17.1: Introduction to Martingales

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

Another case that arises frequently is when we have a second stochastic process \( \bs{Y} = \{Y_t: t \in T\} \) on \( (\Omega, \mathscr{F}, \P) \) with values in a general measure space \( (\Omega, \mathscr{F}, \P) \), and \( \mathfrak{F} \) is the...
natural filtration associated with \( \sigma(Y) \). So in this case, our main assumption is that \( \sigma(X_t) \) is measurable with respect to \( \sigma(Y_s: s \in T, s \le t) \) for \( t \in T \).

The theory of martingales is beautiful, elegant, and mostly accessible in discrete time, when \( T = \N \). 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 \( X_t \) 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 \( \sigma \)-algebra.

The process \( X_t \) is a martingale with respect to \( \mathfrak{F} \) if \( \E(X_t \mid \mathcal{F}_s) = X_s \) for all \( s, t \in T \) with \( s \le t \).

In the special case that \( \mathfrak{F} \) is the natural filtration associated with \( X_t \), we simply say that \( X_t \) is a martingale, without reference to the filtration. In the special case that we have a second stochastic process \( Y_t \) and \( \mathfrak{F} \) is the natural filtration associated with \( Y_t \), we say that \( X_t \) is a martingale with respect to \( \sigma(Y) \).

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 \( X_t \) 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 \( t \). If \( X_t \) is a martingale, then the expected value at a future time, given all of our information, is the present value.
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 \( \mathbb{E}\left(X_t \mid \mathscr{F}_s\right) \ge X_s \) for all \( s, t \in T \) with \( s \le t \).
2. \( \mathbf{X} \) is a super-martingale with respect to \( \mathfrak{F} \) if \( \mathbb{E}\left(X_t \mid \mathscr{F}_s\right) \le X_s \) for all \( s, t \in T \) with \( s \le t \).

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 \( \sigma \)-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, \( \mathbb{N} = \{0, 1, 2, \ldots\} \) for fixed \( n \in \mathbb{N} \). In the section on backwards martingales, \( \mathbb{N} = \{0, 1, 2, \ldots\} \) or \( \mathbb{N} = \{-\infty, 0, 1, 2, \ldots\} \). 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. \( (\bs{X}) \) is a martingale with respect to \( \mathfrak{F} \) if and only if \( \E(X_{n+1} \mid \mathscr{F}_n) = X_n \) for all \( n \in \N \).
2. \( (\bs{X}) \) is a sub-martingale with respect to \( \mathfrak{F} \) if and only if \( \E(X_{n+1} \mid \mathscr{F}_n) \ge X_n \) for all \( n \in \N \).
3. \( (\bs{X}) \) is a super-martingale with respect to \( \mathfrak{F} \) if and only if \( \E(X_{n+1} \mid \mathscr{F}_n) \le X_n \) for all \( n \in \N \).

**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 \( k, n \in \N \) with \( k < n \). Then \( k \le n - 1 \) so \( \mathscr{F}_k \subseteq \mathscr{F}_{n-1} \) and hence \( \E(X_n \mid \mathscr{F}_k) = \E\left[\E(X_n \mid \mathscr{F}_{n-1}) \mid \mathscr{F}_k\right] = \E(X_{n-1} \mid \mathscr{F}_k) \). Repeating the argument, we get to \( \E(X_n \mid \mathscr{F}_k) = \E(X_{k+1} \mid \mathscr{F}_k) = X_k \). 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 \( \langle s, t \rangle \in T \) with \( s \le t \).

1. If \( (\bs{X}) \) is a martingale with respect to \( \mathfrak{F} \) then \( \E(X_s) = \E(X_t) \).
2. If \( (\bs{X}) \) is a sub-martingale with respect to \( \mathfrak{F} \) then \( \E(X_s) \le \E(X_t) \).
3. If \( (\bs{X}) \) is a super-martingale with respect to \( \mathfrak{F} \) then \( \E(X_s) \ge \E(X_t) \).

**Proof**

The results follow directly from the definitions, and the critical fact that \( \E(\E(X_t \mid \mathscr{F}_s)) = \E(X_t) \) for \( s, t \in T \).

