Informally a martingale is simply a stochastic process \( M_t \) defined on some probability space \((\Omega, \mathcal{F}, P)\) and indexed by some ordered set \( T \) 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 \( T \) of possible indices \( t \in T \) 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 of \( \sigma \)-algebras (or filtration) \( \{\mathcal{F}_t\} \subset \mathcal{F} \), possibly those generated by a process \( \{X_s : s \leq t\} \) or even by the martingale itself, \( \mathcal{F}_{M_t} = \sigma\{M_s : s \leq t\} \). We require that \( E|M_t| < \infty \) for each \( t \) (so the conditional expectation below is well-defined) and that

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

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 Markov chains) it is only necessary to take \( s = t + 1 \), and usually we take \( \mathcal{F}_t = \sigma[X_i : i \leq t] \) and write

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

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 \( T = \mathbb{Z}_+ \), and more challenging for continuous time \( T = \mathbb{R}_+ \), so our text (Resnick 1998, chap. 10) covers only integer-time martingales.
1 Optional Stopping Theorem:

A random "time" \( \tau \in T \) is an \( \mathcal{F}_t \)-stopping time or Markov time if \( [\tau \leq t] \in \mathcal{F}_t \) for each \( t \in T \); informally, \( \tau \) "doesn’t depend on the future." If \( \tau \) 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:

\[
\begin{align*}
\mathbb{E}[M_{(t+1) \wedge \tau} | \mathcal{F}_t] &= \mathbb{E}[M_{\tau} 1_{[\tau \leq t]} + M_{t+1} 1_{[\tau > t]} | \mathcal{F}_t] \\
&= M_{\tau} 1_{[\tau \leq t]} + 1_{[\tau > t]} \mathbb{E}[M_{t+1} | \mathcal{F}_t] \\
&= M_{\tau} 1_{[\tau \leq t]} + 1_{[\tau > t]} M_t \\
&= M_{t \wedge \tau}.
\end{align*}
\]

1.1 Application: Simple Random Walks

Fix \( 0 < p < 1 \) and let \( \{ \xi_j \} \) be iid \( \pm 1 \)-valued random variables with \( \mathbb{P}[\xi_j = 1] = p \) and \( \mathbb{P}[\xi_j = -1] = q = 1 - p \), set \( \mathcal{F}_n = \sigma \{ \xi_j : j \leq n \} \), and let \( x \in \mathbb{Z} \). Set

\[
X_n = x + \sum_{j \leq n} \xi_j, \quad (1)
\]

a random walk that is either symmetric (if \( p = \frac{1}{2} \)) or not (if \( p \neq \frac{1}{2} \)). Set \( \mu := (p-q) \) and consider for \( n \in \mathbb{Z}_+ = \{0, 1, \ldots \} \) the three processes

\[
\begin{align*}
M_n^{(1)} &= X_n - \mu n \\
M_n^{(2)} &= (X_n - \mu n)^2 - 4pq n \\
M_n^{(3)} &= (q/p)X_n
\end{align*}
\]

Verify that each of these is a martingale by computing \( \mathbb{E}[M_{n+1}^{(i)} | \mathcal{F}_n] = M_n^{(i)} \) and applying the tower property and induction. For integers \( a \leq x \) and \( b \geq x \), and verify that \( \tau := \inf \{ t \geq 0 : X_t \notin (a, b) \} \) is a stopping time, finite a.s. by Borel-Cantelli. Let \( W = [\tau < \infty] \cap [X_\tau = b] \) be the event that \( X_t \) exits \( (a, b) \) to the right, i.e., that \( X_t \geq b \) before \( X_t \leq a \). If \( p = \frac{1}{2} = q \) (the symmetric case) then by DCT

\[
x = \mathbb{E}[M_0^{(1)}] = \lim_{t \to \infty} \mathbb{E}[M_{t \wedge \tau}^{(1)}] \\
= \mathbb{E}[M_{\tau}^{(1)}] = b \mathbb{P}[W] + a \mathbb{P}[W^c] \\
= (b-a) \mathbb{P}[W] + a, \quad \text{so}
\]

\[
\mathbb{P}[W] = \frac{x-a}{b-a}. \quad (3)
\]
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 unfair game, i.e., if $p \neq q$, then $(p/q)^b \neq (p/q)^a$ and again by DCT,

