Martingales

1. Introduction

Def: A sequence $Y = \{Y_n : n \geq 0\}$ is a *martingale* with respect to the sequence $X = \{X_n : n \geq 0\}$ if, for all $n \geq 0$,

- $\mathbb{E}|Y_n| < \infty$,
- $\mathbb{E}(Y_{n+1}|X_0, X_1, \cdots, X_n) = Y_n$.

*Warning note: conditional expectations are ubiquitous in this chapter, Remember that they are random variablesm and that formulae of the form $\mathbb{E}(A|B) = C$ generally hold only 'almost surely'. We shall omit the term 'almost surely' throughout the chapter.

*We will introduce a general definition of martingale later.

Examples:

1. Simple random walk

   A particle jumps either one step to the right or one step to the left, with corresponding probabilities p and $q(= 1 - p)$. Assuming the usual independence of different moves, it is clear that the position $S_n = X_1 + X_2 + \cdots + X_n$ of the particle after n steps satisfies $\mathbb{E}|S_n| \leq n$ and

   $$\mathbb{E}(S_{n+1}|X_1, X_2, \cdots, X_n) = S_n + (p - q),$$

   whence it is easily seen that $Y_n = S_n - n(p - q)$ defines a martingale with respect to X.

2. The martingale

   The following gambling strategy is called a martingale.

   A gambler has a large fortune. He wagers \$1 on an evens bet. If he loses then he wagers \$2 on the next bet. If he loses the first n plays, then he bets $\$2^n$ on the $(n+1)$th. He is bound to win sooner or later, say on the Tth bet, at which point he ceases to play, and leaves with his profit of $2^T - (1 + 2 + 2^2 + \cdots + 2^{T-1})$. Thus, following this strategy, he is assured an ultimate profit. This sounds like a good policy.

   Writing Y_n for the accumulated gain of the gambler after the nth play (losses count negative), we have that $Y_0 = 0$ and $|Y_n| \leq 1 + 2 + \cdots + 2^{n-1} = 2^n - 1$. Furthermore, $Y_{n+1} = Y_n$ if the gambler has stopped by time $n + 1$, and

   $$Y_{n+1} = \begin{cases} Y_n - 2^n & \text{with probability } \frac{1}{2}, \\ Y_n + 2^n & \text{with probability } \frac{1}{2}, \end{cases}$$

otherwise, implying that $\mathbb{E}(Y_{n+1}|Y_1, Y_2, \cdots, Y_n) = Y_n$. Therefore Y is a martingale (with repect to itself).

*Remark: This martingale possesses a particularly disturbing feature. The random time T has a geometric distribution, $P(T = n) = (1/2)^n$ for $n \geq 1$, so that the mean loss of the gambler just before his ultimate win is

$$\sum_{n=1}^{\infty}(\frac{1}{2})^n(1 + 2 + \cdots + 2^{n-1})$$

which equals infinity. Do not follow this strategy unless your initial capital is considerably greater than that of the casino.

3. De Moivre's martingale

A simple random walk on the set $\{0, 1, \cdots, N\}$ stops when it first hits either of the absorbing barriers at 0 and at N; what is the probability that it stops at the barrier 0?

We first demonstrate a straightforward way to calculate the specified probability. Let p_k be the probability of ultimate ruin starting from k. We have

$$p_k = p \cdot p_{k+1} + q \cdot p_{k-1} \quad \text{if} \quad 1 \leq k \leq N - 1$$

with boundary conditions $p_0 = 1, p_N = 0$. The solution of the difference equation with the boundary conditions is given by

$$p_k = \frac{(q/p)^k - (q/p)^N}{1 - (q/p)^N}.$$

Abraham de Moivre made use of a martingale to answer the 'gambler's ruin' question. Write $X_1, X_2, \cdots$ for the steps of the walk, and S_n for the position after n steps, where $S_0 = k$. Define $Y_n = (q/p)^{S_n}$ where $p = P(X_i = 1), p + q = 1$, and $0 < p < 1$. We claim that

$$\mathbb{E}(Y_{n+1}|X_1, X_2, \cdots, X_n) = Y_n \quad \text{for all } n. \qquad (1)$$

