

17.1: Introduction to Martingales

Basic Theory

Basic Assumptions

For our basic ingredients, we start with a stochastic process $\mathbf{X} = \{X_t : t \in T\}$ on an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$, having state space \mathbb{R} , and where the index set T (representing time) is either \mathbb{N} (discrete time) or $[0, \infty)$ (continuous time). So to review what all this means, Ω is the sample space, \mathcal{F} the σ -algebra of events, \mathbb{P} the probability measure on (Ω, \mathcal{F}) , and X_t is a random variable with values in \mathbb{R} for each $t \in T$. Next, we have a filtration $\mathfrak{F} = \{\mathcal{F}_t : t \in T\}$, and we assume that \mathbf{X} is *adapted* to \mathfrak{F} . To review again, \mathfrak{F} is an increasing family of sub σ -algebras of \mathcal{F} , so that $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ for $s, t \in T$ with $s \leq t$, and X_t is measurable with respect to \mathcal{F}_t for $t \in T$. We think of \mathcal{F}_t as the collection of events up to time $t \in T$, thus encoding the information available at time t . Finally, we assume that $\mathbb{E}(|X_t|) < \infty$, so that the mean of X_t exists as a real number, for each $t \in T$.

There are two important special cases of the basic setup. The simplest case, of course, is when $\mathcal{F}_t = \sigma\{X_s : s \in T, s \leq t\}$ for $t \in T$, so that \mathfrak{F} is the *natural filtration* associated with \mathbf{X} . Another case that arises frequently is when we have a second stochastic process $\mathbf{Y} = \{Y_t : t \in T\}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in a general measure space (S, \mathcal{S}) , and \mathfrak{F} is the natural filtration associated with \mathbf{Y} . So in this case, our main assumption is that X_t is measurable with respect to $\sigma\{Y_s : s \in T, s \leq t\}$ for $t \in T$.

The theory of martingales is beautiful, elegant, and mostly accessible in discrete time, when $T = \mathbb{N}$. But as with the theory of Markov processes, martingale theory is technically much more complicated in continuous time, when $T = [0, \infty)$. In this case, additional assumptions about the continuity of the sample paths $t \mapsto X_t$ and the filtration $t \mapsto \mathcal{F}_t$ are often necessary in order to have a nice theory. Specifically, we will assume that the process \mathbf{X} is right continuous and has left limits, and that the filtration \mathfrak{F} is right continuous and complete. These are the *standard assumptions* in continuous time.

Definitions

For the basic definitions that follow, you may need to review conditional expected value with respect to a σ -algebra.

The process \mathbf{X} is a *martingale* with respect to \mathfrak{F} if $\mathbb{E}(X_t | \mathcal{F}_s) = X_s$ for all $s, t \in T$ with $s \leq t$.

In the special case that \mathfrak{F} is the natural filtration associated with \mathbf{X} , we simply say that \mathbf{X} is a martingale, without reference to the filtration. In the special case that we have a second stochastic process $\mathbf{Y} = \{Y_t : t \in T\}$ and \mathfrak{F} is the natural filtration associated with \mathbf{Y} , we say that \mathbf{X} is a martingale with respect to \mathbf{Y} .

The term *martingale* originally referred to a portion of the harness of a horse, and was later used to describe gambling strategies, such as the one used in the Petersburg paradox, in which bets are doubled when a game is lost. To interpret the definitions above in terms of gambling, suppose that a gambler is at a casino, and that X_t represents her fortune at time $t \in T$ and \mathcal{F}_t the information available to her at time t . Suppose now that $s, t \in T$ with $s < t$ and that we think of s as the current time, so that t is a future time. If \mathbf{X} is a martingale with respect to \mathfrak{F} then the games are fair in the sense that the gambler's expected fortune at the future time t is the same as her current fortune at time s . To venture a bit from the casino, suppose that X_t is the price of a stock, or the value of a stock index, at time $t \in T$. If \mathbf{X} is a martingale, then the expected value at a future time, given all of our information, is the present value.



Figure 17.1.1: An English-style breastplate with a running martingale attachment. By Danielle M., CC BY 3.0, from Wikipedia

But as we will see, martingales are useful in probability far beyond the application to gambling and even far beyond financial applications generally. Indeed, martingales are of fundamental importance in modern probability theory. Here are two related definitions, with equality in the martingale condition replaced by inequalities.

Suppose again that the process \mathbf{X} and the filtration \mathfrak{F} satisfy the [basic assumptions](#) above.

1. \mathbf{X} is a *sub-martingale* with respect to \mathfrak{F} if $\mathbb{E}(X_t | \mathcal{F}_s) \geq X_s$ for all $s, t \in T$ with $s \leq t$.

2. \mathbf{X} is a *super-martingale* with respect to \mathfrak{F} if $\mathbb{E}(X_t | \mathcal{F}_s) \leq X_s$ for all $s, t \in T$ with $s \leq t$.

In the gambling setting, a sub-martingale models games that are favorable to the gambler on average, while a super-martingale models games that are unfavorable to the gambler on average. To venture again from the casino, suppose that X_t is the price of a stock, or the value of a stock index, at time $t \in T$. If \mathbf{X} is a sub-martingale, the expected value at a future time, given all of our information, is greater than the present value, and if \mathbf{X} is a super-martingale then the expected value at the future time is less than the present value. One hopes that a stock index is a sub-martingale.

Clearly \mathbf{X} is a martingale with respect to \mathfrak{F} if and only if it is both a sub-martingale and a super-martingale. Finally, recall that the conditional expected value of a random variable with respect to a σ -algebra is itself a random variable, and so the equations and inequalities in the definitions should be interpreted as holding with probability 1. In this section generally, statements involving random variables are assumed to hold with probability 1.

The conditions that define martingale, sub-martingale, and super-martingale make sense if the index set T is any totally ordered set. In some applications that we will consider later, $T = \{0, 1, \dots, n\}$ for fixed $n \in \mathbb{N}_+$. In the section on backwards martingales, $T = \{-n : n \in \mathbb{N}\}$ or $T = (-\infty, 0]$. In the case of discrete time when $T = \mathbb{N}$, we can simplify the definitions slightly.

Suppose that $\mathbf{X} = \{X_n : n \in \mathbb{N}\}$ satisfies the basic assumptions above.

1. \mathbf{X} is a martingale with respect to \mathfrak{F} if and only if $\mathbb{E}(X_{n+1} | \mathcal{F}_n) = X_n$ for all $n \in \mathbb{N}$.
2. \mathbf{X} is a sub-martingale with respect to \mathfrak{F} if and only if $\mathbb{E}(X_{n+1} | \mathcal{F}_n) \geq X_n$ for all $n \in \mathbb{N}$.
3. \mathbf{X} is a super-martingale with respect to \mathfrak{F} if and only if $\mathbb{E}(X_{n+1} | \mathcal{F}_n) \leq X_n$ for all $n \in \mathbb{N}$.

