_t := 2^t \mathbf{1}_{\{X \ge t\}}$. Then M_t is a martingale starting at $M_0 = 1$, $\tau = X + 1 = \inf\{t : M_t = 0\}$ is a stopping time with finite expectation $\mathsf{E}[\tau] = 2$, but

$$\mathsf{E}[M_\tau] = 0 \ne 1 = \mathsf{E}[M_0].$$

Again M_t is not UI, and this time $\mathsf{E}\tau < \infty$ but $|M_s - M_t|$ is not bounded linearly in $|s-t|$.

Theorem 3 (Backwards Martingale Convergence Theorem) *Let $\{M_t\}$ be a martingale indexed by $\mathbb{Z}$ or $\mathbb{R}$ (or just the negative half-line $\mathbb{Z}_-$ or $\mathbb{R}_-$). Then, without* any *further conditions, there exists a random variable $M_{-\infty} \in L_1(\Omega, \mathcal{F}, \mathsf{P})$ such that*

$$\lim_{t\to-\infty} M_t = M_{-\infty} \text{ a.s. } \textit{and in } L_1(\Omega, \mathcal{F}, \mathsf{P}).$$

The strong law of large numbers for *i.i.d.* L_1 random variables X_n is a corollary: for $n \in \mathbb{N}$, define $S_n := \sum_{j=1}^n X_j$ and $M_{-n} = \bar{X}_n = S_n/n$. Verify that M_t is a martingale for the filtration $\mathcal{F}_t = \sigma\{M_s : s \le t\}$ (note X_n is $\mathcal{F}_{-n+1}$-measurable but *not* $\mathcal{F}_{-n}$-measurable), and that f$M_{-\infty}$ is in the tail field and hence (by Kolmogorov's 0/1 law) is almost-surely constant. Evidently the constant is μ, so $X_n \to \mu$ *a.s.* as $n \to \infty$. □

4 Martingale Problem for Markov Chains

In Section (1.1) we found a particular function $\phi(x) = (q/p)^x$ which, when evaluated along the random walk X_n, would yield a process $M_n^{(3)} = \phi(X_n)$ that was a martingale. In this section we consider the general question of finding functions $\phi(\cdot)$ for which $\phi(X_t)$ is a martingale for specified *Markov chains* X_t— or, more general still, of how to create martingales from processes of the form $\phi(X_t) - A_t$ for "any" function ϕ.

A discrete time *Markov chain* is a process X_n indexed by the nonnegative integers $n \in \mathcal{T} := \mathbb{N}_0$ and taking values in a discrete state space $\mathcal{S}$ with the property that, for each $n \in \mathcal{T}$, the conditional probability $\mathsf{P}[A \mid \mathcal{F}_n]$ of any "future" event $A \in \mathcal{F}^n := \sigma\{X_t : t \ge n\}$, given the "past" $\mathcal{F}_n := \sigma\{X_t : t \le n\}$, depends only on the "present" X_n— *i.e.*, is $\sigma(X_n)$-measurable. Random walks (like the simple random walk of Section (1.1)) are important examples of Markov chains, but others abound. The distribution of a Markov Chain is determined by the *initial distribution* $p_j^{(0)} = \mathsf{P}[X_0 = j]$ for $j \in \mathcal{S}$ and the *transition matrix* $P_{jk}^{(t)} = \mathsf{P}[X_{t+1} = k \mid X_t = j]$ for all $t \in \mathcal{T}$ and pairs $j, k \in \mathcal{S}$. In the important *stationary* case $P_{jk}^{(t)} = P_{jk}$ doesn't depend on t, so $p_j^{(0)} = \mathsf{P}[X_t = j]$ for every $t \in \mathcal{T}$ and n-step transition probabilities $\mathsf{P}[X_{t+n} = k \mid X_t = j] = P_{jk}^n$ are given by simple matrix powers.