If S_n equals 0 or N then the process has stopped by time n, implying that $S_{n+1} = S_n$ and therefore $Y_{n+1} = Y_n$. If on the other hand $0 < S_n < N$, then

$$\mathbb{E}(Y_{n+1}|X_1, \cdots, X_n) = \mathbb{E}\left((q/p)^{S_n + X_{n+1}}|X_1, \cdots, X_n\right) = (q/p)^{S_n}[p(q/p) + q(p/q)^{-1}] = Y_n,$$

and the claim is proved. It follows, by taking expectations of (1), that $\mathbb{E}(Y_{n+1}) = \mathbb{E}(Y_n)$ for all n, and hence $\mathbb{E}|Y_n| = \mathbb{E}|Y_0| = (q/p)^k$ for all n. In particular Y is a martingale (with respect to the sequence X).

Let T be the number of steps before the absorption of the partical at either 0 or N. De Moivre argued as follows: $\mathbb{E}(Y_n) = (q/p)^k$ for all n, and therefore $\mathbb{E}(Y_T) = (q/p)^k$. If you are willing to accept this remark, then the answer to the original question is a simple consequence, as follows. Expanding $\mathbb{E}(Y_T)$, we have that

$$\mathbb{E}(Y_T) = (q/p)^0 p_k + (q/p)^N(1 - p_k).$$

However, $\mathbb{E}(Y_T) = (q/p)^k$ by assumption, and therefore

$$p_k = \frac{\rho^k - \rho^N}{1 - \rho^N}, \quad \text{where} \quad \rho = q/p$$

so long as $\rho \neq 1$, in agreement with the calculation at the beginning.

*This is a very attractive method, which relies on the statement that $\mathbb{E}(Y_T) = \mathbb{E}(Y_0)$ for a certain type of random variable T. A major part of our investigation of martingales will be to determine conditions on such random variables T which ensure that the desired statements are true.

4. Markov chains

   Let X be a discrete-time Markov chain taking values in the countable state space S with transition matrix $\mathbf{P}$. Suppose that $\psi : S \to S$ is bounded and harmonic, which is to say that

$$\sum_{j \in S} p_{ij}\psi(j) = \psi(i) \quad \text{for all } i \in S.$$

   It is easy seen that $Y = \{\psi(X_n) : n \geq 0\}$ is a martingale with respect to X:

$$\mathbb{E}(\psi(X_{n+1})|X_1, \cdots, X_n) = \mathbb{E}(\psi(X_{n+1})|X_n) = \sum_{j \in S} p_{X_n,j}\psi(j) = \psi(X_n).$$

   More generally, suppose that ψ is a right eigenvector of $\mathbf{P}$, which is to say that there exists $\lambda(\neq 0)$ such that

$$\sum_{j \in S} p_{ij}\psi(j) = \lambda\psi(i), \quad i \in S.$$

   Then

$$\mathbb{E}(\psi(X_{n+1})|X_1, \cdots, X_n) = \lambda\psi(X_n),$$

   implying that $\lambda^{-n}\psi(X_n)$ defines a martingale so long as $\mathbb{E}|\psi(X_n)| < \infty$ for all n.

Next we will give a general definition of a martingale. Before proceeding to the deifinition, we recall the most general form of conditional expectation and some other terminnology.

Def: Let Y be a random variable on the probability space $(\Omega, \mathcal{F}, P)$ having finite mean, and let $\mathcal{G}$ be a sub-σ-field of $\mathcal{F}$. The *conditionla expectation* of Y given $\mathcal{G}$, written $\mathbb{E}(Y|\mathcal{G})$, is a $\mathcal{G}$-measurable random variable satisfying

$$\mathbb{E}([Y - \mathbb{E}(Y|\mathcal{G})]I_G) = 0 \quad \text{for all events } G \in \mathcal{G},$$

where I_G is the indicator function of G.

Def: Suppose that $\mathcal{F} = \{\mathcal{F}_0, \mathcal{F}_1, \cdots\}$ is a sequence of sub-σ-fields of $\mathcal{F}$; we call $\mathcal{F}$ a filtration if $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$ for all n. A sequence $Y = \{Y_n : n \geq 0\}$ is said to be adapted to the filtration $\mathcal{F}$ if Y_n is $\mathcal{F}_n$-measurable for all n. Given a filtration $\mathcal{F}$, we normally write $\mathcal{F}_\infty = \lim_{n\to\infty} \mathcal{F}_n$ for the smalles σ-field containing $\mathcal{F}_n$ for all n.