Proof

The conditions in the definitions clearly imply the conditions here, so we just need to show the opposite implications. Thus, assume that the condition in (a) holds and suppose that $k, n \in \mathbb{N}$ with $k < n$. Then $k \leq n-1$ so $\mathcal{F}_k \subseteq \mathcal{F}_{n-1}$ and hence

$$\mathbb{E}(X_n | \mathcal{F}_k) = \mathbb{E}[\mathbb{E}(X_n | \mathcal{F}_{n-1}) | \mathcal{F}_k] = \mathbb{E}(X_{n-1} | \mathcal{F}_k) \quad (17.1.1)$$

Repeating the argument, we get to

$$\mathbb{E}(X_n | \mathcal{F}_k) = \mathbb{E}(X_{k+1} | \mathcal{F}_k) = X_k \quad (17.1.2)$$

The proof for sub and super-martingales is analogous, with inequalities replacing the last equality.

The relations that define martingales, sub-martingales, and super-martingales hold for the ordinary (unconditional) expected values.

Suppose that $s, t \in T$ with $s \leq t$.

1. If \mathbf{X} is a martingale with respect to \mathfrak{F} then $\mathbb{E}(X_s) = \mathbb{E}(X_t)$.
2. If \mathbf{X} is a sub-martingale with respect to \mathfrak{F} then $\mathbb{E}(X_s) \leq \mathbb{E}(X_t)$.
3. If \mathbf{X} is a super-martingale with respect to \mathfrak{F} then $\mathbb{E}(X_s) \geq \mathbb{E}(X_t)$.

Proof

The results follow directly from the definitions, and the critical fact that $\mathbb{E}[\mathbb{E}(X_t | \mathcal{F}_s)] = \mathbb{E}(X_t)$ for $s, t \in T$.

So if \mathbf{X} is a martingale then \mathbf{X} has constant expected value, and this value is referred to as the mean of \mathbf{X} .

Examples

The goal for the remainder of this section is to give some classical examples of martingales, and by doing so, to show the wide variety of applications in which martingales occur. We will return to many of these examples in subsequent sections. Without further ado, we assume that all random variables are real-valued, unless otherwise specified, and that all expected values mentioned below exist in \mathbb{R} . Be sure to try the proofs yourself before expanding the ones in the text.

Constant Sequence

Our first example is rather trivial, but still worth noting.

Suppose that $\mathfrak{F} = \{\mathcal{F}_t : t \in T\}$ is a filtration on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and that X is a random variable that is measurable with respect to \mathcal{F}_0 and satisfies $\mathbb{E}(|X|) < \infty$. Let $X_t = X$ for $t \in T$. Then $\mathbf{X} = \{X_t : t \in T\}$ is a martingale with respect to \mathfrak{F} .

Proof

Since X is measurable with respect to \mathcal{F}_0 , it is measurable with respect to \mathcal{F}_t for all $t \in T$. Hence \mathbf{X} is adapted to \mathfrak{F} . If $s, t \in T$ with $s \leq t$, then $\mathbb{E}(X_t | \mathcal{F}_s) = \mathbb{E}(X | \mathcal{F}_s) = X = X_s$.

Partial Sums

For our next discussion, we start with one of the most basic martingales in discrete time, and the one with the simplest interpretation in terms of gambling. Suppose that $\mathbf{V} = \{V_n : n \in \mathbb{N}\}$ is a sequence of independent random variables with $\mathbb{E}(|V_k|) < \infty$ for $k \in \mathbb{N}$. Let

$$X_n = \sum_{k=0}^n V_k, \quad n \in \mathbb{N} \quad (17.1.3)$$

so that $\mathbf{X} = \{X_n : n \in \mathbb{N}\}$ is simply the *partial sum process* associated with \mathbf{V} .

For the partial sum process \mathbf{X} ,

1. If $\mathbb{E}(V_n) \geq 0$ for $n \in \mathbb{N}_+$ then \mathbf{X} is a sub-martingale.
2. If $\mathbb{E}(V_n) \leq 0$ for $n \in \mathbb{N}_+$ then \mathbf{X} is a super-martingale.
3. If $\mathbb{E}(V_n) = 0$ for $n \in \mathbb{N}_+$ then \mathbf{X} is a martingale.

Proof

Let $\mathcal{F}_n = \sigma\{X_0, X_1, \dots, X_n\} = \sigma\{V_0, V_1, \dots, V_n\}$ for $n \in \mathbb{N}$. Note first that

$$\mathbb{E}(|X_n|) \leq \sum_{k=0}^n \mathbb{E}(|V_k|) < \infty, \quad n \in \mathbb{N} \quad (17.1.4)$$

Next,

$$\mathbb{E}(X_{n+1} | \mathcal{F}_n) = \mathbb{E}(X_n + V_{n+1} | \mathcal{F}_n) = \mathbb{E}(X_n | \mathcal{F}_n) + \mathbb{E}(V_{n+1} | \mathcal{F}_n) = X_n + \mathbb{E}(V_{n+1}), \quad n \in \mathbb{N} \quad (17.1.5)$$

The last equality holds since X_n is measurable with respect to \mathcal{F}_n and V_{n+1} is independent of \mathcal{F}_n . The results now follow from the definitions.

In terms of gambling, if $X_0 = V_0$ is the gambler's initial fortune and V_i is the gambler's net winnings on the i th game, then X_n is the gamblers net fortune after n games for $n \in \mathbb{N}_+$. But partial sum processes associated with independent sequences are important far beyond gambling. In fact, much of classical probability deals with partial sums of independent and *identically distributed* variables. The entire chapter on Random Samples explores this setting.

Note that $\mathbb{E}(X_n) = \sum_{k=0}^n \mathbb{E}(V_k)$. Hence condition (a) is equivalent to $n \mapsto \mathbb{E}(X_n)$ increasing, condition (b) is equivalent to $n \mapsto \mathbb{E}(X_n)$ decreasing, and condition (c) is equivalent to $n \mapsto \mathbb{E}(X_n)$ constant. Here is another martingale associated with the partial sum process, known as the *second moment martingale*.

Suppose that $\mathbb{E}(V_k) = 0$ for $k \in \mathbb{N}_+$ and $\text{var}(V_k) < \infty$ for $k \in \mathbb{N}$. Let

$$Y_n = X_n^2 - \text{var}(X_n), \quad n \in \mathbb{N} \quad (17.1.6)$$

Then $\mathbf{Y} = \{Y_n : n \in \mathbb{N}\}$ is a martingale with respect to \mathbf{X} .