$$
(q/p)^x = \lim_{t \to \infty} E[M_{(3)}^t] = E[M_{(3)}^\tau]
$$

$$
= (q/p)^b P[W] + (q/p)^a P[W^c]
$$

$$
= \left[\frac{(q/p)^b}{(q/p)^a}\right] P[W] + (q/p)^a, \text{ so}
$$

$$
P[W] = \frac{(q/p)^x - (q/p)^a}{(q/p)^b - (q/p)^a}
$$

$$
= \frac{(p/q)^{b-x} - (p/q)^{b-a}}{1 - (p/q)^{b-a}}
$$

$$
\approx (p/q)^{b-x} \text{ if } b \gg a \text{ and } p < \frac{1}{2} < q.
$$

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 $P[W] = (0.910 - 0.9100)/(1 - 0.9100) = 0.34866$, and the chance of reaching $100$ before $0$ starting at $x = 50$ is $P[W] = (0.950 - 0.9100)/(1 - 0.9100) = 0.00513$, while these would be 90% and 50% in a fair game.

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

$$
x^2 = M_{(2)}^0 = \lim_{t \to \infty} E[M_{(2)}^t] = E[M_{(2)}^\tau]
$$

$$
= E[X_{\tau}^2 - \tau]
$$

$$
= b^2 P[W] + a^2 P[W^c] - E[\tau]
$$

$$
= \frac{b^2(x - a) + a^2(b - x)}{b - a} - E[\tau]
$$

$$
= (a + b)x - ab - E[\tau] \text{ so}
$$

$$
E[\tau] = (a + b)x - ab - x^2 = (b - x)(x - a)
$$

so 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 $E[\tau]$. 


from $M^{(1)}_\tau$:

$$x = M^{(1)}_0 = \lim_{t \to \infty} E[M^{(1)}_{t,\tau}] = E[M^{(1)}_\tau]$$

$$= E[X_\tau - \mu \tau]$$

$$= \frac{b((q/p)^x - (q/p)^a) + a((q/p)^b - (q/p)^x)}{(q/p)^b - (q/p)^a} - \mu E[\tau],$$

$$E_\tau = \frac{(b-x)((q/p)^x - (q/p)^a) + (a-x)((q/p)^b - (q/p)^x)}{\mu((q/p)^b - (q/p)^a)}$$

$$= \frac{(b-x)((p/q)^b-x - (p/q)^b-a) - (x-a)[1 - (p/q)^b-a]}{(p-q)[1 - (p/q)^b-a]}$$

or approximately $E_\tau \approx (x-a)/(q-p)$ for $a \ll b$ and $p < q$; for US roulette, $E_\tau = 1047.5$ for $x = 90$ (with a slim 35% chance of winning) and $E_\tau = 940.258$ 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 $P[W] = 37\%$ or $87\%$ with $E[\tau] = 24.46$ or $10.17$, respectively, much closer to the values $50\%$, $90\%$ for $P[W]$ and $25$, $10$ for $E_\tau$ in a fair game. Even faster (and more favorable) is the optimal strategy of bold play; for $x = 50$ this amounts to betting all $50$ at once ($E[W] = 9/19 = 47.37\%$, $E_\tau = 1$) while for $x = 90$, $E[W] = 87.94\%$.

Upon taking the limit as $a \to -\infty$ in Equations (3, 4) we find that $P[X_t \geq b$ for any $t < \infty]$ has probability one if $p \geq \frac{1}{2}$, but for $p < \frac{1}{2}$ the probability $(p/q)^{b-x} < 1$; thus even an infinitely-rich patron has only a $0.910^{10} = 34.8678\%$ chance of increasing a $90$ stake to $100$ in US roulette. The expected time to reach $b > x$ is infinite for $p \leq \frac{1}{2}$, but for $p > \frac{1}{2}$ the expected time is finite, $E[\tau] = (b-x)/(p-q) < \infty$.