Def: Let $\mathcal{F}$ be a filtration of the probability space $(\Omega, \mathcal{F}, P)$, and let Y be a sequence of random variables which is adapted to $\mathcal{F}$. We call the pair $(Y, \mathcal{F}) = \{(Y_n, \mathcal{F}_n) : n \geq 0\}$ a *martingale* if, for all $n \geq 0$,

- $\mathbb{E}|Y_n| < \infty$,
- $\mathbb{E}(Y_{n+1}|\mathcal{F}_n) = Y_n$.

The former definition of martingale is retrieved by choosing $\mathcal{F}_n = \sigma(X_0, X_1, \cdots, X_n)$, the smallest σ-field with respect to which each of the variables $X_0, X_1, \cdots, X_n$ is measurable.

There are many cases of interest in which the martingale condition $\mathbb{E}(Y_{n+1}|\mathcal{F}_n) = Y_n$ does not hold, being replaced instead by an inequality: $\mathbb{E}(Y_{n+1}|\mathcal{F}_n) \geq Y_n$ for all n, or $\mathbb{E}(Y_{n+1}|\mathcal{F}_n) \leq Y_n$ for all n. Sequences satisfying such inequalities have many of the properties of martingales, and we have special names for them.

Def: Let $\mathcal{F}$ be a filtration of the probability space $(\Omega, \mathcal{F}, P)$, and let Y be a sequence of random variables which is adapted to $\mathcal{F}$. We call the pair $(Y, \mathcal{F}) = \{(Y_n, \mathcal{F}_n) : n \geq 0\}$ a *submartingale* if, for all $n \geq 0$,

- $\mathbb{E}(Y_n^+) < \infty$,
- $\mathbb{E}(Y_{n+1}|\mathcal{F}_n) \geq Y_n$,

or a *supermartingale* if, for all $n \geq 0$,

- $\mathbb{E}(Y_n^-) < \infty$,
- $\mathbb{E}(Y_{n+1}|\mathcal{F}_n) \leq Y_n$.

2. Martingale differences and Hoeffding's inequality

Def: Let $(Y, \mathcal{F})$ be a martingale. The sequence of *martingale differences* is the sequence $D = \{D_n : n \geq 1\}$ defined by $D_n = Y_n - Y_{n-1}$, so that

$$Y_n = Y_0 + \sum_{i=1}^{n} D_i.$$

Note that the sequence D is such that D_n is $\mathcal{F}_n$-measurable, $\mathbb{E}|D_n| < \infty$, and $\mathbb{E}(D_{n+1}|\mathcal{F}_n) = 0$ for all n.

Theorem (Hoeffding's inequality): Let $(Y, \mathcal{F})$ be a martingale, and suppose that there exists a sequence $K_1, K_2, \cdots$ of real numbers such that $P(|Y_n - Y_{n-1}| \leq K_n) = 1$ for all n. Then

$$P(|Y_n - Y_0| \geq x) \leq 2\exp\left(-\frac{x^2}{2\sum_{i=1}^n K_i^2}\right), \quad x > 0.$$

That is to say, if the martingale differnces are bounded (almost surely) then there is only a small chance of a large deviation of Y_n from its initial value Y_0.

proof: Applying Markov's inequality, we have for $\theta > 0$,

$$P(Y_n - Y_0 \geq x) \leq e^{-\theta x}\mathbb{E}(e^{\theta(Y_n - Y_0)}).$$

If $\psi > 0$, and D is a random variable having mean 0 and satisfying $P(|D| \leq 1) = 1$, then we obtain

$$\mathbb{E}(e^{\psi D}) \leq \mathbb{E}\left(\frac{1}{2}(1-D)e^{-\psi} + \frac{1}{2}(1+D)e^{\psi}\right) = \frac{1}{2}(e^{-\psi} + e^{\psi}) < e^{\frac{1}{2}\psi^2}$$

by the convexity of $g(d) = e^{\psi d}$ and a comparison of the coefficients of ψ^{2n} for $n \geq 0$.