Proof

Again, let $\mathcal{F}_n = \sigma\{X_0, X_1, \dots, X_n\}$ for $n \in \mathbb{N}$. Since the sequence \mathbf{V} is independent, note that

$$\text{var}(X_n) = \text{var}\left(\sum_{k=0}^n V_k\right) = \sum_{k=0}^n \text{var}(V_k) \quad (17.1.7)$$

Also, $\text{var}(V_k) = \mathbb{E}(V_k^2)$ since $\mathbb{E}(V_k) = 0$ for $k \in \mathbb{N}_+$. In particular, $\mathbb{E}(|Y_n|) < \infty$ for $n \in \mathbb{N}$. Next for $n \in \mathbb{N}$,

$$\begin{aligned} \mathbb{E}(Y_{n+1} | \mathcal{F}_n) &= \mathbb{E}[X_{n+1}^2 - \text{var}(X_{n+1}) | \mathcal{F}_n] = \mathbb{E}[(X_n + V_{n+1})^2 - \text{var}(X_{n+1}) | \mathcal{F}_n] \\ &= \mathbb{E}[X_n^2 + 2X_n V_{n+1} + V_{n+1}^2 - \text{var}(X_{n+1}) | \mathcal{F}_n] = X_n^2 + 2X_n \mathbb{E}(V_{n+1}) + \mathbb{E}(V_{n+1}^2) - \text{var}(X_{n+1}) \end{aligned}$$

since X_n is measurable with respect to \mathcal{F}_n and V_{n+1} is independent of \mathcal{F}_n . But $\mathbb{E}(V_{n+1}) = 0$ and $\mathbb{E}(V_{n+1}^2) - \text{var}(X_{n+1}) = -\text{var}(X_n)$. Hence we have $\mathbb{E}(Y_{n+1} | \mathcal{F}_n) = X_n^2 - \text{var}(X_n) = Y_n$ for $n \in \mathbb{N}$.

So under the assumptions in this theorem, both \mathbf{X} and \mathbf{Y} are martingales. We will generalize the results for partial sum processes below in the discussion on [processes with independent increments](#).

Martingale Difference Sequences

In the last discussion, we saw that the partial sum process associated with a sequence of independent, mean 0 variables is a martingale. Conversely, every martingale in discrete time can be written as a partial sum process of *uncorrelated* mean 0 variables. This representation gives some significant insight into the theory of martingales generally. Suppose that $\mathbf{X} = \{X_n : n \in \mathbb{N}\}$ is a martingale with respect to the filtration $\mathcal{F} = \{\mathcal{F}_n : n \in \mathbb{N}\}$.

Let $V_0 = X_0$ and $V_n = X_n - X_{n-1}$ for $n \in \mathbb{N}_+$. The process $\mathbf{V} = \{V_n : n \in \mathbb{N}\}$ is the *martingale difference sequence* associated with \mathbf{X} and

$$X_n = \sum_{k=0}^n V_k, \quad n \in \mathbb{N} \quad (17.1.8)$$

As promised, the martingale difference variables have mean 0, and in fact satisfy a stronger property.

Suppose that $\mathbf{V} = \{V_n : n \in \mathbb{N}\}$ is the martingale difference sequence associated with \mathbf{X} . Then

1. \mathbf{V} is adapted to \mathcal{F} .
2. $\mathbb{E}(V_n | \mathcal{F}_k) = 0$ for $k, n \in \mathbb{N}$ with $k < n$.
3. $\mathbb{E}(V_n) = 0$ for $n \in \mathbb{N}_+$.

Proof

1. Of course $V_0 = X_0$ is measurable with respect to \mathcal{F}_0 . For $n \in \mathbb{N}_+$, X_n and X_{n-1} , and hence V_n are measurable with respect to \mathcal{F}_n .
2. Let $k \in \mathbb{N}$. By the martingale and adapted properties,

$$\mathbb{E}(V_{k+1} | \mathcal{F}_k) = \mathbb{E}(X_{k+1} | \mathcal{F}_k) - \mathbb{E}(X_k | \mathcal{F}_k) = X_k - X_k = 0 \quad (17.1.9)$$

Next by the tower property,

$$\mathbb{E}(V_{k+2} | \mathcal{F}_k) = \mathbb{E}[\mathbb{E}(V_{k+2} | \mathcal{F}_{k+1}) | \mathcal{F}_k] = 0 \quad (17.1.10)$$

Continuing (or using induction) gives the general result.

3. Since \mathbf{X} is a martingale, it has constant mean, as noted above. Hence $\mathbb{E}(V_n) = \mathbb{E}(X_n) - \mathbb{E}(X_{n-1}) = 0$ for $n \in \mathbb{N}_+$. We could also use part (b).

Also as promised, if the martingale variables have finite variance, then the martingale difference variables are uncorrelated.

Suppose again that $\mathbf{V} = \{V_n : n \in \mathbb{N}\}$ is the martingale difference sequence associated with the martingale \mathbf{X} . Assume that $\text{var}(X_n) < \infty$ for $n \in \mathbb{N}$. Then \mathbf{V} is an uncorrelated sequence. Moreover,

$$\text{var}(X_n) = \sum_{k=0}^n \text{var}(V_k) = \text{var}(X_0) + \sum_{k=1}^n \mathbb{E}(V_k^2), \quad n \in \mathbb{N} \quad (17.1.11)$$

Proof

Let $k, n \in \mathbb{N}$ with $k < n$. To show that V_k and V_n are uncorrelated, we just need to show that $\mathbb{E}(V_k V_n) = 0$ (since $\mathbb{E}(V_n) = 0$). But by the previous result,

$$\mathbb{E}(V_k V_n) = \mathbb{E}[\mathbb{E}(V_k V_n | \mathcal{F}_k)] = \mathbb{E}[V_k \mathbb{E}(V_n | \mathcal{F}_k)] = 0 \quad (17.1.12)$$

Finally, the variance of a sum of uncorrelated variables is the sum of the variances. Since V_k has mean 0, $\text{var}(V_k) = \mathbb{E}(V_k^2)$ for $k \in \mathbb{N}_+$. Hence the formula for $\text{var}(X_n)$ holds.

We now know that a discrete-time martingale is the partial sum process associated with a sequence of uncorrelated variables. Hence we might hope that there are martingale versions of the fundamental theorems that hold for a partial sum process associated with an independent sequence. This turns out to be true, and is a basic reason for the importance of martingales.

Discrete-Time Random Walks

Suppose that $\mathbf{V} = \{V_n : n \in \mathbb{N}\}$ is a sequence of independent random variables with $\{V_n : n \in \mathbb{N}_+\}$ identically distributed. We assume that $\mathbb{E}(|V_n|) < \infty$ for $n \in \mathbb{N}$ and we let a denote the common mean of $\{V_n : n \in \mathbb{N}_+\}$. Let $\mathbf{X} = \{X_n : n \in \mathbb{N}\}$ be the partial sum process associated with \mathbf{V} so that

$$X_n = \sum_{i=0}^n V_i, \quad n \in \mathbb{N} \quad (17.1.13)$$

This setting is a special case of the more general partial sum process considered [above](#). The process \mathbf{X} is sometimes called a (discrete-time) *random walk*. The initial position $X_0 = V_0$ of the walker can have an arbitrary distribution, but then the *steps* that the walker takes are independent and identically distributed. In terms of gambling, $X_0 = V_0$ is the initial fortune of the gambler playing a sequence of independent and identical games. If V_i is the amount won (or lost) on game $i \in \mathbb{N}_+$, then X_n is the gambler's net fortune after n games.