Let X_n be a stationary Markov chain with transition matrix P on a discrete (but not necessarily finite) state space $\mathcal{S}$. Then for $\phi(X_n)$ to be a martingale we need for each $j \in \mathcal{S}$

$$\begin{aligned} 0 &= \mathsf{E}[\phi(X_1) - \phi(X_0) \mid X_0 = j] \\ &= \mathcal{A}\phi(j) := \sum_{k \ne j} P_{jk}[\phi(k) - \phi(j)], \end{aligned}$$

for the operator $\mathcal{A}$ called the *generator* of the process. In this case ϕ is said to be *harmonic*. Even if ϕ is *not* harmonic, we can still construct a martingale by subtracting precisely the right thing:

$$M_\phi(t) := \phi(X_t) - \sum_{0 \le s < t} \mathcal{A}\phi(X_s)$$

will always be a martingale, starting at $\phi(X_0)$. In fact, this property characterizes the Markov chain X_t completely, and is the modern way of *defining* the Markov process.

4.1 Martingale Problems

In both discrete and continuous time, the most powerful and general way known for constructing Markov processes and exploring their properties is to view them as solutions to a *Martingale Problem.* We describe it for discretely-distributed processes X_t, but similar characterizations apply to Markov processes with continuous marginal distributions.

4.2 Discrete Time

Let $P^{(t)}_{jk}$ be a (possibly time-dependent) Markov transition matrix on a state space $\mathcal{S}$ indexed by $\mathcal{T} = \mathbb{N}_0$ or $\mathcal{T} = \mathbb{Z}$, so $(\forall j, k \in \mathcal{S})$ and $(\forall t \in \mathcal{T})$,

$$P^{(t)}_{jk} \geq 0 \qquad \text{and} \qquad \sum_{k\in\mathcal{S}} P^{(t)}_{jk} = 1.$$

Then an $\mathcal{S}$-valued process X_t indexed by $t \in \mathcal{T}$ is a Markov chain with transition matrix $P^{(t)}_{jk}$ if and only if it solves the discrete-time Martingale Problem: for all bounded functions $\phi : \mathcal{S} \to \mathbb{R}$, the process

$$M_\phi(t) := \phi(X_t) - \phi(X_0) - \sum_{0\leq s<t} \sum_{j\neq i=X_s} P^{(s)}_{ij} \left[\phi(j) - \phi(i)\right]$$

must be a martingale indexed by $t \in \mathcal{T}$. In the homogeneous case where $P^{(t)}_{jk} \equiv P_{jk}$ doesn't depend on t, the n-step transition probability is simply the matrix power P^n, and the operator

$$\mathrm{G}\phi(i) = \sum_{j\neq i} P_{ij}[\phi(j) - \phi(i)]$$

is called the *generator* of the process. The function ϕ is called *harmonic* if $\mathrm{G}\phi \equiv 0$, in which case $\phi(X_t)$ itself is a martingale.

4.2.1 Example: Simple Random Walks

For the *symmetric* random walk on $\mathbb{Z}$, for example, $\mathrm{G}\phi(x) = \frac{1}{2}[\phi(x+1) - 2\phi(x) + \phi(x-1)]$, half the second-difference operator, so all affine functions $\phi(x) = a + bx$ (and only they) are harmonic. Now we'll consider asymmetric walks.

Let X_t be the simple random walk (1) starting at $X_0 = x$ with $\mathsf{P}[\xi_j = 1] = p$ and $\mathsf{P}[\xi_j = -1] = q := (1-p)$ with $0 < p < 1$. To be harmonic a function ϕ must satisfy $0 \equiv \mathcal{A}\phi(x) = p[\phi(x+1) - \phi(x)] - q[\phi(x) - \phi(x-1)]$, so by induction $[\phi(x) - \phi(x-1)] = (q/p)^x[\phi(1) - \phi(0)]$. Summing the geometric series shows that all solutions are of the form $\phi(x) = a + b(q/p)^x$ for $p \neq q$, and (as before) $\phi(x) = a + bx$ for $p = q = \frac{1}{2}$.