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

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

Proof

Since $X$ is measurable with respect to $\mathscr{F}_0$, it is measurable with respect to $\mathscr{F}_t$ for all $t \in T$. Hence $\{X_t: t \in T\}$ is adapted to $\mathfrak{F}$. If $s, t \in T$ with $s \le t$, then $\E(X_t \mid \mathscr{F}_s) = \E(X \mid \mathscr{F}_s) = X = X_s$.

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 $\{V_n: n \in \N\}$ is a sequence of independent random variables with $\E(|V_k|) < \infty$ for $k \in \N$. Let $X_n = \sum_{k=0}^n V_k$, $n \in \N$ so that $\{X_n: n \in \N\}$ is simply the partial sum process associated with $\{V_n: n \in \N\}$.

For the partial sum process $\{X_n: n \in \N\}$,

1. If $\E(V_n) \ge 0$ for $n \in \N_+$ then $\{X_n: n \in \N\}$ is a super-martingale.
2. If $\E(V_n) \le 0$ for $n \in \N_+$ then $\{X_n: n \in \N\}$ is a sub-martingale.
3. If $\E(V_n) = 0$ for $n \in \N_+$ then $\{X_n: n \in \N\}$ is a martingale.

Proof

Let $\mathscr{F}_n = \sigma\{X_0, X_1, \ldots, X_n\}$ for $n \in \N$. Note first that $\E(|X_n|) \le \sum_{k=0}^n \E(|V_k|) < \infty$. Next, $\E(\left(X_{n+1} \mid \mathscr{F}_n\right) = \E(\left(X_n + V_{n+1} \mid \mathscr{F}_n\right) = \E(X_n + \E(V_{n+1} \mid \mathscr{F}_n)) \in \mathfrak{F}_n$. The last equality holds since $X_n$ is measurable with respect to $\mathfrak{F}_n$ and $V_{n+1}$ is independent of $\mathfrak{F}_n$. 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 $n \in \N_+$. 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 $\E(X_n) = \sum_{k=0}^n \E(V_k)$. Hence condition (a) is equivalent to $n \mapsto \E(X_n)$ increasing, condition (b) is equivalent to $n \mapsto \E(X_n)$ decreasing, and condition (c) is equivalent to $n \mapsto \E(X_n)$ constant. Here is another martingale associated with the partial sum process, known as the second moment martingale.

Suppose that $\E(V_k) = 0$ for $k \in \N_+$ and $\var(V_k) < \infty$ for $k \in \N$. Let $Y_n = X_n^2 - \var(X_n)$, $n \in \N$. Then $\{Y_n: n \in \N\}$ is a martingale with respect to $\{X_n: n \in \N\}$.

Proof

Again, let $\mathfrak{F}_n = \sigma\{X_0, X_1, \ldots, X_n\}$ for $n \in \N$. Since the sequence $\{Y_n\}$ is
independent, note that \( \text{Var}(X_n) = \text{Var}\left(\sum_{k=0}^n V_k\right) = \sum_{k=0}^n \text{Var}(V_k) \). Also, \( \text{Var}(V_k) = \text{E}(V_k^2) \) since \( \text{E}(V_k) = 0 \) for \( k \in \mathbb{N}_+ \). In particular, \( \text{E}(Y_n) = \text{E}\left(\sum_{k=0}^n V_k\right) = \sum_{k=0}^n \text{E}(V_k) \) since \( \text{E}(V_k) = 0 \) for \( k \in \mathbb{N}_+ \). Next for \( n \in \mathbb{N}_+ \), let \( \text{Var}(X_n) = \text{Var}\left(\sum_{k=0}^n V_k\right) \). Also, \( \text{Var}(X_{n+1}) = \text{Var}\left(\sum_{k=0}^n V_k + V_{n+1}\right) = \sum_{k=0}^n \text{Var}(V_k) + \text{Var}(V_{n+1}) \). Hence we have \( \text{E}(Y_{n+1}) = \text{E}(X_n) + \text{Var}(X_{n+1}) \).

So under the assumptions in this theorem, both \( \text{E}(X_0) \) and \( \text{E}(Y_0) \) 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 \( \text{E}(X_n) = \text{E}(Y_n) = 0 \) for \( n \in \mathbb{N}_+ \). Then for \( n \in \mathbb{N}_+ \), let \( V_n = X_n - X_{n-1} \). The process \( \text{E}(V_n) = 0 \) for \( n \in \mathbb{N}_+ \). Hence we have \( \text{E}(Y_{n+1}) = \text{E}(X_n) + \text{Var}(X_{n+1}) \).