For the random walk \mathbf{X} ,

1. \mathbf{X} is a martingale if $a = 0$.
2. \mathbf{X} is a sub-martingale if $a \geq 0$.
3. \mathbf{X} is a super-martingale if $a \leq 0$.

For the second moment martingale, suppose that V_n has common mean $a = 0$ and common variance $b^2 < \infty$ for $n \in \mathbb{N}_+$, and that $\text{var}(V_0) < \infty$.

Let $Y_n = X_n^2 - \text{var}(V_0) - b^2 n$ for $n \in \mathbb{N}$. Then $\mathbf{Y} = \{Y_n : n \in \mathbb{N}\}$ is a martingale with respect to \mathbf{X} .

Proof

This follows from the corresponding result for a general partial sum process, above, since

$$\text{var}(X_n) = \sum_{k=0}^n \text{var}(V_k) = \text{var}(V_0) + b^2 n, \quad n \in \mathbb{N} \quad (17.1.14)$$

We will generalize the results for discrete-time random walks below, in the discussion on [processes with stationary, independent increments](#).

Partial Products

Our next discussion is similar to the one on partial sum processes [above](#), but with products instead of sums. So suppose that $\mathbf{V} = \{V_n : n \in \mathbb{N}\}$ is an independent sequence of nonnegative random variables with $\mathbb{E}(V_n) < \infty$ for $n \in \mathbb{N}$. Let

$$X_n = \prod_{i=0}^n V_i, \quad n \in \mathbb{N} \quad (17.1.15)$$

so that $\mathbf{X} = \{X_n : n \in \mathbb{N}\}$ is the *partial product process* associated with \mathbf{X} .

For the partial product process \mathbf{X} ,

1. If $\mathbb{E}(V_n) = 1$ for $n \in \mathbb{N}_+$ then \mathbf{X} is a martingale with respect to \mathbf{V}
2. If $\mathbb{E}(V_n) \geq 1$ for $n \in \mathbb{N}_+$ then \mathbf{X} is a sub-martingale with respect to \mathbf{V}
3. If $\mathbb{E}(V_n) \leq 1$ for $n \in \mathbb{N}_+$ then \mathbf{X} is a super-martingale with respect to \mathbf{V}

Proof

Let $\mathcal{F}_n = \sigma\{V_0, V_1, \dots, V_n\}$ for $n \in \mathbb{N}$. Since the variables are independent,

$$\mathbb{E}(X_n) = \prod_{i=0}^n \mathbb{E}(V_i) < \infty \quad (17.1.16)$$

Next,

$$\mathbb{E}(X_{n+1} | \mathcal{F}_n) = \mathbb{E}(X_n V_{n+1} | \mathcal{F}_n) = X_n \mathbb{E}(V_{n+1} | \mathcal{F}_n) = X_n \mathbb{E}(V_{n+1}) \quad n \in \mathbb{N} \quad (17.1.17)$$

since X_n is measurable with respect to \mathcal{F}_n and V_{n+1} is independent of \mathcal{F}_n . The results now follow from the definitions.

As with random walks, a special case of interest is when $\{V_n : n \in \mathbb{N}_+\}$ is an identically distributed sequence.

The Simple Random Walk

Suppose now that that $\mathbf{V} = \{V_n : n \in \mathbb{N}\}$ is a sequence of independent random variables with $\mathbb{P}(V_i = 1) = p$ and $\mathbb{P}(V_i = -1) = 1 - p$ for $i \in \mathbb{N}_+$, where $p \in (0, 1)$. Let $\mathbf{X} = \{X_n : n \in \mathbb{N}\}$ be the partial sum process associated with \mathbf{V} so that

$$X_n = \sum_{i=0}^n V_i, \quad n \in \mathbb{N} \quad (17.1.18)$$

Then \mathbf{X} is the simple random walk with parameter p , and of course, is a special case of the more general random walk studied [above](#). In terms of gambling, our gambler plays a sequence of independent and identical games, and on each game, wins €1 with probability p and loses €1 with probability $1 - p$. So if V_0 is the gambler's initial fortune, then X_n is her net fortune after n games.

For the simple random walk,

1. If $p > \frac{1}{2}$ then \mathbf{X} is a sub-martingale.
2. If $p < \frac{1}{2}$ then \mathbf{X} is a super-martingale.
3. If $p = \frac{1}{2}$ then \mathbf{X} is a martingale.

Proof

Note that $\mathbb{E}(V_n) = p - (1 - p) = 2p - 1$ for $n \in \mathbb{N}_+$, so the results follow from the theorem [above](#).

So case (a) corresponds to favorable games, case (b) to unfavorable games, and case (c) to fair games.

Open the simulation of the simple symmetric random. For various values of the number of trials n , run the simulation 1000 times and note the general behavior of the sample paths.

Here is the second moment martingale for the simple, symmetric random walk.

Consider the simple random walk with parameter $p = \frac{1}{2}$, and let $Y_n = X_n^2 - \text{var}(V_0) - n$ for $n \in \mathbb{N}$. Then $\mathbf{Y} = \{Y_n : n \in \mathbb{N}\}$ is a martingale with respect to \mathbf{X}

Proof

Note that $\mathbb{E}(V_i) = 0$ and $\text{var}(V_i) = 1$ for each $i \in \mathbb{N}_+$, so the result follows from the general result above.

But there is another martingale that can be associated with the simple random walk, known as *De Moivre's martingale* and named for one of the early pioneers of probability theory, Abraham De Moivre.

For $n \in \mathbb{N}$ define

$$Z_n = \left(\frac{1-p}{p} \right)^{X_n} \quad (17.1.19)$$

Then $\mathbf{Z} = \{Z_n : n \in \mathbb{N}\}$ is a martingale with respect to \mathbf{X} .

Proof

Note that

$$Z_n = \prod_{k=0}^n \left(\frac{1-p}{p} \right)^{V_k}, \quad n \in \mathbb{N} \quad (17.1.20)$$

and

$$\mathbb{E} \left[\left(\frac{1-p}{p} \right)^{V_k} \right] = \left(\frac{1-p}{p} \right)^1 p + \left(\frac{1-p}{p} \right)^{-1} (1-p) = 1, \quad k \in \mathbb{N}_+ \quad (17.1.21)$$

So the result follows from the theorem [above](#) on partial products.