### 1.1.1 Other Random Walks

More generally we can construct a process $X_n$ as in (1) for any iid $\{\xi\} \subset L_2$ and martingales $M^{(j)}_n$ as in (2), with $\mu = E\xi_j$ in (2a), replacing $4pq$ with $\sigma^2 = V\xi_j$ in (2b), and replacing $(q/p)$ with $e^{t^*}$ where $t^* \neq 0$ is the solution to $M(t^*) = 1$ for the MGF $M(t)$ of $\xi_j$ ($t^* < 0$ if $\mu > 0$, $t^* > 0$ if $\mu < 0$). Now the probabilities of Equations (3, 4) and expectations of Equations (5, 6) will only be approximate, since $X_\tau$ won’t be exactly $a$ or $b$. Abraham Wald (1945) studied the discrepancy in some detail, motivated by the following application.
1.1.2 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 \( \sigma \)-finite reference measure) \( f_0 \) and \( f_1 \), the likelihood ratio (against the Null) for the first \( n \) observations is

\[
\Lambda_n = \prod_{j \leq n} \frac{f_1(Y_j)}{f_0(Y_j)}.
\]

In Wald’s Sequential Probability Ratio Test (SPRT), one observes data sequentially until \( \Lambda_n \) passes an upper boundary \( 1 < U < \infty \) (in which case the null hypothesis \( H_0 : Y_j \overset{iid}{\sim} f_0(y) \, dy \) is rejected) or a lower boundary \( 0 < L < 1 \) (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 \( \tau = \inf \{ n : \Lambda_n \not\in (L, U) \} = \inf \{ n : X_n \not\in (a = \log L, b = \log U) \} \) is Wald’s stopping time, so the results of Section (1.1.1) apply. In addition, \( \Lambda_n \) itself is a martingale under \( f_0 \), and \( \Lambda_n^{-1} \) under \( f_1 \), giving convenient tools for bounding the probability of incorrect hypothesis-test results or the expected duration of a sequential test. A Bayesian with prior \( P[H_0] = \pi_0 \) would report posterior probability \( P[H_0 \mid \text{Data}] = (1 + \frac{\pi_1}{\pi_0} \Lambda_\tau)^{-1} \), or about \( \pi_0/(\pi_0 + \pi_1 a) \) if \( X_\tau \leq a \) and \( \pi_0/(\pi_0 + \pi_1 b) \) if \( X_\tau \geq b \), 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 \( P[\text{Reject } H_0 \mid H_0] \leq \alpha \) and \( P[\text{Reject } H_0 \mid H_1] \geq 1 - \beta \), 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 \( t \), the number of times it passes from below \( a \) to above \( b \); then:

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

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

$$
\alpha_n = \inf\{t > \beta_{n-1} : M_t \leq a\}
$$

$$
\beta_n = \inf\{t > \alpha_n : M_t \geq b\},
$$

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 $\alpha_1$ that $M_t \leq a$, $Y_t$ accumulates the increments of $M_t$ until the first time $\beta_1$ that $M_t \geq b$; the process continues if the martingale $M_t \leq a$ again falls below $a$ (at time $\alpha_2$), 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. Then

$$
Y_t = \sum_{n \in \mathbb{N}} [M_{t \wedge \beta_n} - M_{t \wedge \alpha_n}] 
\geq (b-a)\nu_{[a,b]}^{(t)} + [M_t - a]
$$

$$
\mathbb{E}Y_t \geq (b-a)\mathbb{E}\nu_{[a,b]}^{(t)} + \mathbb{E}[M_t - a]
\geq (b-a)\mathbb{E}\nu_{[a,b]}^{(t)} - \mathbb{E}(M_t - a) -
\geq (b-a)\mathbb{E}\nu_{[a,b]}^{(t)} - \mathbb{E}|M_t| - |a|.
$$

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

The important conclusion is that $\mathbb{E}\nu_{[a,b]}^{(t)} < \infty$. If $M_t$ is uniformly bounded in $L_1$, $\mathbb{E}|M_t| \leq c < \infty$ for all $t \in T$, then by Fatou’s lemma we even have $\mathbb{E}\nu_{[a,b]}^{(t)} \leq [c + |a|]/(b-a) < \infty$, so the number of times $\nu_{[a,b]}^{(t)} < \infty$ that $M_t$ ever crosses the interval $[a,b]$ is almost-surely finite— leading to

**Theorem 1 (Martingale Path Regularity)** Let $M_t^0$ be a martingale with index set $T = \mathbb{R}_+$. Then with probability one, $M_t^0$ has limits from the left and from the right at every point $t \in T$, 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.

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 $E|M_t| \leq c < \infty$ for all $t \in T$. Then there exists a random variable $M_\infty \in L_1$ such that $M_t \rightarrow M_\infty$ a.s. as $t \rightarrow \infty$. If $\{M_t\}$ is Uniformly Integrable (for example, if $E|M_t|^p \leq c_p < \infty$ for some $p > 1$), then also $M_t \rightarrow M_\infty$ in $L_1$.