Let \( \text{E}(V_0) = 0 \) and \( \text{E}(V_n) = X_n - X_{n-1} \) for \( n \in \mathbb{N}_+ \). The process \( \text{E}(V_n) = 0 \) for \( n \in \mathbb{N}_+ \). Hence we have \( \text{E}(Y_{n+1}) = \text{E}(Y_n) + \text{Var}(X_{n+1}) \).

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

Suppose that \( \text{E}(V_n) = 0 \) for \( n \in \mathbb{N}_+ \). Then

1. \( \text{Var}(V_n) \) is adapted to \( \mathfrak{F}_n \).
2. \( \text{E}(V_n) = 0 \) for \( n \in \mathbb{N} \), and hence \( \text{E}(V_n) = 0 \) for \( n \in \mathbb{N}_+ \).
3. \( \text{E}(V_n) = 0 \) for \( n \in \mathbb{N}_+ \)

**Proof**

1. Of course \( \text{E}(V_0) = 0 \) is measurable with respect to \( \mathfrak{F}_0 \). For \( n \in \mathbb{N}_+ \), \( \text{E}(V_n) = 0 \) and \( \text{E}(X_{n-1}) \), and hence \( \text{E}(V_n) = 0 \) for \( n \in \mathbb{N} \).
2. \( \text{E}(V_{n+1}) = \text{E}(X_n) \) for \( n \in \mathbb{N}_+ \). By the martingale and adapted properties, \( \text{E}(V_{n+1}) = \text{E}(X_n) \) for \( n \in \mathbb{N}_+ \). Next by the tower property, \( \text{E}(V_{n+1}) = \text{E}(X_n) \) for \( n \in \mathbb{N}_+ \). Continuing (or using induction) gives the general result.
3. Since \( \text{E}(V_n) = 0 \) for \( n \in \mathbb{N}_+ \), it has constant mean, as noted above. Hence \( \text{E}(V_n) = 0 \) for \( n \in \mathbb{N}_+ \). We could also use part (a).

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

Suppose again that \( \text{Var}(V_n) = 0 \) for \( n \in \mathbb{N}_+ \). Then for \( n \in \mathbb{N}_+ \), we have \( \text{Var}(V_n) = 0 \) for \( n \in \mathbb{N}_+ \).

Assume that \( \var(X_n) \lt \infty \) for \( n \in \mathbb{N} \). Then \( \{ \bs V \} \) is an uncorrelated sequence. Moreover, \[ \var(X_n) = \sum_{k=0}^n \var(V_k) = \var(X_0) + \sum_{k=1}^n \E(V_k^2), \quad \text{quad } n \in \mathbb{N} \]

**Proof**

Let \( \{ k, \, n \in \mathbb{N} \} \) with \( \{ k \lt n \} \). To show that \( \{ V_k \} \) and \( \{ V_n \} \) are uncorrelated, we just need to show that \( \E(V_k V_n) = 0 \) (since \( E(V_n) = 0 \)). But by the previous result, \[ \E(V_k V_n) = \E[E(V_k V_n \mid \mathcal{F}_k)] = \E[V_k \E(V_n \mid \mathcal{F}_k)] = 0 \] Finally, the variance of a sum of uncorrelated variables is the sum of the variances. Since \( \{ V_k \} \) has mean 0, \( \var(V_k) = \E(V_k^2) \) for \( k \in \mathbb{N}_+ \). Hence the formula for \( \var(X_n) \) 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 \( \{ \bs V = \{V_n: n \in \mathbb{N}\} \} \) is a sequence of independent random variables with \( \{ \{V_n: n \in \mathbb{N}_+\} \} \) identically distributed. We assume that \( \{ \E(V_n) \} \lt \infty \) for \( n \in \mathbb{N} \) and we let \( \{ a \} \) denote the common mean of \( \{ V_n: n \in \mathbb{N}_+ \} \). Let \( \{ \bs X = \{X_n: n \in \mathbb{N}\} \} \) be the partial sum process associated with \( \{ \bs V \} \) so that \[ X_n = \sum_{i=0}^n V_i, \quad n \in \mathbb{N} \] This setting is a special case of the more general partial sum process considered above. The process \( \{ \bs X \} \) 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 \( \{ \bs X \} \),