This and the martingale maximal inequality lead to simple proofs of things about the random walk— for example, if $p < q$ (so X_t is more likely to decrease than increase) and $a > x$, then for $t > 0$,

$$\begin{aligned}\mathsf{P}[\sup_{0\le s\le t} X_s \ge a] &= \mathsf{P}[\sup_{0\le s\le t} (q/p)^{X_s} \ge (q/p)^a] \\ &\le \frac{(q/p)^x}{(q/p)^a} = (p/q)^{a-x}.\end{aligned}$$

Taking the supremum over all $t > 0$ (since the bound doesn't depend on t), we see that the probability of *ever* exceeding a decreases geometrically. With a little more work, we can find exceedence probabilities for lines $a + bt$ too:

Let $b \in \mathbb{R}$ and set $Y_t := X_t - bt$ where X_t is the simple random walk of Section (1.1). Then Y too is a Markov chain, and the function $\phi(x) = r^x$ will be harmonic for Y if r satisfies

$$\begin{aligned}0 = \mathcal{A}\phi(x) &= p\phi(x+1-b) - \phi(x) + q\phi(x-1-b) \\ &= r^{x-1-b}[pr^2 - r^{1+b} + q].\end{aligned}$$

The term in brackets

$$h(r) := pr^2 - r^{1+b} + q$$

vanishes at $r = 1$ and tends to infinity as $r \to \pm\infty$. Its derivative at $r = 1$ is $h'(1) = (\mu - b)$ for $\mu = (p - q) = (2p - 1)$; if this doesn't vanish, then there must exist another root $r_* \ne 1$ of $h(r_*) = 0$ for which $\mathcal{A}\phi \equiv 0$ and hence $M_\phi(t) := r_*^{X_t - bt}$ is a martingale starting at $M_\phi(0) = r_*^x$. By the Martingale Maximal Inequality (MMI, Theorem 4 on *p.* 13), for any $a, b \in \mathbb{R}$,

$$\mathsf{P}\left[\sup_{0\le s\le t} \{X_s - bs\} \ge a\right] = \mathsf{P}\left[\sup_{0\le s\le t} \{r_*^{Y_s}\} \ge r_*^a\right] \le r_*^{x-a}, \tag{7}$$

giving a bound for the probability that the random walk X_s *ever* crosses the line $y = a + bs$ (since the bound doesn't depend on $t < \infty$). In the Roulette example, with $p = 9/19$ and $b = 0$ we have $r_* = q/p = 10/9$, so (7) implies

$$\mathsf{P}[X_t \text{ ever exceeds } a] \le (9/10)^{a-x},$$

the same bound as before. Now, however, we have new results like

$$\mathsf{P}[X_t \text{ ever exceeds } (a + t/2)] \le (3.382975)^{x-a}$$

for a symmetric random walk and $a \ge x$, since $r_* \approx 3.382975$ is the solution $r \ne 1$ to $h(r) = [\frac{1}{2}r^2 - r^{3/2} + \frac{1}{2}] = 0$.

4.2.2 General Random Walks

Now let $\{\xi_j\}$ be iid from any distribution with a MGF $M(t) = \mathsf{E}[e^{t\xi_j}]$ that is finite in some interval around zero. Let $X_n := x + \sum_{j\le n} \xi_j$ be a random walk starting at $x \in \mathbb{R}$, and let

$a, b \in \mathbb{R}$. Then for any $t \in \mathbb{R}$ for which $M(t)$ is finite,

$$Y_n := \exp\{tX_n - n \log M(t)\}$$

is a martingale and, for any t_* such that $M(t_*) = e^{t_* b}$, so is

$$Y_n^* := \exp\{t_*(X_n - nb)\}.$$

By the MMI,

$$\begin{aligned}\mathsf{P}[X_n \text{ ever exceeds } a + b\,n] &= \mathsf{P}\left[\sup_{n\geq 0}(X_n - n\,b) \geq a\right] \\ &= \mathsf{P}\left[\sup_{n\geq 0} Y_n^* \geq e^{t^* a}\right] \leq \exp\{t^*(x-a)\}.\end{aligned}$$

For example, if $\xi_j \overset{\text{iid}}{\sim} \mathsf{No}(\mu, \sigma^2)$ then $M(t) = e^{t\mu + t^2\sigma^2/2}$ is finite for all $t \in \mathbb{R}$ and the equation

$$M(t_*) = e^{t_*\mu + t_*^2\sigma^2/2} = e^{t_* b}$$

is satisfied for $t_* = 0$ or $t_* = 2(b-\mu)/\sigma^2$. The first of these gives a trivial bound but the second gives

$$\mathsf{P}[X_n \text{ ever exceeds } a + b\,n] \leq \exp\left\{2(b-\mu)(x-a)/\sigma^2\right\}$$

or, for $x = \mu = 0 < a$, simply $\exp\{-2ab/\sigma^2\}$. This same bound, as it happens, applies to Brownian motion with drift. Exercise: Find a bound for the probability that a unit-rate Poisson random walk X_t ever exceeds $1 + 2t$ (Ans: $\exp(-1.256431) = 0.2846682$).