The Beta-Bernoulli Process

Recall that the beta-Bernoulli process is constructed by randomizing the success parameter in a Bernoulli trials process with a beta distribution. Specifically we have a random variable P that has the beta distribution with parameters $a, b \in (0, \infty)$, and a sequence of indicator variables $\mathbf{X} = (X_1, X_2, \dots)$ such that given $P = p \in (0, 1)$, \mathbf{X} is a sequence of independent variables with $\mathbb{P}(X_i = 1) = p$ for $i \in \mathbb{N}_+$. As usual, we couch this in reliability terms, so that $X_i = 1$ means *success* on trial i and $X_i = 0$ means *failure*. In our study of this process, we showed that the finite-dimensional distributions are given by

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \frac{a^{[k]} b^{[n-k]}}{(a+b)^{[n]}}, \quad n \in \mathbb{N}_+, (x_1, x_2, \dots, x_n) \in \{0, 1\}^n \quad (17.1.22)$$

where we use the ascending power notation $r^{[j]} = r(r+1) \cdots (r+j-1)$ for $r \in \mathbb{R}$ and $j \in \mathbb{N}$. Next, let $\mathbf{Y} = \{Y_n : n \in \mathbb{N}\}$ denote the partial sum process associated with \mathbf{X} , so that once again,

$$Y_n = \sum_{i=1}^n X_i, \quad n \in \mathbb{N} \quad (17.1.23)$$

Of course Y_n is the number of success in the first n trials and has the beta-binomial distribution defined by

$$\mathbb{P}(Y_n = k) = \binom{n}{k} \frac{a^{[k]} b^{[n-k]}}{(a+b)^{[n]}}, \quad k \in \{0, 1, \dots, n\} \quad (17.1.24)$$

Now let

$$Z_n = \frac{a + Y_n}{a + b + n}, \quad n \in \mathbb{N} \quad (17.1.25)$$

This variable also arises naturally. Let $\mathcal{F}_n = \sigma\{X_1, X_2, \dots, X_n\}$ for $n \in \mathbb{N}$. Then as shown in the section on the beta-Bernoulli process, $Z_n = \mathbb{E}(X_{n+1} | \mathcal{F}_n) = \mathbb{E}(P | \mathcal{F}_n)$. In statistical terms, the second equation means that Z_n is the *Bayesian estimator* of the unknown success probability p in a sequence of Bernoulli trials, when p is modeled by the random variable P .

$\mathbf{Z} = \{Z_n : n \in \mathbb{N}\}$ is a martingale with respect to \mathbf{X} .

Proof

Note that $0 \leq Z_n \leq 1$ so $\mathbb{E}(Z_n) < \infty$ for $n \in \mathbb{N}$. Next,

$$\mathbb{E}(Z_{n+1} | \mathcal{F}_n) = \mathbb{E}\left[\frac{a + Y_{n+1}}{a + b + n + 1} \mid \mathcal{F}_n\right] = \frac{\mathbb{E}[a + (Y_n + X_{n+1}) | \mathcal{F}_n]}{a + b + n + 1} = \frac{a + Y_n + \mathbb{E}(X_{n+1} | \mathcal{F}_n)}{a + b + n + 1} \quad (17.1.26)$$

As noted above, $\mathbb{E}(X_{n+1} | \mathcal{F}_n) = (a + Y_n)/(a + b + n)$. Substituting into the displayed equation above and doing a bit of algebra we have

$$\mathbb{E}(Z_{n+1} | \mathcal{F}_n) = \frac{(a + Y_n) + (a + Y_n)/(a + b + n)}{a + b + n + 1} = \frac{a + Y_n}{a + b + n} = Z_n \quad (17.1.27)$$

Open the beta-Binomial experiment. Run the simulation 1000 times for various values of the parameters, and compare the empirical probability density function with the true probability density function.

Pólya's Urn Process

Recall that in the simplest version of Pólya's urn process, we start with an urn containing a red and b green balls. At each discrete time step, we select a ball at random from the urn and then replace the ball and add c new balls of the same color to the urn. For the parameters, we need $a, b \in \mathbb{N}_+$ and $c \in \mathbb{N}$. For $i \in \mathbb{N}_+$, let X_i denote the color of the ball selected on the i th draw, where 1 means red and 0 means green. The process $\mathbf{X} = \{X_n : n \in \mathbb{N}_+\}$ is a classical example of a sequence of exchangeable yet dependent variables. Let $\mathbf{Y} = \{Y_n : n \in \mathbb{N}\}$ denote the partial sum process associated with \mathbf{X} , so that once again,

$$Y_n = \sum_{i=1}^n X_i, \quad n \in \mathbb{N} \quad (17.1.28)$$

Of course Y_n is the total number of red balls selected in the first n draws. Hence at time $n \in \mathbb{N}$, the total number of red balls in the urn is $a + cY_n$, while the total number of balls in the urn is $a + b + cn$ and so the *proportion* of red balls in the urn is

$$Z_n = \frac{a + cY_n}{a + b + cn} \quad (17.1.29)$$

$\mathbf{Z} = \{Z_n : n \in \mathbb{N}\}$ is a martingale with respect to \mathbf{X} .

Indirect proof

If $c = 0$ then $Z_n = a/(a+b)$ for $n \in \mathbb{N}$ so \mathbf{Z} is a constant martingale. If $c \in \mathbb{N}_+$ then \mathbf{Z} is equivalent to the beta-Bernoulli process with parameters a/c and b/c . Moreover,

$$Z_n = \frac{a + cY_n}{a + b + cn} = \frac{a/c + Y_n}{a/c + b/c + n}, \quad n \in \mathbb{N} \quad (17.1.30)$$

So \mathbf{Z} is a martingale by the previous theorem.

Direct Proof

Trivially, $0 \leq Z_n \leq 1$ so $\mathbb{E}(Z_n) < \infty$ for $n \in \mathbb{N}$. Let $\mathcal{F}_n = \sigma\{X_1, X_2, \dots, X_n\}$. For $n \in \mathbb{N}$,

$$\mathbb{E}(Z_{n+1} | \mathcal{F}_n) = \mathbb{E}\left[\frac{a + cY_{n+1}}{a + b + c(n+1)} \mid \mathcal{F}_n\right] = \frac{\mathbb{E}[a + c(Y_n + X_{n+1}) | \mathcal{F}_n]}{a + b + c(n+1)} = \frac{a + cY_n + c\mathbb{E}(X_{n+1} | \mathcal{F}_n)}{a + b + cn + c} \quad (17.1.31)$$

since Y_n is measurable with respect to \mathcal{F}_n . But the probability of selecting a red ball on draw $n+1$, given the history of the process up to time n , is simply the proportion of red balls in the urn at time n . That is,

$$\mathbb{E}(X_{n+1} | \mathcal{F}_n) = \mathbb{P}(X_{n+1} = 1 | \mathcal{F}_n) = Z_n = \frac{a + cY_n}{a + b + cn} \quad (17.1.32)$$

Substituting and simplifying gives $\mathbb{E}(Z_{n+1} | \mathcal{F}_n) = Z_n$.

Open the simulation of Pólya's Urn Experiment. Run the simulation 1000 times for various values of the parameters, and compare the empirical probability density function of the number of red ball selected to the true probability density function.

Processes with Independent Increments.