Proof. Define $M_\infty \equiv \lim_{t \rightarrow \infty} M_t$ and $M_\infty \equiv \limsup_{t \rightarrow \infty} M_t$, and suppose (for contradiction) that $P[M_\infty = M_\infty] < 1$. Then there exist numbers $a < b$ for which $0 < P[M_\infty < a < b < M_\infty]$. But $
u_{[a,b]}^{(\infty)} = \infty$ on the event $[M_\infty < a < b < M_\infty]$, contradicting $E\nu_{[a,b]}^{(\infty)} \leq (c + |a|)/(b - a) < \infty$. The result about UI martingales now follows by the UI convergence theorem.

Corollary 1 Let $M_t$ be a martingale and $\tau$ a stopping time. Then

$$E[M_0] = E[M_\tau]$$

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

Proof. Obviously $M_\tau = \lim_{t \rightarrow \infty} M_{t \wedge \tau}$ 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

$$E[S_\tau] = 1 \neq 0 = E[S_0]$$

so the conclusion of Corollary 1 fails. Note that $S_n$ is not UI here, and $|S_s - S_t| \leq |s - t|$ is linearly bounded, but $E\tau = \infty$. For another example, let $X \sim Ge(\frac{1}{2})$ be a geometric random variable with $P[X = x] = 2^{1-x-1}$ for $x \in \mathbb{Z}_+$, and set $M_t \equiv 2^t1_{(X \geq t)}$; 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 $E[\tau] = 2$, but

$$E[M_\tau] = 0 \neq 1 = E[M_0].$$

Again $M_t$ is not UI, and this time $E\tau < \infty$ 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 \( M_{-\infty} \in L_1(\Omega, \mathcal{F}, P) \) such that

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

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 \( M_{-n} = X_n = S_n/n \); verify that \( M_t \) is a martingale for the filtration \( \mathcal{F}_t = \sigma\{M_s : s \leq t\} \) (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 \( \mu \), so \( X_n \to \mu \text{ a.s. as } n \to \infty \).

4 Martingale Problem for Markov Chains:

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 the Martingale Problem. We describe it for discretely-distributed processes \( X_t \), but similar characterizations apply to Markov processes with continuous marginal distributions.

4.1 Discrete Time

Let \( P_{jk}^{(t)} \) be a (possibly time-dependent) Markov transition matrix on a state space \( S \) indexed by \( T = \mathbb{Z}_+ \) or \( T = \mathbb{Z} \), so

\[
(\forall j, k \in S, t \in T) \; P_{jk}^{(t)} \geq 0 \quad \text{and} \quad (\forall j \in S, t \in T) \sum_{k \in S} P_{jk}^{(t)} = 1.
\]

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

\[
M_\phi(t) \equiv \phi(X_t) - \phi(X_0) - \sum_{0 \leq s < t} \sum_{j \neq i} P_{ij}^{(s)} [\phi(j) - \phi(i)]
\]

must be a martingale indexed by \( t \in T \). In the homogeneous case where \( P_{jk}^{(t)} \equiv P_{jk} \) doesn’t depend on \( t \), the \( n \)-step transition probability is simply
the matrix power $P^n$, and the operator
\[
\mathcal{G}\phi(i) = \sum_{j \neq i} P_{ij} [\phi(j) - \phi(i)]
\]
is called the generator of the process. If $\phi$ is harmonic, i.e., $\mathcal{G}\phi \equiv 0$, then $\phi(X_t)$ is a martingale. For the symmetric random walk on $\mathbb{Z}$, for example, $\mathcal{G}\phi(x) = \frac{1}{2} [\phi(x + 1) - 2\phi(x) + \phi(x - 1)]$, half the second-difference operator, so only affine functions $\phi(x) = mx + b$ are harmonic.