1. \( \{ \bs X \} \) is a martingale if \( \{ a = 0 \} \).
2. \( \{ \bs X \} \) is a sub-martingale if \( \{ a \ge 0 \} \).
3. \( \{ \bs X \} \) is a super-martingale if \( \{ a \le 0 \} \)

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

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

**Proof**

This follows from the corresponding result for a general partial sum process, above, since \[ \var(X_n) = \sum_{k=0}^n \var(V_k) = \var(V_0) + \sum_{k=1}^n b^2, \quad \text{quad } n \in \mathbb{N} \]

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


Updated: Thu, 18 Feb 2021 00:32:34 GMT

Powered by 7
Our next discussion is similar to the one on partial sum processes above, but with products instead of sums. So suppose that \( \{ \text{bs} V = \{V_n: n \in \mathbb{N}\} \} \) is an independent sequence of nonnegative random variables with \( \{ \text{E}(V_n) \mid \text{lt infinity} \} \) for \( n \in \mathbb{N} \). Let \( X_n = \prod_{i=0}^{n} V_i \), quad \( n \in \mathbb{N} \) so that \( \{ \text{bs} X = \{X_n: n \in \mathbb{N}\} \} \) is the partial product process associated with \( \{ \text{bs} V \} \).

For the partial product process \( \{ \text{bs} X \} \),

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

Proof

Let \( \{ \text{mathscr{F}}_n = \{ \text{sigma}\{V_0, V_1, \ldots, V_n\}\} \} \) for \( n \in \mathbb{N} \). Since the variables are independent, \( \{ \text{E}(X_n) = \prod_{i=0}^{n} \text{E}(V_i) \mid \text{lt infinity} \} \). Next, \( \{ \text{E}(X_{n+1}) \mid \text{mathscr{F}}_n \} = \text{E}(X_n \cdot V_{n+1}) \mid \text{mathscr{F}}_n \) is measurable with respect to \( \{ \text{mathscr{F}}_n \} \) and \( \{ V_{n+1} \} \) is independent of \( \{ \text{mathscr{F}}_n \} \). 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 \( \{ \text{bs}\{V\} = \{V_n: n \in \mathbb{N}\} \} \) is a sequence of independent random variables with \( \{ \text{P}(V_i = 1) = p \} \) and \( \{ \text{P}(V_i = -1) = 1 - p \} \) for \( i \in \mathbb{N}_+ \), where \( \{ p \in (0, 1) \} \). Let \( \{ \text{bs}\{X\} = \{X_n: n \in \mathbb{N}\} \} \) be the partial sum process associated with \( \{ \text{bs}\{V\} \} \) so that \( \{ X_n = \sum_{i=0}^{n} V_i \} \) quad \( n \in \mathbb{N} \). Then \( \{ \text{bs}\{X\} \} \) 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 \( \{ p \leq \frac{1}{2} \} \) then \( \{ \text{bs}\{X\} \} \) is a sub-martingale.
2. If \( \{ p \leq \frac{1}{2} \} \) then \( \{ \text{bs}\{X\} \} \) is a sub-martingale.
3. If \( \{ p = \frac{1}{2} \} \) then \( \{ \text{bs}\{X\} \} \) is a martingale.

Proof

Note that \( \{ \text{E}(V_n) = p \cdot (1 - p) = 2p - 1 \} \) 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 \( Y_n = X_n^2 - \var(V_0) - n \) for \( n \in \N \). Then \( \{Y_n: n \in \N\} \) is a martingale with respect to \( \{\bs(X)\} \)

Proof

Note that \( \| \E(V_i) = 0 \) and \( \| \var(V_i) = 1 \) for each \( i \in \N_+ \), 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 \( n \in \N \) define \[ Z_n = \left(\frac{1 - p}{p}\right)^{X_n}, \quad n \in \N \] Then \( \{Z_n: n \in \N\} \) is a martingale with respect to \( \{\bs(X)\} \).