Our first example above concerned the [partial sum process](#) \mathbf{X} associated with a sequence of independent random variables \mathbf{V} . Such processes are the only ones in discrete time that have *independent increments*. That is, for $m, n \in \mathbb{N}$ with $m \leq n$, $X_n - X_m$ is independent of (X_0, X_1, \dots, X_m) . The [random walk process](#) has the additional property of *stationary increments*. That is, the distribution of $X_n - X_m$ is the same as the distribution of $X_{n-m} - X_0$ for $m, n \in \mathbb{N}$ with $m \leq n$. Let's consider processes in discrete or continuous time with these properties. Thus, suppose that $\mathbf{X} = \{X_t : t \in T\}$ satisfying the [basic assumptions](#) above relative to the filtration $\mathcal{F} = \{\mathcal{F}_s : s \in T\}$. Here are the two definitions.

The process \mathbf{X} has

1. *Independent increments* if $X_t - X_s$ is independent of \mathcal{F}_s for all $s, t \in T$ with $s \leq t$.
2. *Stationary increments* if $X_t - X_s$ has the same distribution as $X_{t-s} - X_0$ for all $s, t \in T$.

Processes with stationary and independent increments were studied in the Chapter on Markov processes. In continuous time (with the continuity assumptions we have imposed), such a process is known as a *Lévy process*, named for Paul Lévy, and also as a *continuous-time random walk*. For a process with independent increments (not necessarily stationary), the connection with martingales depends on the mean function m given by $m(t) = \mathbb{E}(X_t)$ for $t \in T$.

Suppose that $\mathbf{X} = \{X_t : t \in [0, \infty)\}$ has independent increments.

1. If m is increasing then \mathbf{X} is a sub-martingale.
2. If m is decreasing then \mathbf{X} is a super-martingale.
3. If m is constant then \mathbf{X} is a martingale

Proof

The proof is just like the one above for partial sum processes. Suppose that $s, t \in [0, \infty)$ with $s < t$. Then

$$\mathbb{E}(X_t | \mathcal{F}_s) = \mathbb{E}[X_s + (X_t - X_s) | \mathcal{F}_s] = \mathbb{E}(X_s | \mathcal{F}_s) + \mathbb{E}(X_t - X_s | \mathcal{F}_s) \quad (17.1.33)$$

But X_s is measurable with respect to \mathcal{F}_s and $X_t - X_s$ is independent of \mathcal{F}_s . So

$$\mathbb{E}(X_t | \mathcal{F}_s) = X_s + \mathbb{E}(X_t - X_s) = X_s + m(t) - m(s) \quad (17.1.34)$$

Compare this theorem with the corresponding theorem for the [partial sum process](#) above. Suppose now that $\mathbf{X} = \{X_t : t \in [0, \infty)\}$ is a stochastic process as above, with mean function m , and let $Y_t = X_t - m(t)$ for $t \in [0, \infty)$. The process $\mathbf{Y} = \{Y_t : t \in [0, \infty)\}$ is sometimes called the *compensated process* associated with \mathbf{X} and has mean function 0. If \mathbf{X} has independent increments, then clearly so does \mathbf{Y} . Hence the following result is a trivial corollary to our previous theorem.

Suppose that \mathbf{X} has independent increments. The compensated process \mathbf{Y} is a martingale.

Next we give the *second moment martingale* for a process with independent increments, generalizing the [second moment martingale](#) for a partial sum process.

Suppose that $\mathbf{X} = \{X_t : t \in T\}$ has independent increments with constant mean function and with $\text{var}(X_t) < \infty$ for $t \in T$. Then $\mathbf{Y} = \{Y_t : t \in T\}$ is a martingale where

$$Y_t = X_t^2 - \text{var}(X_t), \quad t \in T \quad (17.1.35)$$

Proof

The proof is essentially the same as for the partial sum process in discrete time. Suppose that $s, t \in T$ with $s < t$. Note that $\mathbb{E}(Y_t | \mathcal{F}_s) = \mathbb{E}(X_t^2 | \mathcal{F}_s) - \text{var}(X_t)$. Next,

$$X_t^2 = [(X_t - X_s) + X_s]^2 = (X_t - X_s)^2 + 2(X_t - X_s)X_s + X_s^2 \quad (17.1.36)$$

But $X_t - X_s$ is independent of \mathcal{F}_s , X_s is measurable with respect to \mathcal{F}_s , and $\mathbb{E}(X_t - X_s) = 0$ so

$$\mathbb{E}(X_t^2 | \mathcal{F}_s) = \mathbb{E}[(X_t - X_s)^2] + 2X_s \mathbb{E}(X_t - X_s) + X_s^2 = \mathbb{E}[(X_t - X_s)^2] + X_s^2 \quad (17.1.37)$$

But also by independence and since $X_t - X_s$ has mean 0,

$$\text{var}(X_t) = \text{var}[(X_t - X_s) + X_s] = \text{var}(X_s) + \text{var}(X_t - X_s)^2 = \text{var}(X_s) + \mathbb{E}[(X_t - X_s)^2] \quad (17.1.38)$$

Putting the pieces together gives

$$\mathbb{E}(Y_t | \mathcal{F}_s) = X_s^2 - \text{var}(X_s) = Y_s \quad (17.1.39)$$

Of course, since the mean function is constant, \mathbf{X} is also a martingale. For processes with independent *and* stationary increments (that is, random walks), the last two theorems simplify, because the mean and variance functions simplify.

Suppose that $\mathbf{X} = \{X_t : t \in T\}$ has stationary, independent increments, and let $a = \mathbb{E}(X_1 - X_0)$. Then

1. \mathbf{X} is a martingale if $a = 0$
2. \mathbf{X} is a sub-martingale if $a \geq 0$
3. \mathbf{X} is a super-martingale if $a \leq 0$

Proof

Recall that the mean function m is given by $m(t) = \mathbb{E}(X_0) + at$ for $t \in T$, so the result follows from the [corresponding result](#) for a process with independent increments.

Compare this result with the [corresponding one](#) above for discrete-time random walks. Our next result is the second moment martingale. Compare this with the [second moment martingale](#) for discrete-time random walks.

Suppose that $\mathbf{X} = \{X_t : t \in T\}$ has stationary, independent increments with $\mathbb{E}(X_0) = \mathbb{E}(X_1)$ and $b^2 = \mathbb{E}(X_1^2) < \infty$. Then $\mathbf{Y} = \{Y_t : t \in T\}$ is a martingale where

$$Y_t = X_t^2 - \text{var}(X_0) - b^2 t, \quad t \in T \quad (17.1.40)$$

Proof

Recall that if $\mathbb{E}(X_0) = \mathbb{E}(X_1)$ then \mathbf{X} has constant mean function. Also, $\text{var}(X_t) = \text{var}(X_0) + b^2 t$, so the result follows from the corresponding result for a process with independent increments.