Writing $D_n = Y_n - Y_{n-1}$, we have that

$$\mathbb{E}(e^{\theta(Y_n - Y_0)}) = \mathbb{E}(e^{\theta(Y_{n-1} - Y_0)}e^{\theta D_n}).$$

By conditioning on $\mathcal{F}_{n-1}$, we obtain

$$\mathbb{E}(e^{\theta(Y_n - Y_0)}|\mathcal{F}_{n-1}) = e^{\theta(Y_{n-1} - Y_0)}\mathbb{E}(e^{\theta D_n}|\mathcal{F}_{n-1}) \leq e^{\theta(Y_{n-1} - Y_0)}\exp\left(\frac{1}{2}\theta^2 K_n^2\right)$$

where we have used the fact that $Y_{n-1} - Y_0$ is $\mathcal{F}_{n-1}$-measurable, in addition to the second inequality in the proof applied to the random variable D_n/K_n. We take expectation of the above inequality and iterate to find that

$$\mathbb{E}(e^{\theta(Y_n - Y_0)}) \leq \mathbb{E}(e^{\theta(Y_{n-1} - Y_0)})\exp\left(\frac{1}{2}\theta^2 K_n^2\right) \leq \exp\left(\frac{1}{2}\theta^2\sum_{i=1}^n K_i^2\right).$$

Therefore, we have

$$P(Y_n - Y_0 \geq x) \leq e^{-\theta x}\mathbb{E}(e^{\theta(Y_n - Y_0)}) \leq \exp\left(-\theta x + \frac{1}{2}\theta^2\sum_{i=1}^n K_i^2\right) \leq \exp\left(-\frac{x^2}{2\sum_{i=1}^n K_i^2}\right), \quad x > 0.$$

The same argument is valid with $Y_n - Y_0$ replaced by $Y_0 - Y_n$, and the claim of the theorem follows by adding the two (identical) bounds together.

Examples:

1. Bin packing

   The bin packing problem is a basic problem of operations research.

Given n objects with sizes $x_1, x_2, \cdots, x_n$, and an unlimited collection of bins each of size 1, what is the minimum number of bins required in order to pack the objects?

In the randomized version of this problem, we suppose that the objects have independent random sizes $X_1, X_,2, \cdots$ having some common distribution on $[0, 1]$. Let B_n be the (random) number of bins required in order to pack $X_1, X_2, \cdots, X_n$ efficiently; that is, B_n is the minimum number of bins of unit capacity such that the sum of the sizes of the objects in any given bin does not exceed its capacity.

It may be shown that B_n grows approximately linearly in n, in that there exists a positive constant β such that $n^{-1}B_n \to \beta$ a.s. and in mean square as $n \to \infty$. We shall not prove this here, but note its consequence:

$$\frac{1}{n}\mathbb{E}(B_n) \to \beta \quad \text{as} \quad n \to \infty.$$

The next question might be to ask how close B_n is to its mean value $\mathbb{E}(B_n)$, and Hoeffding's inequality may be brought to bear here.

For $i \le n$, let $Y_i = \mathbb{E}(B_n|\mathcal{F}_i)$, where $\mathcal{F}_i$ is the σ-field generated by $X_1, X_2, \cdots, X_i$. It is easily seen that $(Y, \mathcal{F})$ is a martingale by the tower property of expectation:

$$\mathbb{E}(Y_{i+1}|\mathcal{F}_i) = \mathbb{E}\left(\mathbb{E}(B_n|\mathcal{F}_{i+1})|\mathcal{F}_i\right) = \mathbb{E}(B_n|\mathcal{F}_i) = Y_i.$$

Furthermore, $Y_n = B_n$ and $Y_0 = \mathbb{E}(B_n)$ since $\mathcal{F}_0$ is the trivial σ-field $\{\emptyset, \Omega\}$.