Proof

Note that \[ \| Z_n = \prod_{k=0}^n \left(\frac{1 - p}{p}\right)^{V_k}, \quad \text{quad n \in \N} \] and \( \| \E[\left(\frac{1 - p}{p}\right)^{V_k}] = 1 \) for \( k \in \N_+ \). 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 \( \bs{Y} = (Y_1, Y_2, \ldots) \) such that given \( P = p \in (0, 1) \), \( \bs{Y} \) is a sequence of independent variables with \( P(Y_i = 1) = p \) for \( i \in \N_+ \). 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 \[ P(Y_n = k) = \binom{n}{k} \frac{a^k b^{n-k}}{(a + b)^n}, \quad k \in \{0, 1, \ldots, n\} \] Now let \[ Z_n = \frac{a + Y_n}{a + b + n}, \quad n \in \N \] This variable also arises naturally. Let \( \mathscr{F}_n = \sigma\{X_1, X_2, \ldots, X_n\} \) for \( n \in \N \). Then as shown in the section on the beta-Bernoulli process, \( \| Z_n = \E(X_{n+1} \mid \mathscr{F}_n) \) and \( \{Z_n \} \) is a martingale with respect to \( \{\bs(X)\} \).

Proof

Note that \( \| \E(Z_n) = 0 \) and \( \| \var(Z_n) = 1 \) for \( n \in \N \). Next, \( \| \E[\left(a + Y_{n+1}\right) \mid \mathscr{F}_n] = \E\left(\frac{a + Y_{n+1}}{a + b + n}\right) = \frac{a + \E\left(Y_{n+1}\right)}{a + b + n} \). 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 \( P \).

\[ \| \bs(Z) = \{Z_n: n \in \N\} \] is a martingale with respect to \( \{\bs(X)\} \).

Proof

Note that \( \| \E(Z_n) = 0 \) and \( \| \var(Z_n) = 1 \) for \( n \in \N \). Next, \( \| \E[\left(a + Y_{n+1}\right) \mid \mathscr{F}_n] = \E\left(\frac{a + Y_{n+1}}{a + b + n}\right) = \frac{a + \E\left(Y_{n+1}\right)}{a + b + n} \). 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 \( P \).

https://stats.libretexts.org/Bookshelves/Probability_Theory/Book%3A_Probability_Mathematical_Statistics_and_Stochastic_Pr…
Updated: Thu, 18 Feb 2021 00:32:34 GMT
Powered by
\[ \frac{a + Y_n + \mathbb{E}(X_{n+1} \mid \mathcal{F}_n)}{a + b + n + 1} = \frac{a + Y_n + (a + Y_n) / (a + b + n)}{a + b + n + 1} = \frac{a + Y_n}{a + b + n} = Z_n \]

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

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 \( \{X_n: n \in \mathbb{N}_+\} \) is a classical example of a sequence of exchangeable yet dependent variables. Let \( \{Y_n: n \in \mathbb{N}\} \) denote the partial sum process associated with \( \{X_n\} \), so that once again, \( Y_n = \sum_{i=1}^n X_i \). 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 \( a + c Y_n \), while the total number of balls in the urn is \( a + b + c n \) and so the *proportion* of red balls in the urn is \( Z_n = \frac{a + c Y_n}{a + b + c n} \). \( \{Z_n: n \in \mathbb{N}\} \) is a martingale with respect to \( \{X_n\} \).

*Indirect proof*

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

*Direct Proof*

Trivially, \( 0 \le Z_n \le 1 \) so \( \mathbb{E}(Z_n) \le \mathbb{E}(1) = 1 \) for \( n \in \mathbb{N}_+ \). \( \mathbb{E}(Z_{n+1} \mid \mathcal{F}_n) = \mathbb{E}(\mathbb{E}(a + c Y_{n+1} \mid \mathcal{F}_n) \mid \mathcal{F}_n) = \mathbb{E}(a + c Y_{n+1}) = \frac{a + Y_n + (a + Y_n) / (a + b + n)}{a + b + n + 1} \) 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, \( \mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = Z_n = \frac{a + c Y_n}{a + b + c n} \). Substituting and simplifying gives \( \mathbb{E}(Z_{n+1} \mid \mathcal{F}_n) = Z_n \).