In discrete time, as we have mentioned several times, all of these results reduce to the earlier results for [partial sum processes](#) and [random walks](#). In continuous time, the Poisson processes, named of course for Simeon Poisson, provides examples. The standard (homogeneous) Poisson counting process $\mathbf{N} = \{N_t : t \in [0, \infty)\}$ with constant rate $r \in (0, \infty)$ has stationary, independent increments and mean function given by $m(t) = rt$ for $t \in [0, \infty)$. More generally, suppose that $r : [0, \infty) \rightarrow (0, \infty)$ is piecewise continuous (and non-constant). The non-homogeneous Poisson counting process $\mathbf{N} = \{N_t : t \in [0, \infty)\}$ with rate function r has independent increments and mean function given by

$$m(t) = \int_0^t r(s) ds, \quad t \in [0, \infty) \quad (17.1.41)$$

The increment $N_t - N_s$ has the Poisson distribution with parameter $m(t) - m(s)$ for $s, t \in [0, \infty)$ with $s < t$, so the process does not have stationary increments. In all cases, m is increasing, so the following results are corollaries of our general results:

Let $\mathbf{N} = \{N_t : t \in [0, \infty)\}$ be the Poisson counting process with rate function $r : [0, \infty) \rightarrow (0, \infty)$. Then

1. \mathbf{N} is a sub-martingale
2. The compensated process $\mathbf{X} = \{N_t - m(t) : t \in [0, \infty)\}$ is a martingale.

Open the simulation of the Poisson counting experiment. For various values of r and t , run the experiment 1000 times and compare the empirical probability density function of the number of arrivals with the true probability density function.

We will see further examples of processes with stationary, independent increments in continuous time (and so also examples of continuous-time martingales) in our study of Brownian motion.

Likelihood Ratio Tests

Suppose that (S, \mathcal{S}, μ) is a general measure space, and that $\mathbf{X} = \{X_n : n \in \mathbb{N}\}$ is a sequence of independent, identically distributed random variables, taking values in S . In statistical terms, \mathbf{X} corresponds to *sampling* from the common distribution, which is usually not completely known. Indeed, the central problem in statistics is to draw inferences about the distribution from observations of \mathbf{X} . Suppose now that the underlying distribution either has probability density function g_0 or probability density function g_1 , with respect to μ . We assume that g_0 and g_1 are positive on S . Of course the common special cases of this setup are

- S is a measurable subset of \mathbb{R}^n for some $n \in \mathbb{N}_+$ and $\mu = \lambda_n$ is n -dimensional Lebesgue measure on S .
- S is a countable set and $\mu = \#$ is counting measure on S .

The likelihood ratio test is a hypothesis test, where the null and alternative hypotheses are

- H_0 : the probability density function is g_0 .
- H_1 : the probability density function is g_1 .

The test is based on the test statistic

$$L_n = \prod_{i=1}^n \frac{g_0(X_i)}{g_1(X_i)}, \quad n \in \mathbb{N} \quad (17.1.42)$$

known as the *likelihood ratio test statistic*. Small values of the test statistic are evidence in favor of the alternative hypothesis H_1 . Here is our result.

Under the alternative hypothesis H_1 , the process $\mathbf{L} = \{L_n : n \in \mathbb{N}\}$ is a martingale with respect to \mathbf{X} , known as the *likelihood ratio martingale*.

Proof

Let $\mathcal{F}_n = \sigma\{X_1, X_2, \dots, X_n\}$. For $n \in \mathbb{N}$,

$$\mathbb{E}(L_{n+1} \mid \mathcal{F}_n) = \mathbb{E}\left[L_n \frac{g_0(X_{n+1})}{g_1(X_{n+1})} \mid \mathcal{F}_n\right] = L_n \mathbb{E}\left[\frac{g_0(X_{n+1})}{g_1(X_{n+1})}\right] \quad (17.1.43)$$

Since L_n is measurable with respect to \mathcal{F}_n and $g_0(X_{n+1})/g_1(X_{n+1})$ is independent of \mathcal{F}_n . But under H_1 , and using the change of variables formula for expected value, we have

$$\mathbb{E}\left[\frac{g_0(X_{n+1})}{g_1(X_{n+1})}\right] = \int_S \frac{g_0(x)}{g_1(x)} g_1(x) d\mu(x) = \int_S g_0(x) d\mu(x) = 1 \quad (17.1.44)$$

This result also follows essentially from the theorem [above](#) on partial products. The sequence $\mathbf{Z} = (Z_1, Z_2, \dots)$ given by $Z_i = g_0(X_i)/g_1(X_i)$ for $i \in \mathbb{N}_+$ is independent and identically distributed, and as just shown, has mean 1 under H_1 .

Branching Processes

In the simplest model of a *branching process*, we have a system of *particles* each of which can die out or split into new particles of the same type. The fundamental assumption is that the particles act independently, each with the same offspring distribution on \mathbb{N} . We will let f denote the (discrete) probability density function of the number of offspring of a particle, m the mean of the distribution, and ϕ the probability generating function of the distribution. Thus, if U is the number of children of a particle, then $f(n) = \mathbb{P}(U = n)$ for $n \in \mathbb{N}$, $m = \mathbb{E}(U)$, and $\phi(t) = \mathbb{E}(t^U)$ defined at least for $t \in (-1, 1]$.

Our interest is in *generational time* rather than *absolute time*: the original particles are in generation 0, and recursively, the children a particle in generation n belong to generation $n+1$. Thus, the stochastic process of interest is $\mathbf{X} = \{X_n : n \in \mathbb{N}\}$ where X_n is the

number of particles in the n th generation for $n \in \mathbb{N}$. The process \mathbf{X} is a Markov chain and was studied in the section on discrete-time branching chains. In particular, one of the fundamental problems is to compute the probability q of extinction starting with a single particle:

$$q = \mathbb{P}(X_n = 0 \text{ for some } n \in \mathbb{N} \mid X_0 = 1) \quad (17.1.45)$$

Then, since the particles act independently, the probability of extinction starting with $x \in \mathbb{N}$ particles is simply q^x . We will assume that $f(0) > 0$ and $f(0) + f(1) < 1$. This is the interesting case, since it means that a particle has a positive probability of dying without children and a positive probability of producing more than 1 child. The fundamental result, you may recall, is that q is the smallest fixed point of ϕ (so that $\phi(q) = q$) in the interval $[0, 1]$. Here are two martingales associated with the branching process:

Each of the following is a martingale with respect to \mathbf{X} .

1. $\mathbf{Y} = \{Y_n : n \in \mathbb{N}\}$ where $Y_n = X_n/m^n$ for $n \in \mathbb{N}$.
2. $\mathbf{Z} = \{Z_n : n \in \mathbb{N}\}$ where $Z_n = q^{X_n}$ for $n \in \mathbb{N}$.

Proof

Let $\mathcal{F}_n = \sigma\{X_0, X_1, \dots, X_n\}$. For $n \in \mathbb{N}$, note that X_{n+1} can be written in the form

$$X_{n+1} = \sum_{i=1}^{X_n} U_i \quad (17.1.46)$$

where $\mathbf{U} = (U_1, U_2, \dots)$ is a sequence of independent variables, each with PDF f (and hence mean μ and PGF ϕ), and with \mathbf{U} independent of \mathcal{F}_n . Think of U_i as the number of children of the i th particle in generation n .