Now, let $B_n(i)$ be the minimal number of bins required in order to pack all the objects except the i th. Since the objects are packed efficiently, we msut have $B_n(i) \le B_n \le B_n(i) + 1$. Taking conditional expectations given $\mathcal{F}_{i-1}$ and $\mathcal{F}_i$ we obtain

$$\begin{gathered}\mathbb{E}(B_n(i)|\mathcal{F}_{i-1}) \le Y_{i-1} \le \mathbb{E}(B_n(i)|\mathcal{F}_{i-1}) + 1, \\ \mathbb{E}(B_n(i)|\mathcal{F}_i) \le Y_i \le \mathbb{E}(B_n(i)|\mathcal{F}_i) + 1.\end{gathered}$$

However, $\mathbb{E}(B_n(i)|\mathcal{F}_{i-1}) = \mathbb{E}(B_n(i)|\mathcal{F}_i)$, since we are not required to pack the ith object, and hence knowledge of X_i is irrelevant. It follows from that $|Y_i - Y_{i-1}| \le 1$. We may now apply Hoeffding's inequality to find that

$$P(|B_n - \mathbb{E}(B_n)| \ge x) \le 2\exp(-\frac{1}{2}x^2/n), \quad x > 0.$$

For example, setting $x = \varepsilon n$, we see that the chance that B_n deviates from its mean by εn (or more) decays exponentially in n as $n \to \infty$. Using $\frac{1}{n}\mathbb{E}(B_n) \to \beta$, we have

$$P(|B_n - \beta n| \ge \varepsilon n) \le 2\exp\left(-\frac{1}{2}\varepsilon^2 n(1 + o(1))\right).$$

2. Travelling salesman problem

A travelling salesman is required to visit n towns but may choose his route. How does he find the shortest possible route, and how long is it?

Here is a randomized version of the problem. Let $P_i = (U_i, V_i), i = 1, \cdots, n$ be independent and uniformly distributed points in the unit square $[0,1]^2$. It is required to tour these points using an airplane. If we tour them in the order $P_{\pi(1)}, P_{\pi(2)}, \cdots, P_{\pi(n)}$, for some permutation π of $\{1, 2, \cdots, n\}$, the total length of the journey is

$$d(\pi) = \sum_{i=1}^{n-1} |P_{\pi(i+1)} - P_{\pi(i)}| + |P_{\pi(n)} - P_{\pi(1)}|.$$

The shortest tour has length $D_n = \min_\pi d(\pi)$. It turns out that the asymptotic behavior of D_n for large n is given as follows: there exists a positive constant τ such that $D_n/\sqrt{n} \to \tau$ a.s. and in mean square. We shall not prove this, but note the consequence that

$$\frac{1}{\sqrt{n}} \mathbb{E}(D_n) \to \tau \quad \text{as} \quad n \to \infty.$$

How close is D_n to its mean?

Set $Y_i = \mathbb{E}(D_n|\mathcal{F}_i)$ for $i \le n$, where $\mathcal{F}_i$ is the σ-field generated by $P_1, P_2, \cdots, P_i$. As before, $(Y, \mathcal{F})$ is a martingale, and $Y_n = D_n$, $Y_0 = \mathbb{E}(D_n)$.

Let $D_n(i)$ be the minimal tour-length through the points $P_1, P_2, \cdots, P_{i-1}, P_{i+1}, \cdots, P_n$, and note that $\mathbb{E}(D_n(i)|\mathcal{F}_i) = \mathbb{E}(D_n(i)|\mathcal{F}_{i-1})$. We have

$$D_n(i) \le D_n \le D_n(i) + 2Z_i, \quad i \le n-1,$$

where Z_i is the shortest distance from P_i to one of the points $P_{i+1}, P_{i+2}, \cdots, P_n$.

We take conditional expectations to obtain

$$\mathbb{E}(D_n(i)|\mathcal{F}_{i-1}) \le Y_{i-1} \le \mathbb{E}(D_n(i)|\mathcal{F}_{i-1}) + 2\mathbb{E}(Z_i|\mathcal{F}_{i-1}),$$
$$\mathbb{E}(D_n(i)|\mathcal{F}_i) \le Y_i \le \mathbb{E}(D_n(i)|\mathcal{F}_i) + 2\mathbb{E}(Z_i|\mathcal{F}_i),$$

and hence

$$|Y_i - Y_{i-1}| \le 2\max\{\mathbb{E}(Z_i|\mathcal{F}_i), \mathbb{E}(Z_i|\mathcal{F}_{i-1})\}, \quad i \le n-1.$$