4.3 Continuous Time

Similar bounds are available for Markov processes indexed by continuous time $\mathcal{T} = \mathbb{R}_+$, such as Brownian motion and continuous-time Markov chains.

Let $Q_{jk}^{(t)}$ be a (possibly time-dependent) continuous-time Markov transition rate matrix on a discrete state space $\mathcal{S}$, *i.e.*, a family of matrices on $\mathcal{S} \times \mathcal{S}$ that for each $t \in \mathcal{T}$ satisfies

$$(\forall j \neq k \in \mathcal{S})\ Q_{jk}^{(t)} \geq 0 \qquad \text{and} \qquad (\forall j \in \mathcal{S})\ \sum_{k\in\mathcal{S}} Q_{jk}^{(t)} = 0.$$

Then an $\mathcal{S}$-valued process X_t is a Markov chain with rate matrix $Q_{jk}^{(t)}$ if and only if it solves the continuous-time Martingale Problem: for all bounded functions $\phi : \mathcal{S} \to \mathbb{R}$, the process

$$M_\phi(t) := \phi(X_t) - \int_0^t \Big[\sum_{j\neq i = X_s} Q_{ij}^{(s)}\,[\phi(j) - \phi(i)]\Big]\,ds$$

must be a martingale starting at $M_\phi(0) = \phi(x)$. In the homogeneous case where $Q^{(t)}_{jk} \equiv Q_{jk}$ doesn't depend on t, the time-t transition probability is simply the matrix exponential $P^t = \exp(t\,Q) = \sum_{n\ge 0} \frac{t^n}{n!}\, Q^n$. The operator

$$\mathrm{G}\phi(i) := \sum_{j\in\mathcal{S}} Q_{ij}[\phi(j) - \phi(i)]$$

is called the *(infinitesimal) generator* of the process, and M_ϕ can be written

$$M_\phi(t) := \phi(X_t) - \int_0^t \mathrm{G}\phi(X_s)\,ds.$$

If ϕ is harmonic, then $\phi(X_t)$ is a martingale. A similar approach works for processes with continuous marginal distribution: for Brownian Motion in $\mathbb{R}^d$, for example, $\mathrm{G}\phi(x) = \frac{1}{2}\Delta\phi(x)$, half the Laplacian, illustrating why functions that satisfy $\mathrm{G}\phi \equiv 0$ are called *harmonic*.

4.3.1 Example: SII Jump Processes

The unit-rate Poisson process $N(t)$ is characterized by its initial value of 0 and its generator $\mathrm{G}\phi(x) = [\phi(x+1) - \phi(x)]$. The sum

$$X_t = \sum_j \; u_j\, N_j(\nu_j\, t)$$

of independent Poisson processes with rates $\nu_j > 0$ and jump sizes $u_j \in \mathbb{R}$ is also a continuous time Markov process, with generator given by

$$\begin{aligned} \mathrm{G}\phi(x) &= \sum_j [\phi(x+u_j) - \phi(x)]\,\nu_j \\ &= \int_{\mathbb{R}} [\phi(x+u) - \phi(x)]\,\nu(du) \end{aligned} \tag{8}$$

for $\phi \in \mathcal{C}_b^1(\mathbb{R})$, for the discrete measure $\nu(du) := \sum_j u_j \delta_{\nu_j}(du)$. The log ch.f. is

$$\log \mathsf{E} e^{i\omega X_t} = \int_{\mathbb{R}} \left[e^{i\omega u} - 1\right]\,\nu(du). \tag{9}$$