### 4.1.1 Example

Let $X_t$ be the simple random walk (1) starting at $x = 0$ with $P[\xi_j = 1] = p$ and $P[\xi_j = -1] = q = 1 - p$ with $0 < p < 1$; then for any $\epsilon \in \mathbb{R}$,
\[
Y_t = X_t - \epsilon t
\]
is a Markov process with generator
\[
\mathcal{G}\phi(y) = E[\phi(Y_{t+1}) - \phi(Y_t) \mid Y_t = y] = p\phi(y + 1 - \epsilon) - \phi(y) + q\phi(y - 1 - \epsilon),
\]
a finite-difference approximation to a second-order differential operator. Inspired by the exponential form of the continuous-time solutions to $\mathcal{G}\phi \equiv 0$, let’s search for harmonic functions of the form $\phi(y) = r^y$ for $r > 0$:
\[
0 \equiv \mathcal{G}\phi(y) = pr^{y+1-\epsilon} - r^y + qr^{y-1-\epsilon} = r^{y-1-\epsilon} [pr^2 - r^{1+\epsilon} + q]
\]
Set $\mu = p - q = 2p - 1$. The term in brackets (say, $h(r)$) vanishes for $r = 1$—but, since $h(0) = q > 0$, $h(1) = 0$, $h'(1) = \mu - \epsilon$, and $h(r) \to +\infty$ as $r \to \infty$, for any $-1 < \epsilon < \mu$ (resp., $\mu < \epsilon < 1$) there also exists $0 < r_* < 1$ (resp., $1 < r_* < \infty$) for which $h(r_*) = pr_*^2 - r_*^{1+\epsilon} + q$ vanishes and hence for which
\[
M_\phi(t) = r_*^{X_t - \epsilon t}
\]

is a martingale starting at $M_\phi(0) = 1$. By the Maximal Inequality (Theorem 4), for any $\lambda > 0$ and $\mu < \epsilon < 1$,
\[
P[ \sup_{0 \leq s \leq t} r_*^{X_s - \epsilon s} > \lambda] \leq 1/\lambda, \quad \text{i.e.,}
\]
\[
P[ \sup_{0 \leq s \leq t} \{X_s - \epsilon s\} > \log(\lambda)/\log(r_*)] \leq 1/\lambda, \quad \text{or, with } \lambda = r_*^b,
\]
\[
P[ \sup_{0 \leq s \leq t} \{X_s - \epsilon s - b\} > 0] \leq r_*^{-b}, \quad (7)
\]
giving a bound for the probability that the random walk \( X_s \) ever crosses the line \( y = \epsilon s + b \) (since the bound doesn’t depend on \( t < \infty \)). In the Roulette example, with \( p = 9/19 \), for \( \epsilon = 0 > \mu = \frac{-1}{19} \) we have \( r_* = q/p = 10/9 \), so (7) implies
\[
P[(x + X_t) \text{ ever exceeds } b] \leq (9/10)^{b-x},
\]
the same bound we found before, but now we have new results like
\[
P[X_t \text{ ever exceeds } (b + t/2)] \leq \frac{1}{3.382975^b}
\]
for a symmetric random walk and \( b \geq 0 \), since \( r_* \approx 3.382975 \) is the solution \( r > 1 \) to \( h(r) = \left[ r^2 - 2r^{3/2} + 1 \right] = 0 \).

### 4.2 Continuous Time

Let \( Q_{jk}^{(t)} \) be a (possibly time-dependent) continuous-time Markov transition rate matrix on a state space \( \mathcal{S} \) so
\[
(\forall j \neq k \in \mathcal{S}, t \in \mathcal{T}) \ Q_{jk}^{(t)} \geq 0 \quad \text{and} \quad (\forall j \in \mathcal{S}, t \in \mathcal{T}) \sum_{k \in \mathcal{S}} Q_{jk}^{(t)} = 0.
\]
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 \( \phi : \mathcal{S} \rightarrow \mathbb{R} \), the process
\[
M_{\phi}(t) = \phi(X_t) - \phi(X_0) - \int_0^t \left[ \sum_{j \neq i \in \mathcal{S}} Q_{ij}^{(s)} \left[ \phi(j) - \phi(i) \right] \right] ds
\]
must be a martingale. In the homogeneous case where \( Q_{jk}^{(t)} \equiv Q_{jk} \) doesn’t depend on \( t \), the time-\( t \) transition probability is simply the matrix exponential \( P^t = \exp(t Q) \), and the operator
\[
\mathfrak{S}\phi(i) = \sum_{j \neq i} Q_{ij} \left[ \phi(j) - \phi(i) \right]
\]
is called the (infinitesimal) generator of the process. If \( \phi \) is harmonic, i.e., \( \mathfrak{S}\phi \equiv 0 \), 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, \( \mathfrak{S}\phi(x) = \frac{1}{2} \Delta \phi(x) \), half the Laplacian, illustrating why functions that satisfy \( \mathfrak{S}\phi \equiv 0 \) are called harmonic.
4.2.1 Example