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* \( \{Z_n\} \) associated with a sequence of independent variables.
random variables \( \mathcal{V} \). 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 \((X_0, X_1, \ldots, X_m)\). 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 \( \mathcal{V}(X) = \{X_t : t \in \mathbb{N}\} \) satisfying the basic assumptions above relative to the filtration \( \mathcal{F} = \{\mathcal{F}_s : s \in \mathbb{T}\} \). Here are the two definitions.

The process \( \mathcal{V}(X) \) has

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

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 \( t \in \mathbb{T} \).

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

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

**Proof**

The proof is just like the one above for partial sum processes. Suppose that \((X, \mathcal{F})\) 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 \( \mathcal{V}(Y) = \{Y_t : t \in [0, \infty)\} \) is sometimes called the compensated process associated with \( \mathcal{V}(X) \) and has mean function 0. If \( \mathcal{V}(X) \) has independent increments, then clearly so does \( \mathcal{V}(Y) \).

Hence the following result is a trivial corollary to our previous theorem.

Suppose that \( \mathcal{V}(X) \) has independent increments. The compensated process \( \mathcal{V}(Y) \) 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 \( \mathcal{V}(X) = \{X_t : t \in [0, \infty)\} \) has independent increments with constant mean function and with \( \mathbb{V}(X_t) \) for \( t \in T \). Then \( \mathcal{V}(Y) = \{Y_t : t \in [0, \infty)\} \) is a martingale where \( Y_t = X_t^2 - \mathbb{V}(X_t) \) for \( t \in T \).
Proof

The proof is essentially the same as for the partial sum process in discrete time. Suppose that \( \{ s, t \in T \} \) with \( s \lt t \)). Note that \( \{ E(Y_t \mid \mathscr{F}_s) = E(X_t^2 \mid \mathscr{F}_s) - \var(X_t) \). Next, \( X_t^2 = (X_t - X_s + X_s)^2 = (X_t - X_s)^2 + 2(X_t - X_s)X_s + X_s^2 \) but \( (X_t - X_s) \) is independent of \( \mathscr{F}_s \), \( X_s \) is measurable with respect to \( \mathscr{F}_s \), and \( E(X_t - X_s) = 0 \) so \( E(X_t^2 \mid \mathscr{F}_s) = E[(X_t - X_s)^2 + 2X_sE(X_t - X_s) + X_s^2] = E[(X_t - X_s)^2 + X_s^2] \) but also by independence and since \( X_t - X_s \) has mean 0, \( \var(X_t) = \var[(X_t - X_s) + X_s] = \var(X_s) + \var[(X_t - X_s)^2] \). Putting the pieces together gives \( E(Y_t \mid \mathscr{F}_s) = X_s^2 - \var(X_s) = Y_s \).

Of course, since the mean function is constant, \( \{ \text{bs X} \} \) 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 \( \{ \text{bs X} = \{X_t: t \in \text{T}\} \} \) has stationary, independent increments, and let \( \{ a = E(X_1 - X_0) \} \). Then

1. \( \{ \text{bs X} \} \) is a martingale if \( a = 0 \)
2. \( \{ \text{bs X} \} \) is a sub-martingale if \( a \ge 0 \)
3. \( \{ \text{bs X} \} \) is a super-martingale if \( a \le 0 \)

Proof

Recall that the mean function \( m(t) \) is given by \( m(t) = E(X_0) + a t \) 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 \( \{ \text{bs X} = \{X_t: t \in \text{T}\} \} \) has stationary, independent increments with \( E(X_0) = E(X_1) \) and \( b^2 = E(X_1^2) \lt \infty \). Then \( \{ \text{bs Y} = \{Y_t: t \in \text{T}\} \} \) is a martingale where \( Y_t = X_t^2 - \var(X_0) - b^2 t, \quad t \in T \)