Actually Eqns (8, 9) continue to be well-defined and determine the distribution of a Markov process X_t with stationary independent increments (SII) for any finite Borel measure $\nu(du)$ on $\mathbb{R}$ or, since both integrands vanish to first order at zero, even for infinite "Lévy measures" $\nu(du)$ that satisfy the "local L_1 condition"

$$\int_{\mathbb{R}} (1 \wedge |u|)\;\nu(du) < \infty. \tag{10}$$

One example is the *gamma process* $X_t \sim \mathsf{Ga}(\alpha dt, \beta)$ whose Lévy measure is given by $\nu(du) = \alpha\, u^{-1}\, e^{-\beta u}\, \mathbf{1}_{\{u>0\}}\, du$, and whose independent increments

$$[X_t - X_s] \sim \mathsf{Ga}(\alpha(t-s),\ \beta)$$

have gamma distributions. Another is the *symmetric α-stable (SαS) process* $X_t \sim \mathsf{St}(\alpha, 0, \gamma t, 0)$ with $\nu(du) = \frac{\alpha\gamma}{\pi}\,\Gamma(\alpha)\sin(\frac{\pi\alpha}{2})\,|u|^{-\alpha-1}\,du$, with α-stable increments. Eqn (10) is only satisfied for $0 < \alpha < 1$, but the approach can be extended to cover the entire range of $0 < \alpha < 2$ (including the Cauchy, $\alpha = 1$) using "compensation". Ask me if you'd like to know more.

5 Maximal Inequalities

Under mild conditions, the suprema of martingales over finite and even infinite intervals may be bounded; this makes them extremely useful for bounding the growth of processes. The usual bounds are of two kinds: bounds on the probability that a martingale M_t (or its absolute value $|M_t|$) exceeds a fixed number $\lambda \in \mathbb{R}$ in some prescribed time interval, and bounds on the expectation of the supremum of $|M_t|^p$ over some interval, for real numbers $p \geq 1$. Here are a few illustrative results.

Theorem 4 *Let M_t be a martingale and let $t \in \mathcal{T}$. Then for any $\lambda > 0$,*

$$\mathsf{P}\left[\sup_{0\le s\le t} M_s \ge \lambda\right] \le \lambda^{-1}\mathsf{E}M_t^+$$
$$\mathsf{P}\left[\sup_{0\le s\le t} |M_s| \ge \lambda\right] \le \lambda^{-1}\mathsf{E}|M_t|$$

Proof. Let $\tau = \inf\{t \geq 0 :\ M_t \geq \lambda\}$. Since both M_t and $M_{t\wedge\tau}$ are martingales,

$$\begin{aligned}
\mathsf{E}M_t &= \mathsf{E}M_{t\wedge\tau} \\
&= \mathsf{E}\left\{M_\tau \mathbf{1}_{[\tau\le t]} + M_t \mathbf{1}_{[\tau>t]}\right\} \\
&\ge \mathsf{E}\left\{\lambda \mathbf{1}_{[\tau\le t]} + M_t \mathbf{1}_{[\tau>t]}\right\} \\
&= \lambda\mathsf{P}[\tau \le t] + \mathsf{E}\left\{M_t \mathbf{1}_{[\tau>t]}\right\}, \quad \text{so} \\
\mathsf{E}[M_t \mathbf{1}_{[\tau\le t]}] &\ge \lambda\mathsf{P}[\tau\le t] \quad \text{and hence} \\
\mathsf{P}\left\{\sup_{0\le s\le t} M_s \ge \lambda\right\} &= \mathsf{P}[\tau \le t] \\
&\le \lambda^{-1}\mathsf{E}[M_t \mathbf{1}_{[\tau\le t]}] \\
&\le \lambda^{-1}\mathsf{E}[M_t^+ \mathbf{1}_{[\tau\le t]}] \\
&\le \lambda^{-1}\mathsf{E}[M_t^+],
\end{aligned}$$

proving the first assertion. Since $-M_t$ is also a martingale, we also have:

$$\mathsf{P}\left\{\inf_{0\le s\le t} M_s \le -\lambda\right\} \le \lambda^{-1}\mathsf{E}[M_t^-]; \quad \text{adding these together,}$$
$$\mathsf{P}\left\{\sup_{0\le s\le t} |M_s| \ge \lambda\right\} \le \lambda^{-1}\mathsf{E}[|M_t|].$$