The unit-rate Poisson process $N(t)$ is characterized by its initial value and its generator $\mathcal{G}_\phi(x) = [\phi(x + 1) - \phi(x)];$ the sum

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

of independent Poisson processes with rates $\nu_j$ and jump sizes $u_j$ is also a continuous-time Markov process, with generator given for $\phi \in C^1(\mathbb{R})$ by

$$\mathcal{G}_\phi(x) = \sum_j [\phi(x + u_j) - \phi(x)] \nu_j$$

$$= \int_\mathbb{R} [\phi(x + u) - \phi(x)] \nu(du)$$

(8)

for the discrete measure $\nu(du) \equiv \sum_j u_j \delta_{\nu_j}(du)$, with log ch.f.

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

(9)

Actually Equations (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”

$$\int_\mathbb{R} (1 \land |u|) \nu(du) < \infty.$$

(10)

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

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

have gamma distributions; another is the symmetric $\alpha$-stable (S$\alpha$S) process $X_t \sim \text{St}(\alpha, 0, \gamma t, 0)$ with $\nu(du) = \frac{\alpha}{\pi} \frac{\Gamma(\alpha)}{2} \sin(\frac{\pi \alpha}{2}) |u|^{-\alpha-1} du$, with $\alpha$-stable increments. Equation (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”.

Page 11 December 2, 2010
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 $t \in T$. Then for any $\lambda > 0$,

$$
P \left[ \sup_{0 \leq s \leq t} M_s \geq \lambda \right] \leq \lambda^{-1} \mathbb{E} M_t^+
$$

$$
P \left[ \sup_{0 \leq s \leq t} |M_s| \geq \lambda \right] \leq \lambda^{-1} \mathbb{E} |M_t|
$$

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

$$
\mathbb{E} M_t = \mathbb{E} M_{t \wedge \tau} = \mathbb{E} \left\{ M_t \mathbf{1}_{[\tau \leq t]} + M_t \mathbf{1}_{[\tau > t]} \right\} \\
\geq \mathbb{E} \left\{ \lambda \mathbf{1}_{[\tau \leq t]} + M_t \mathbf{1}_{[\tau > t]} \right\} = \lambda \mathbb{P} [\tau \leq t] + \mathbb{E} \left\{ M_t \mathbf{1}_{[\tau > t]} \right\},
$$

so

$$
\mathbb{E} [M_t \mathbf{1}_{[\tau \leq t]}] \geq \lambda \mathbb{P} [\tau \leq t] \quad \text{and hence}
$$

$$
P \left\{ \sup_{0 \leq s \leq t} M_s \geq \lambda \right\} = P[\tau \leq t] \leq \lambda^{-1} \mathbb{E} [M_t \mathbf{1}_{[\tau \leq t]}] \leq \lambda^{-1} \mathbb{E} [M_t^+ \mathbf{1}_{[\tau \leq t]}] \leq \lambda^{-1} \mathbb{E} [M_t^+],
$$

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

$$
P \left\{ \inf_{0 \leq s \leq t} M_s \leq -\lambda \right\} \leq \lambda^{-1} \mathbb{E} [-M_t^-]; \quad \text{adding these together,}
$$

$$
P \left\{ \sup_{0 \leq s \leq t} |M_s| \geq \lambda \right\} \leq \lambda^{-1} \mathbb{E} [|M_t|].
$$

\[ \]
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
\[
\mathbb{P}\{ |M|_t^* \geq \lambda \} \leq \lambda^{-1} \mathbb{E}\left[ |M_t| \mathbb{1}_{\{|M|_t^* \geq \lambda\}} \right].
\] (11)

**Theorem 5** For any martingale \( M_t \) and any real numbers \( p > 1, q = \frac{p}{p-1} \),
\[
\| \sup_{s \leq t} |M_s| \|_p \leq q \sup_{s \leq t} \| M_s \|_p.
\]

**Proof.**

By Fubini’s theorem,
\[
\mathbb{E}(|M|_t^*)^p = \int_0^\infty p\lambda^{p-1} \mathbb{P}( |M|_t^* \geq \lambda ) d\lambda
\]
\[
\leq \int_0^\infty p\lambda^{p-1} \lambda^{-1} \mathbb{E}\left[ |M_t| \mathbb{1}_{\{|M|_t^* \geq \lambda\}} \right] d\lambda
\]
\[
= \mathbb{E} \int_0^{\|M|_t^*} p\lambda^{p-2} |M_t| d\lambda
\]
\[
= \frac{p}{p-1} \mathbb{E}(|M|_t^*)^{p-1} |M_t|.
\]