Proof

Recall that if \( E(X_0) = E(X_1) \) then \( \{ \text{bs X} \} \) has constant mean function. Also, \( \var(X_t) = \var(X_0) + b^2 t \), 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 \( \{ \text{bs N} = \{N_t: t \in [0, \infty)\} \} \) with constant rate \( r \in (0, \infty) \) has stationary, independent increments and mean function given by \( m(t) = r t \) for \( t \in [0, \infty) \). More generally, suppose that \( r: [0, \infty) \to (0, \infty) \) is piecewise continuous (and non-constant). The non-homogeneous Poisson counting process \( \{ \text{bs N} = \{N_t: t \in [0, \infty)\} \} \) with rate function \( r \) has independent increments and mean function given by \( m(t) = \int_0^t r(s) \, ds \), \( \var(N_t) = \int_0^t r(s) \, ds \). The increment \( N_t - N_s \) has the Poisson distribution with parameter \( m(t) - m(s) \) for \( s, t \in [0, \infty) \) with \( s \lt t \), so the process does not have stationary increments. In all cases, \( m(t) \) is increasing, so the following results are corollaries of our general results:
Let \( \{N_t: t \in [0, \infty)\} \) be the Poisson counting process with rate function \( r: [0, \infty) \to (0, \infty) \). Then

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

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 \( (S, \mathcal{S}, \mu) \) is a general measure space, and that \( \{X_n: n \in \mathbb{N}\} \) is a sequence of independent, identically distributed random variables, taking values in \( (S, \mathcal{S}) \). In statistical terms, \( \{X_n\} \) 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 \( \{X_n\} \). Suppose now that the underlying distribution either has probability density function \( g_0 \) or probability density function \( g_1 \), with respect to \( \mu \). 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 \( L_n = \prod_{i=1}^n \frac{g_0(X_i)}{g_1(X_i)}, n \in \mathbb{N} \) 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 \( \{L_n: n \in \mathbb{N}\} \) is a martingale with respect to \( \{X_n\} \), known as the likelihood ratio martingale.

**Proof**

Let \( \mathfrak{m}(F_n) = \sigma(X_1, X_2, \ldots, X_n) \). For \( n \in \mathbb{N} \), \( \mathfrak{m}(L_{\{X_n\}}) \mid \mathfrak{m}(F_n) \) is independent of \( X_n \). By the change of variables formula for expected value, we have \( \mathbb{E}[\prod_{i=1}^n \frac{g_0(X_i)}{g_1(X_i)}] = \int_S \frac{g_0(x)}{g_1(x)} g_1(x) \, d\mu(x) = 1 \). This result also follows essentially from the
theorem above on partial products. The sequence \( \{Z = (Z_1, Z_2, \ldots) \} \) 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 \( \phi \) 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 \( \{X_n : n \in \mathbb{N}\} \) where \( X_n \) is the number of particles in the \( n \)th generation for \( n \in \mathbb{N} \). The process \( \{X_n\} \) 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: \( q = \mathbb{P}(X_n = 0 \mid X_0 = 1) \). 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 \( f(0) < 1 \) 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 \( \phi \) (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 \( \{X_n\} \).

1. \( \{Y_n = X_n / m^n : n \in \mathbb{N}\} \) where \( Y_n = X_n / m^n \) for \( n \in \mathbb{N} \).
2. \( \{Z_n = q^{X_n} : n \in \mathbb{N}\} \) where \( Z_n = q^{X_n} \) for \( n \in \mathbb{N} \).

#### Proof

Let \( \mathbf{F} = \sigma(X_0, X_1, \ldots) \). For \( n \in \mathbb{N} \), note that \( \{X_{n+1}\} \) can be written in the form \( \{X_{n+1} = U_{n+1} X_n \} \) where \( \{U_{n+1} \mid \mathbf{F}_n\} \) is a sequence of independent variables, each with PDF \( f \) (and hence mean \( m \) and PGF \( \phi \)), and with \( \{U_{n+1} \mid \mathbf{F}_n\} \) independent of \( \text{mathscr{F}}_n \).

Think of \( U_{n+1} \) as the number of children of the \( n \)th particle in generation \( n \).

1. For \( n \in \mathbb{N} \), \( \{\mathbf{E}(X_{n+1} \mid \mathbf{F}_n) = \mathbf{E}(U_{n+1} X_n \mid \mathbf{F}_n) = \mathbf{E}(U_{n+1} \mid \mathbf{F}_n) X_n = m X_n \} \) since \( X_n \) is a positive random variable.
2. For \( n \in \mathbb{N} \), \( \{\mathbf{E}(X_{n+1} \mid \mathbf{F}_n) = \mathbf{E}(U_{n+1} \mid \mathbf{F}_n) X_n = m X_n \} \) since \( X_n \) is a positive random variable.

### 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 \( \sigma \)-algebras in a given filtration, and thus accumulating information about the random variable.