In order to estimate the right hand side here, let $Q \in [0,1]^2$, and let $Z_i(Q)$ be the shortest distance from Q to the closest of a collection of $n-i$ points chosen uniformly at random from the unit square. If $Z_i(Q) > x$ then no point lies within the circle $\mathcal{C}(x, Q)$ having radius x nad center at Q. Note that $\sqrt{2}$ is the largest possible distance between two points in the square. Now, there exists c such that, for all $x \in (0, \sqrt{2}]$, the intersection of $\mathcal{C}(x, Q)$ with the unit square has area at least cx^2, uniformly in Q. Therefore

$$P(Z_i(Q) > x) \le (1 - cx^2)^{n-i}, \quad 0 < x \le \sqrt{2}.$$

Integrating over x, we find that

$$\mathbb{E}(Z_i(Q)) \le \int_0^{\sqrt{2}} (1 - cx^2)^{n-i}\, dx \le \int_0^{\sqrt{2}} e^{-cx^2(n-i)}\, dx = \frac{1}{\sqrt{n-i}} \int_0^{\sqrt{2(n-i)}} e^{-cy^2}\, dy < \frac{C}{\sqrt{n-i}}$$

for some constant C. So, we deduce that the random variables $\mathbb{E}(Z_i|\mathcal{F}_i)$ and $\mathbb{E}(Z_i|\mathcal{F}_{i-1})$ are smaller than $C/\sqrt{n-i}$, whence

$$|Y_i - Y_{i-1}| \le \frac{2C}{\sqrt{n-i}} \quad \text{for } i \le n-1.$$

For the case $i = n$, we use the trivial bound $|Y_n - Y_{n-1}| \le 2\sqrt{2}$.

Applying Hoeffding's inequality, we obtain

$$P(|D_n - \mathbb{E}(D_n)| \ge x) \le 2\exp\left(-\frac{x^2}{2(8+\sum_{i=1}^{n-1} 4C^2/i)}\right) \le 2\exp(-Ax^2/\log n), \quad x > 0,$$

for some positive constant A. Combining this with $\frac{1}{\sqrt{n}}\mathbb{E}(D_n) \to \tau$, we find that

$$P(|D_n - \tau\sqrt{n}| \ge \varepsilon\sqrt{n}) \le 2\exp(-B\varepsilon^2 n/\log n), \quad \varepsilon > 0,$$

for some positive constant B and all large n.

3. Crossings and convergence

Theorem (Martingale convergence theorem): Let $(Y, \mathcal{F})$ be a submartingale and suppose that $\mathbb{E}(Y_n^+) \le M$ for some M and all n. There exists a random variable Y_∞ such that $Y_n \xrightarrow{a.s.} Y_\infty$ as $n \to \infty$. We have in addition that:

1. Y_∞ has finite mean if $\mathbb{E}|Y_0| < \infty$, and
2. $Y_n \xrightarrow{1} Y_\infty$ if the sequence $\{Y_n : n \ge 0\}$ is uniformly integrable.

*We will prove the theorem after prensenting some of its applications and examples.

*Remark1: Definition of uniformly integrable sequence of random variables: A sequence $X_1, X_2, \cdots$ of random variables is said to be uniformly integrable if

$$\sup_n \mathbb{E}(|X_n| I_{\{|X_n| \ge a\}}) \to 0 \quad \text{as } a \to \infty.$$

*Remark2: It follows that any submartingale or supermartingale $(Y, \mathcal{F})$ converges almost surely if it satisfies $\mathbb{E}|Y_n| \le M$. We also have the following corollary of the martingale convergence theorem.

Theorem: If $(Y, \mathcal{F})$ is either a non-negative supermartingale or a non-positive submartingale, then $Y_\infty = \lim_{n\to\infty} Y_n$ exists almost surely.

proof: If Y is a non-positive submartingale, then $\mathbb{E}(Y_n^+) = 0$, whence the result follows from martingale convergence theorem. For a non-negative supermartingale Y, apply the same argument to $-Y$.