Hölder’s inequality asserts that \( \mathbb{E}(YZ) \leq \{\mathbb{E}Y^p\}^{1/p} \{\mathbb{E}Z^q\}^{1/q} \) 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
\[
\{\mathbb{E}(|M|_t^*)^p\}^{1/p} \leq q \mathbb{E}\{(|M|_t^*)^{p-1}\}^{1/q} \mathbb{E}\{|M_t|^p\}^{1/p}
\]
\[
\{\mathbb{E}(|M|_t^*)^p\}^{1/p} \leq q \| |M|_t^* \|_p \leq q \sup_{0 \leq s \leq t} \| M_s \|_p.
\]

Note that \( q \to \infty \) so the bound blows up since \( p \to 1 \); to achieve an \( L_1 \) bound on \( \mathbb{E}|M|_t^* \) we need something slightly stronger than an \( L_1 \) bound on \( \mathbb{E}|M_t| \) (see below).
In summary: if $M_t$ is a martingale and if $t \in T$ then
\begin{align*}
P[\sup_{s \leq t} M_s \geq \lambda] & \leq \lambda^{-1} E[M_t^+] \\
P[\min_{s \leq t} M_s \leq -\lambda] & \leq \lambda^{-1} E[M_t^-] \\
E[\sup_{s \leq t} |M_s|] & \leq \lambda^{-1} E[|M_t|]
\end{align*}
\begin{align*}
E[\sup_{s \leq t} |M_s|^p] & \leq q^p \sup_{s \leq t} E[|M_s|^p] = q^p E[|M_t|^p] \quad (p > 1) \\
E[\sup_{s \leq t} |M_s|] & \leq e^{-1} \sup_{s \leq t} E[|M_s| \log^+(|M_s|)] \quad (p = 1)
\end{align*}

6 Doob’s Martingale:

Fix any $Y \in L_1(\Omega, \mathcal{F}, P)$ and set $M_t = E[Y \mid \mathcal{F}_t]$, 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:
\begin{itemize}
  \item 1. They have close connections with Markov processes;
  \item 2. Their expectations at stopping times are easy to compute;
  \item 3. They offer a tool for bounding the maxima and minima of processes;
  \item 4. They offer a tool for establishing path regularity of processes;
  \item 5. They offer a tool for establishing the a.s. convergence of certain random sequences;
  \item 6. They are important for modeling economic and statistical time series which are, in some sense, predictions.
\end{itemize}

Examples:
\begin{itemize}
  \item 1. Partial sums $S_n = \sum_{i=1}^n X_i$ 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 $\mathbb{E}X_n = p$. 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 $F_n \in \sigma\{X_1, \ldots, X_n\}$, then $\mathbb{E}[M_{n+1} | F_n] = M_n$ 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: $M_n = (\sum_{i=1}^n Y_i)^2 - n\sigma^2$, where $\mathbb{E}Y_i Y_j = \sigma^2 \delta_{ij}$

4. Radon-Nikodym Derivatives:

$$M_n(\omega) = 2^{-n} \int_{i/2^n}^{(i+1)/2^n} f(x) \, dx, \quad i = [2^n \omega]$$

$$\rightarrow M_\infty(\omega) = f(\omega) \quad a.s.$$ 

5. Leftovers:

- Submartingales: $\mathbb{E}[X_t^+] < \infty$, $X_t \in \mathcal{F}_t$, $X_t \leq \mathbb{E}[X_s | \mathcal{F}_t]$ for $s > t$.

- Supermartingales: If $X_t$ is a submartingale then $Y_t = -X_t$ is a supermartingale, satisfying $Y_t \geq \mathbb{E}[Y_s | \mathcal{F}_t]$ for $s > t$.

- Jensen’s inequality: if $M_t$ is a martingale and if $\phi$ convex with $\mathbb{E}[\phi(M_t)^+] < \infty$, 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: $(\exists Y_\infty \in L_1) \ Y_t \rightarrow Y \ a.s. \ (and, \ if \ \{Y_t\} \ is \ UI, \ also \ in \ L_1)$.

References