1. For $n \in \mathbb{N}$,

$$\mathbb{E}(Y_{n+1} \mid \mathcal{F}_n) = \mathbb{E}\left(\frac{X_{n+1}}{m^{n+1}} \mid \mathcal{F}_n\right) = \frac{1}{m^{n+1}} \mathbb{E}\left(\sum_{i=1}^{X_n} U_i \mid \mathcal{F}_n\right) = \frac{1}{m^{n+1}} m X_n = \frac{X_n}{m^n} = Y_n \quad (17.1.47)$$

2. For $n \in \mathbb{N}$

$$\mathbb{E}(Z_{n+1} \mid \mathcal{F}_n) = \mathbb{E}(q^{X_{n+1}} \mid \mathcal{F}_n) = \mathbb{E}\left(q^{\sum_{i=1}^{X_n} U_i} \mid \mathcal{F}_n\right) = [\phi(q)]^{X_n} = q^{X_n} = Z_n \quad (17.1.48)$$

Doob's Martingale

Our next example is one of the simplest, but most important. Indeed, as we will see later in the section on convergence, this type of martingale is almost universal in the sense that every uniformly integrable martingale is of this type. The process is constructed by conditioning a fixed random variable on the σ -algebras in a given filtration, and thus *accumulating information* about the random variable.

Suppose that $\mathfrak{F} = \{\mathcal{F}_t : t \in T\}$ is a filtration on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and that X is a real-valued random variable with $\mathbb{E}(|X|) < \infty$. Define $X_t = \mathbb{E}(X \mid \mathcal{F}_t)$ for $t \in T$. Then $\mathbf{X} = \{X_t : t \in T\}$ is a martingale with respect to \mathfrak{F} .

Proof

For $t \in T$, recall that $|X_t| = |\mathbb{E}(X \mid \mathcal{F}_t)| \leq \mathbb{E}(|X| \mid \mathcal{F}_t)$. Taking expected values gives $\mathbb{E}(|X_t|) \leq \mathbb{E}(|X|) < \infty$. Suppose that $s, t \in T$ with $s < t$. Using the tower property of conditional expected value,

$$\mathbb{E}(X_t \mid \mathcal{F}_s) = \mathbb{E}[\mathbb{E}(X \mid \mathcal{F}_t) \mid \mathcal{F}_s] = \mathbb{E}(X \mid \mathcal{F}_s) = X_s \quad (17.1.49)$$

The martingale in the last theorem is known as *Doob's martingale* and is named for Joseph Doob who did much of the pioneering work on martingales. It's also known as the *Lévy martingale*, named for Paul Lévy.

Doob's martingale arises naturally in the statistical context of Bayesian estimation. Suppose that $\mathbf{X} = (X_1, X_2, \dots)$ is a sequence of independent random variables whose common distribution depends on an unknown real-valued parameter θ , with values in a parameter space $A \subseteq \mathbb{R}$. For each $n \in \mathbb{N}_+$, let $\mathcal{F}_n = \sigma\{X_1, X_2, \dots, X_n\}$ so that $\mathfrak{F} = \{\mathcal{F}_n : n \in \mathbb{N}_+\}$ is the natural filtration associated with \mathbf{X} . In Bayesian estimation, we model the unknown parameter θ with a random variable Θ taking values in A and having a specified *prior distribution*. The *Bayesian estimator* of θ based on the sample $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$ is

$$U_n = \mathbb{E}(\Theta \mid \mathcal{F}_n), \quad n \in \mathbb{N}_+ \quad (17.1.50)$$

So it follows that the sequence of Bayesian estimators $\mathbf{U} = (U_n : n \in \mathbb{N}_+)$ is a Doob martingale. The estimation referred to in the discussion of the [beta-Bernoulli](#) process above is a special case.

Density Functions

For this example, you may need to review general measures and density functions in the chapter on Distributions. We start with our probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and filtration $\mathfrak{F} = \{\mathcal{F}_n : n \in \mathbb{N}\}$ in discrete time. Suppose now that μ is a finite measure on the sample space (Ω, \mathcal{F}) . For each $n \in \mathbb{N}$, the restriction of μ to \mathcal{F}_n is a measure on the measurable space (Ω, \mathcal{F}_n) , and similarly the restriction of \mathbb{P} to \mathcal{F}_n is a probability measure on (Ω, \mathcal{F}_n) . To save notation and terminology, we will refer to these as μ and \mathbb{P} on \mathcal{F}_n , respectively. Suppose now that μ is absolutely continuous with respect to \mathbb{P} on \mathcal{F}_n for each $n \in \mathbb{N}$. Recall that this means that if $A \in \mathcal{F}_n$ and $\mathbb{P}(A) = 0$ then $\mu(A) = 0$ for every $B \in \mathcal{F}_n$ with $B \subseteq A$. By the Radon-Nikodym theorem, μ has a density function $X_n : \Omega \rightarrow \mathbb{R}$ with respect to \mathbb{P} on \mathcal{F}_n for each $n \in \mathbb{N}_+$. The density function of a measure with respect to a positive measure is known as a *Radon-Nikodym derivative*. The theorem and the derivative are named for Johann Radon and Otto Nikodym. Here is our main result.

$\mathbf{X} = \{X_n : n \in \mathbb{N}\}$ is a martingale with respect to \mathfrak{F} .

Proof

Let $n \in \mathbb{N}$. By definition, X_n is measurable with respect to \mathcal{F}_n . Also, $\mathbb{E}(|X_n|) = \|\mu\|$ (the total variation of μ) for each $n \in \mathbb{N}$. Since μ is a finite measure, $\|\mu\| < \infty$. By definition,

$$\mu(A) = \int_A X_n d\mathbb{P} = \mathbb{E}(X_n; A), \quad A \in \mathcal{F}_n \quad (17.1.51)$$

On the other hand, if $A \in \mathcal{F}_n$ then $A \in \mathcal{F}_{n+1}$ and so $\mu(A) = \mathbb{E}(X_{n+1}; A)$. So to summarize, X_n is \mathcal{F}_n -measurable and $\mathbb{E}(X_{n+1}; A) = \mathbb{E}(X_n; A)$ for all $A \in \mathcal{F}_n$. By definition, this means that $\mathbb{E}(X_{n+1} | \mathcal{F}_n) = X_n$, and so \mathbf{X} is a martingale with respect to \mathfrak{F} .

Note that μ may not be absolutely continuous with respect to \mathbb{P} on \mathcal{F} or even on $\mathcal{F}_\infty = \sigma(\bigcup_{n=0}^\infty \mathcal{F}_n)$. On the other hand, if μ is absolutely continuous with respect to \mathbb{P} on \mathcal{F}_∞ then μ has a density function X with respect to \mathbb{P} on \mathcal{F}_∞ . So a natural question in this case is the relationship between the martingale \mathbf{X} and the random variable X . You may have already guessed the answer, but at any rate it will be given in the section on convergence.

This page titled [17.1: Introduction to Martingales](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist](#) ([Random Services](#)) via [source content](#) that was edited to the style and standards of the LibreTexts platform.