Examples:

1. Random walk

   Consider de Moivre;s martingale of Example 3 in Section 1, namely $Y_n = (q/p)^{S_n}$ where S_n is the position after n steps of the usual simple random walk. The sequence $\{Y_n\}$ is a non-negative martingale, and hence converges almost surely to some finite limit Y as $n \to \infty$. This is not of much interest if $p = q$ since $Y_n = 1$ for all n in this case. Suppose then that $p \neq q$. The random variable Y_n takes values in the set $\{\rho^k : k = 0, \pm 1, \cdots\}$ where $\rho = q/p$. Certainly Y_n cannot converge to any given (possobly random) member of this set, since this would necessarily entail that S_n converges to a finite limit (which is obviously false). Therefore Y_n converges to a limit point of the set, not lying within the set. The only such limit point which is finite is 0, and therefore $Y_n \to 0$ a.s. Hence, $S_n \to -\infty$ a.s. if $p < q$, and $S_n \to \infty$ a.s. if $p > q$. Note that Y_n does not converge in mean, since $\mathbb{E}(Y_n) = \mathbb{E}(Y_0) = 1 \neq 0$ for all n.

2. Doob's martingale (though some ascribe the construction to Lévy)

   Let Z be a random variable on $(\Omega, \mathcal{F}, P)$ such that $\mathbb{E}|Z| < \infty$. Suppose that $\mathcal{F} = \{\mathcal{F}_0, \mathcal{F}_1, \cdots\}$ is a filtration, and write $\mathcal{F}_\infty = \lim_{n\to\infty} \mathcal{F}_n$ for the smallest σ-field containing every $\mathcal{F}_n$. Now define $Y_n = \mathbb{E}(Z|\mathcal{F}_n)$. It is easy seen that $(Y, \mathcal{F})$ is a martingale:

   $$\begin{gathered}\mathbb{E}|Y_n| = \mathbb{E}|\mathbb{E}(Z|\mathcal{F}_n)| \leq \mathbb{E}\left\{\mathbb{E}(|Z||\mathcal{F}_n)\right\} = \mathbb{E}|Z| < \infty, \\ \mathbb{E}(Y_{n+1}|\mathcal{F}_n) = \mathbb{E}[\mathbb{E}(Z|\mathcal{F}_{n+1})|\mathcal{F}_n] = \mathbb{E}(Z|\mathcal{F}_n).\end{gathered}$$

   Furthermore, $\{Y_n\}$ is a uniformly integrable sequence*. It follows by the martingale convergence theorem that $Y_\infty = \lim_{n\to\infty} Y_n$ exists almost surely and in mean. As a matter of a fact, one can argue that $Y_\infty = \mathbb{E}(Z|\mathcal{F}_\infty)$.

   *Uniform integrability of the sequence $\{Y_n\}$: As a consequence of Jensen's inequality, the following holds almost surely:

   $$|Y_n| = |\mathbb{E}(Z|\mathcal{F}_n)| \leq \mathbb{E}\left(|Z||\mathcal{F}_n\right).$$

   So, $\mathbb{E}(|Y_n|I_{\{|Y_n|\geq a\}}) \leq \mathbb{E}(X_n I_{\{X_n \geq a\}})$ where $X_n = \mathbb{E}(|Z||\mathcal{F}_n)$. By the definition of conditional expectation, $\mathbb{E}\left\{(|Z| - X_n)I_{\{X_n\geq a\}}\right\}$, so that

   $$\mathbb{E}(|Y_n|I_{\{|Y_n|\geq a\}}) \leq \mathbb{E}(|Z|I_{\{X_n\geq a\}}).$$

   Now, by Markov's inequality,

   $$P(X_n \geq a) \leq a^{-1}\mathbb{E}(X_n) = a^{-1}\mathbb{E}|Z| \to 0 \quad \text{as } a \to 0, \text{ uniformly in } n.$$

   Using the fact that

   $$\mathbb{E}|Y| < \infty \Leftrightarrow \sup_{A:P(A)<\delta} \mathbb{E}(|Y|I_A) \to 0, \quad \text{as } \delta \to 0,$$

we deduce that $\mathbb{E}(|Z|I_{\{X_n \geq a\}}) \to 0$ as $a \to 0$ uniformly in n, implying that the sequence $\{Y_n\}$ is uniformly integrable.

Reference: Grimmett & Stirzaker *Probability and Random Processes* Third Edition (2001)