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

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 \( \mathbb{E}(|X_t|) \leq \mathbb{E}(|X|) \lt \infty \). Suppose that \( s, t \in T \) with \( s \lt t \). Using the tower property of conditional expected value, \( \mathbb{E}(X_t \mid \mathcal{F}_s) = \mathbb{E}(\mathbb{E}(X \mid \mathcal{F}_t) \mid \mathcal{F}_s) = \mathbb{E}(X \mid \mathcal{F}_s) = X_s \)

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 \( \{X_1, X_2, \ldots\} \) is a sequence of independent random variables whose common distribution depends on an unknown real-valued parameter \( \theta \), with values in a parameter space \( A \subseteq \mathbb{R} \). For each \( n \in \mathbb{N}_+ \), let \( \mathcal{F}_n = \sigma\{X_1, X_2, \ldots, X_n\} \) so that \( \mathcal{F} = \{\mathcal{F}_n : n \in \mathbb{N}_+\} \) is the natural filtration associated with \( \{X_i\} \). In Bayesian estimation, we model the unknown parameter \( \theta \) with a random variable \( \Theta \) taking values in \( A \) and having a specified prior distribution. The Bayesian estimator of \( \theta \) based on the sample \( \{X_1, X_2, \ldots, X_n\} \) is \( U_n = \mathbb{E}(\Theta \mid \mathcal{F}_n) \).

So it follows that the sequence of Bayesian estimators \( \{U_n : n \in \mathbb{N}_+\} \) 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 \( \mu \) is a finite measure on the sample space \( (\Omega, \mathcal{F}) \). For each \( n \in \mathbb{N} \), the restriction of \( \mu \) to \( \mathcal{F}_n \) is a measure on the measurable space \( (\Omega, \mathcal{F}_n) \), and similarly the restriction of \( \mathfrak{F} \) to \( \mathcal{F}_n \) is a probability measure on \( (\Omega, \mathcal{F}_n) \). To save notation and terminology, we will refer to these as \( (\mu, \mathcal{F}) \) and \( (\mathfrak{F}, \mathcal{F}_n) \) respectively. Suppose now that \( \mu \) is absolutely continuous with respect to \( \mathfrak{F} \) for each \( n \in \mathbb{N} \). Recall that this means that if \( \mu(A) = 0 \) then \( \mu(B) = 0 \) for every \( B \in \mathcal{F}_n \). By the Radon-Nikodym theorem, \( \mu \) has a density function \( \frac{d\mu}{d\mathfrak{F}} \) with respect to \( \mathfrak{F} \) 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.

\( \{X_n : n \in \mathbb{N}\} \) is a martingale with respect to \( \mathfrak{F} \).
Proof

Let \( n \in \mathbb{N} \). By definition, \( X_n \) is measurable with respect to \( \mathscr{F}_n \). Also, \( \|E(|X_n|) = \|\mu\| \) (the total variation of \( \mu \)) for each \( n \in \mathbb{N} \). Since \( \|\mu\| \) is a finite measure, \( \|\mu\| < \infty \). By definition, \( \mu(A) = \int_A X_n \, d\mathbb{P} = \E(X_n; A), \quad A \in \mathscr{F}_n \) On the other hand, if \( A \in \mathscr{F}_{n+1} \) then \( \mu(A) = \E(X_{n+1}; A) \). So to summarize, \( X_n \) is \( \mathscr{F}_n \)-measurable and \( \E(X_{n+1}; A) = \E(X_n; A) \) for all \( A \in \mathscr{F}_n \). By definition, this means that \( \E(X_{n+1} \mid \mathscr{F}_n) = X_n \), and so \( \bs X \) is a martingale with respect to \( \mathfrak{F} \).

Note that \( \mu \) may not be absolutely continuous with respect to \( \mathbb{P} \) on \( \mathscr{F}_\infty \) or even on \( \mathscr{F}_\infty \). On the other hand, if \( \mu \) is absolutely continuous with respect to \( \mathbb{P} \) on \( \mathscr{F}_\infty \) then \( \mu \) has a density function \( X \) with respect to \( \mathbb{P} \) on \( \mathscr{F}_\infty \). So a natural question in this case is the relationship between the martingale \( \bs X \) 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.