□

In fact we proved something slightly stronger (which we'll need below). Set $|M|_t^* := \sup_{0\le s\le t}|M_s|$; then

$$\mathsf{P}\left\{|M|_t^* \ge \lambda\right\} \le \lambda^{-1}\mathsf{E}\left[|M_t|\mathbf{1}_{\{|M|_t^*\ge\lambda\}}\right]. \tag{11}$$

Theorem 5 *For any martingale M_t and any real numbers $p>1$ and $q := \frac{p}{p-1} > 1$,*

$$\Big\| \sup_{s\le t} |M_s| \Big\|_p \le q \sup_{s\le t} \|M_s\|_p.$$

Proof.
By Fubini's theorem,

$$\begin{aligned}
\mathsf{E}\big[(|M|_t^*)^p\big] &= \int_0^\infty p\lambda^{p-1}\,\mathsf{P}\big[|M|_t^*\ge\lambda\big]\,d\lambda \\
&\le \int_0^\infty p\lambda^{p-1}\,\lambda^{-1}\mathsf{E}\left[|M_t|\mathbf{1}_{\{|M|_t^*\ge\lambda\}}\right]\,d\lambda \\
&= \mathsf{E}\int_0^{|M|_t^*} p\lambda^{p-2}\,|M_t|\,d\lambda \\
&= \frac{p}{p-1}\mathsf{E}\Big[\big(|M|_t^*\big)^{p-1}|M_t|\Big].
\end{aligned}$$

Hölder's inequality asserts that $\mathsf{E}[Y\,Z] \le \{\mathsf{E}Y^p\}^{1/p}\{\mathsf{E}Z^q\}^{1/q}$ for any nonnegative random variables Y and Z; applying this with $Y=|M_t|$ and $Z=(|M|_t^*)^{p-1}$, and noting $(p-1)q=p$, we get

$$\{\mathsf{E}(|M|_t^*)^p\}^1 \le q\ \mathsf{E}\left\{\big(|M|_t^*)^p\right\}^{1/q}\mathsf{E}\left\{|M_t|^p\right\}^{1/p}$$
$$\{\mathsf{E}(|M|_t^*)^p\}^{1-1/q} = \|\ |M|_t^*\ \|_p \le q\ \|M_t\|_p = q \sup_{0\le s\le t}\|M_s\|_p.$$

□

Note that $q \nearrow \infty$ as $p \searrow 1$, so the bound blows up as p shrinks to one. To achieve an L_1 bound on $\mathsf{E}|M|_t^*$ we need something slightly stronger than an L_1 bound on $\mathsf{E}|M_t|$ (see below).

In summary: if M_t is a martingale and if $t \in \mathcal{T}$ then

$$\begin{aligned}
\mathsf{P}[\sup_{s\le t} M_s \ge \lambda] &\le \lambda^{-1}\mathsf{E}[M_t^+]\\
\mathsf{P}[\min_{s\le t} M_s \le -\lambda] &\le \lambda^{-1}\mathsf{E}[M_t^-]\\
\mathsf{P}[\sup_{s\le t} |M_s| \ge \lambda] &\le \lambda^{-1}\mathsf{E}|M_t|\\
\mathsf{E}\sup_{s\le t} |M_s|^p &\le q^p \sup_{s\le t} \mathsf{E}\big[|M_s|^p\big] = q^p\ \mathsf{E}\big[|M_t|^p\big] \quad (p>1)\\
\mathsf{E}\sup_{s\le t} |M_s| &\le \frac{e}{e-1}\sup_{s\le t} \mathsf{E}\big[|M_s|\log^+(|M_s|)\big] \quad (p=1)
\end{aligned}$$

6 Doob's Martingale

Fix any $Y \in L_1(\Omega, \mathcal{F}, \mathsf{P})$ and let $M_t := \mathsf{E}[Y \mid \mathcal{F}_t]$ be the best prediction of Y available at time t. Then M_t is a uniformly-integrable martingale, and $M_t \to Y$ *a.s.* and in L_1.

