Introduction to Martingales

Robert L. Wolpert
Institute of Statistics and Decision Sciences
Duke University, Durham, NC, USA

We've already encountered and used martingales in this course to help study the hitting-times of Markov processes. Informally a martingale is simply a stochastic process M_t defined on some probability space $(\Omega, \mathcal{F}, \mathsf{P})$ 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 $\mathbb{Z}_+$ 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 (or *filtration*) $\{\mathcal{F}_t\} \subset \mathcal{F}$ of σ-algebras, possibly those generated by a process $\{X_s : s \leq t\}$ or even by the martingale itself, $\mathcal{F}_t = \sigma\{M_s : s \leq t\}$, and require that $\mathsf{E}|M_t| < \infty$ for each t (so the conditional expectation below is well-defined) and that

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

In particular, $\{M_t\}$ 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 the Markov chains we looked at before) it is only necessary to take $s = t + 1$, and usually we take $\mathcal{F}_t = \sigma[X_i : i \leq t]$ 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 $\mathcal{T} = \mathbb{Z}_+$, and more challenging for continuous time $\mathcal{T} = \mathbb{R}_+$.

1. Optional Stopping Theorem:

If τ is a *stopping time* or *Markov time*, *i.e.*, a random time that "doesn't depend on the future" (technically the requirement is that the event $[\tau \le t]$ should be in $\mathcal{F}_t$ for each t), and if M_t is a martingale, then $M_{t\wedge\tau}$ is a martingale too. The proof is easy for integer-time martingales:

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

2. Martingale Path Regularity:

If M_t is a martingale and $a < b$ are real numbers, denote by $\nu^{(t)}_{[a,b]}$ the number of "upcrossings" of the interval $[a, b]$ by M_s prior to time t, the number of times it passes from below a to above b; then:

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

and, in particular, martingale paths don't oscillate infinitely often— thus 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 $1 + \nu^{(t)}_{[a,b]}$ non-zero terms. Then

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

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

The important conclusion is that $\mathsf{E}\nu^{(t)}_{[a,b]} < \infty$. If M_t is *uniformly* bounded in L^1, $\mathsf{E}|M_t| \leq c < \infty$ for all $t \in \mathcal{T}$, then by Fatou's lemma we even have $\mathsf{E}\nu^{(\infty)}_{[a,b]} \leq [c + |a|]/(b-a) < \infty$, so the number of times $\nu^{(\infty)}_{[a,b]} < \infty$ that M_t *ever* crosses the interval $[a,b]$ is almost-surely finite— leading to

Theorem 1 (Martingale Path Regularity) *Let M^0_t be a martingale with index set $\mathcal{T} = \mathbb{R}_+$. Then with probability one, M^0_t 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 $M_t \equiv \lim_{s\searrow t} M^0_s$. If the filtration is right-continuous, $\mathcal{F}_t = \cap_{s>t}\mathcal{F}_s$, then M_t is also a martingale.*

The upcrossing lemma is also the key result needed for proving

3. Martingale Convergence Theorems:

Theorem 2 (Martingale Convergence Theorem) *Let M_t be a martingale satisfying $\mathsf{E}|M_t| \leq c < \infty$ 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 $\mathsf{E}|M_t|^p \leq c_p < \infty$ for some $p > 1$), then also $M_t \to M_\infty$ in L^1.*

Proof. Define $M_\infty \equiv \liminf_{t\to\infty} M_t$ and $M^\infty \equiv \limsup_{t\to\infty} M_t$, and suppose (for contradiction) that $\mathsf{P}[M_\infty = M^\infty] < 1$. Then there exist numbers $a < b$ for which $0 < \mathsf{P}[M_\infty < a < b < M^\infty]$. But $\nu^{(\infty)}_{[a,b]} = \infty$ on the event $[M_\infty < a < b < M^\infty]$, contradicting $\mathsf{E}\nu_{[a,b]}^{(\infty)} \le \big(c + |a|\big)/(b-a) < \infty$. 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 $|M_s - M_t| \le c|s-t|$ for some $c < \infty$ and $\mathsf{E}\tau < \infty$.