7 Summary

To summarize, martingales are important because:

1. They have close connections with Markov processes;
2. Their expectations at stopping times are easy to compute;
3. They offer a tool for bounding the maxima and minima of processes;
4. They offer a tool for establishing path regularity of processes;
5. They offer a tool for establishing the *a.s.* convergence of certain random sequences;
6. They are important for modeling economic and statistical time series which are, in some sense, predictions.

Examples:

1. Partial sums $S_n = \Sigma_{i=1}^n X_i$ of independent mean-zero RV's
2. Stochastic Integrals. For example: let M_n be your "fortune" at time n in a gambling game, and let X_n be an IID Bernoulli sequence with probability $\mathsf{E}X_n = p$. Preceding each time $n+1 \in \mathbb{N}$ you may bet any fraction F_n you like of your (current) fortune M_n on the upcoming Bernoulli event X_{n+1}, at odds $(p : 1-p)$; your new fortune after that bet will be $M_{n+1} = M_n(1-F_n)$ if you lose (*i.e.*, if $X_{n+1} = 0$), and $M_{n+1} = M_n(1+F_n\frac{1-p}{p})$ if you win (*i.e.*, if $X_{n+1} = 1$), or in general $M_{n+1} = M_n(1 - F_n(1 - X_{n+1}/p))$. If

$F_n \in \sigma\{X_1 \cdots X_n\}$ depends only on information available at time n, then $\mathsf{E}[M_{n+1} \mid \mathcal{F}_n] = M_n$ and M_n is a martingale. Hence there is *no possible betting strategy* F_n based only on observed information $\mathcal{F}_n$ that can lead to a positive expected profit, since $\mathsf{E}[M_n - M_0] \equiv 0$. We can represent M_n in the form

$$M_n = M_0 + \sum_{i=0}^{n-1} F_i M_i [Y_{i+1} - Y_i]$$

as the "martingale transform" of the martingale $Y_n := (S_n - np)/p$.

3. Variance of a Sum: $M_n = \left(\sum_{i=1}^n Y_i\right)^2 - n\sigma^2$, where $\mathsf{E} Y_i Y_j = \sigma^2 \delta_{ij}$

4. Radon-Nikodym Derivatives:

$$\begin{aligned} M_n(\omega) &= 2^{-n} \int_{i/2^n}^{(i+1)/2^n} f(x)\,dx, \qquad i = \lfloor 2^n \omega \rfloor \\ &\to M_\infty(\omega) = f(\omega) \quad a.s. \end{aligned}$$

5. Leftovers:

   - Submartingales: $\mathsf{E}[X_t^+] < \infty$, $X_t \in \mathcal{F}_t$, $X_t \le \mathsf{E}[X_s \mid \mathcal{F}_t]$ for $s > t$.
   - Supermartingales: If X_t is a submartingale then $Y_t := (-X_t)$ is a supermartingale, satisfying $Y_t \ge \mathsf{E}[Y_s \mid \mathcal{F}_t]$ for $s > t$.
   - Jensen's inequality: if M_t is a martingale and if ϕ convex with $\mathsf{E}[\phi(M_t)^+] < \infty$, then $X_t = \phi(M_t)$ is a submartingale.
   - Most of the bounds and convergence theorems above extend to sub- or super-martingales.
   - Positive supermartingales always converge: if $Y_t \ge 0$ is a supermartingale, then $(\exists Y_\infty \in L_1)$ $Y_t \to Y$ *a.s.* If $\{Y_t\}$ is UI, also $Y_t \to Y$ in L_1.
   - A martingale is both a submartingale and a supermartingale.

References

Neyman, J. and Pearson, E. (1933), "On the Problem of the Most Efficient Tests of Statistical Hypotheses," *Philosophical Transactions of the Royal Society. Series A*, 231, 289–337.

Resnick, S. I. (1998), *A Probability Path*, Boston, MA: Birkhäuser.

Wald, A. (1945), "Sequential tests of statistical hypotheses," *Annals of Mathematical Statistics*, 16, 117–186.

Wald, A. and Wolfowitz, J. (1948), "Optimal Character of the Sequential Probability Ratio Test," *Annals of Mathematical Statistics*, 19, 326–339.

Last edited: November 15, 2016