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 $\tau \equiv \inf\{t : S_t = 1\}$ 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 $x \in \mathbb{Z}_+$, and set $M_t \equiv 2^t 1_{\{X \ge t\}}$; 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*

$$\lim_{t\to-\infty} M_t = M_{-\infty} \text{ a.s. and in } L^1(\Omega, \mathcal{F}, \mathsf{P}).$$

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 $M_{-\infty}$ is in the tail field and hence (by Kolmogorov's 0/1 law) is almost-surely constant, evidently μ, so $X_n \to \mu$ *a.s.* as $n \to \infty$. □

4. Martingale Problem for Continuous-Time Markov Chains:

Let Q^t_{jk} be a (possibly time-dependent) Markov transition matrix on a state space $\mathcal{S}$. Then an $\mathcal{S}$-valued process X_t is a Markov chain with transition matrix $Q_{jk}(t)$ if and only if, for all bounded functions $\phi : \mathcal{S} \to \mathbb{R}$, the process

$$M_\phi(t) = \phi(X_t) - \phi(X_0) - \int_0^t \Big[\sum_{j \ne i = X_s} Q^s_{ij} \left[\phi(j) - \phi(i)\right] \Big] \, ds$$

is a martingale. Similar characterizations apply to discrete-time Markov chains and to continuous-time Markov processes with non-discrete state space $\mathcal{S}$. This is the most powerful and general way known for *constructing* Markov processes.

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 $p \ge 1$. Here are a few illustrative results.

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

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

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

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

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

$$\begin{aligned}
\mathsf{P}\left\{\inf_{0\leq s\leq t} M_s \leq -\lambda\right\} &\leq \lambda^{-1}\mathsf{E}[M_t^-]; \quad \text{adding these together,} \\
\mathsf{P}\left\{\sup_{0\leq s\leq t} |M_s| \geq \lambda\right\} &\leq \lambda^{-1}\mathsf{E}[|M_t|].
\end{aligned}$$

□

In fact we proved something slightly stronger (which we'll need below). Set $M_t^* \equiv \sup_{0\leq s\leq t} |M_s|$; then

$$\mathsf{P}\{M_t^* \geq \lambda\} \leq \lambda^{-1}\mathsf{E}\left[|M_t| 1_{\{M_t^*\geq\lambda\}}\right]. \tag{1}$$

Theorem 5 *For any martingale M_t and any real number $p > 1$,*

$$\mathsf{E}[\sup_{s\leq t} |M_s|^p] \leq \left(\frac{p}{p-1}\right)^p \sup_{s\leq t} \mathsf{E}[|M_s|^p].$$

Proof.

Set $q = p/(p-1)$, the conjugate exponent, so $\frac{1}{p} + \frac{1}{q} = 1$. By Fubini's theorem,

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

Hölder's inequality asserts that $\mathsf{E}[Y\,Z] \le \{\mathsf{E}Y^p\}^{1/p}\{\mathsf{E}Z^q\}^{1/q}$ for any non-negative 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

$$\begin{aligned}
\mathsf{E}(M_t^*)^p &\le \frac{p}{p-1}\mathsf{E}\left\{\big(M_t^*)^p\right\}^{1/q}\mathsf{E}\left\{|M_t|^p\right\}^{1/p} \\
&\le \left(\frac{p}{p-1}\right)^p \mathsf{E}|M_t|^p.
\end{aligned}$$

□

Note that the bound blows up as $p \to 1$; to achieve a bound on $\mathsf{E}M_t^*$ we need something slightly stronger than a bound on $\mathsf{E}|M_t|$ (see below).

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

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

6. Doob's Martingale:

Fix any $Y \in L^1(\Omega, \mathcal{F}, \mathsf{P})$ and set $M_t = \mathsf{E}[Y \mid \mathcal{F}_t]$, the best prediction of Y available at time t. Then M_t is a uniformly-integrable martingale.

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 maximum and minimum 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 $F_n \in \sigma\{X_1 \cdots X_n\}$, then $\mathsf{E}[M_{n+1} \mid \mathcal{F}_n] = M_n$ and M_n is a martingale. Note that

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

   for 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 $\mathsf{E}Y_iY_j = \sigma^2\delta_{ij}$

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: $\mathsf{E}[X_t^+] < \infty$, $\mathsf{E}[X_s \mid \mathcal{F}_t] \geq X_t$, $X_t \in \mathcal{F}_t$.
- 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.
- Supermartingales: If X_t is a submartingale then $Y_t = -X_t$ is a supermartingale, satisfying $\mathsf{E}[Y_s \mid \mathcal{F}_t] \leq Y_t$.
- Most of the bounds and convergence theorems above extend to sub- or super- martingales.