# MARTINGALES

ANDREW CHANG

Abstract. The concept of a martingale is simple and can be used for various applications in probability theory. We introduce the foundational background and define the concept, develop it with some key theorems, and give some applications in proving important results.

## 1. Introduction

Martingales are sequences of random variables where the expected value of each term, with knowledge of all the previous terms, is equal to the value of the previous term. In other words, at any given time, the conditional expectation of the next value, given all of the past values, is simply the current value. Some examples are the random walk and a gambler making a series of fair bets. (In particular, terminology for martingales has long been oriented around the latter.)

Martingales are related to Markov chains, but not all martingales are Markov chains and not all Markov chains are martingales, as the martingale definition deals with only the expected value being equal, and Markov chains are based only on knowing the next probabilities based on the present.

## 2. Filtrations and Foundation

In the standard definition with full generality, we can define martingales in terms of *$\sigma$-algebras*, and more specifically, a *filtration* of them:

**Definition 2.1.** A *$\sigma$-algebra* (also known as a *$\sigma$-field*) on a set $S$ is a collection of subsets of $S$ that include $S$ and are closed over complement and countable unions.

We use this collection to represent the knowledge at various times in the stochastic process.

**Definition 2.2.** A filtration on the probability space $(\Omega, \mathcal{F}, P)$ is a sequence $\{\mathcal{F}_n : n = 1, 2, \dots\}$ of sub-$\sigma$ fields of $\mathcal{F}$ such that for all n, $\mathcal{F}_n \subset \mathcal{F}_{n+1}$.

The filtration represents the knowledge at successive betting times. Note that it only increases, with each sub-$\sigma$ field being a superset of the last.

Finally, we define what it means for a stochastic process, essentially a collection of random variables, to be *adapted* to a filtration.

*Date*: November 11th, 2020.

**Definition 2.3.** A stochastic process $X = \{X_n,\ n = 1, 2, \dots\}$, is *adapted* to the filtration $(\mathcal{F}_n)$ if for all $n$, $X_n$ is $\mathcal{F}_n$-measurable (i.e. a random variable).

In other words, an adapted process can't "see into the future": $X_n$ isn't known until time $n$.

Although these definitions and concepts may seem contrived, they are broadly used in probability theory and are useful for formalizing some general ideas, so we will use them consistently in this text.

Another convention is that we will consistently use *a.s.* and *a.e.* to denote *almost surely* and *almost everywhere* respectively, due to the cumbersomeness of writing them out each time.

## 3. Martingales

The intuitive meaning of a martingale is essentially that $\mathbb{E}[X_{n+1} \mid X_1, \dots, X_n] = X_n$. However, this is not the most general definition. Using the terminology of filtrations, we can write a martingale as:

**Definition 3.1.** A process $X = \{X_n, \mathcal{F}_n,\ n = 1, 2, \dots\}$, is a *martingale* if for each $n = 1, 2, \dots,$

(1) $\{\mathcal{F}_n,\ n = 1, 2 \dots\}$ is a filtration and X is adapted to $(\mathcal{F}_n)$;
(2) for each n, $X_n$ is integrable;
(3) for each n, $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n$.

We can similarly define the more general *submartingales*:

**Definition 3.2.** A process $X = \{X_n, \mathcal{F}_n,\ n = 1, 2, \dots\}$, is a *submartingale* if for each $n = 1, 2, \dots,$

(1) $\{\mathcal{F}_n,\ n = 1, 2 \dots\}$ is a filtration and X is adapted to $(\mathcal{F}_n)$;
(2) for each n, $X_n$ is integrable;
(3) for each n, $X_n \leq \mathbb{E}[X_{n+1} \mid \mathcal{F}_n]$.

A *supermartingale* is analogous, reversing the inequality.

Since we will only concern ourselves with processes where $n = 1, 2, \dots$ for now, we will usually omit that condition for brevity.

## 4. Stopping Times

One of the first major results about martingales we will examine is *stopping times* (defined much the same as when examining Markov chains), as shown by the notable Optional Stopping Theorem. One application of this result is formally showing that any betting strategy that takes finite time will always lose money, as long as the odds are in favor of the house.

**Theorem 4.1** (Doob's Optional Stopping Theorem)**.** *If $X$ is a supermartingale and $T$ a stopping time, $\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$ if*

(1) *$X$ is bounded,*
(2) *$T$ is bounded,*
(3) *or $\mathbb{E}[T] < \infty$ and $|X_n(\omega) - X_{n-1}(\omega)|$ is bounded by $K$.*

*Proof.* Consider $X_n^T$, defined as taking $X_n$ if $n \leq T$ and $X_T$ otherwise. By induction, $\mathbb{E}[X_n^T - X_0] \leq 0$. If $T$ is bounded by $N$ then $\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$, as the sequence $X_n^T$ converges to $\mathbb{E}[X_T]$ uniformly. If $X$ is bounded then the $X_n^T$ sequence is converging to $X_T$ and are bounded, so the dominated convergence theorem applies, giving $\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$. In the last case we have

$$|X_n^T - X_0| = |\sum_{k=1}^{\min\{T,n\}} (X_k - X_{k-1})| \leq KT.$$

$\mathbb{E}[KT] < \infty$ so using the dominated convergence theorem on $X_n^T - X_0$ gives the desired result. ∎

Note that considering $-X$ proves the theorem in the case of a submartingale. The inequality can be shown to have equality for the martingale case by the fact that it must satisfy both the submartingale and the opposite supermartingale inequality.

*Example.* A simple example is of a gambler who bets \$1 on fair coin flips, with a $\frac{1}{2}$ probability of winning and a $\frac{1}{2}$ probability of losing. Starting with \$30, ey continues betting until either running out of money or reaching \$100, at which point ey goes home. Knowing that this is a martingale, we can apply the Optional Stopping Theorem to find that the expected value at a stopping time, say when ey leaves, is equal to the initial expected value, or simply \$30. Thus, there is a 0.3 chance that the gambler goes home with \$100, and a 0.7 chance ey goes bust.

## 5. Key Lemmas

Several lemmas, generally useful in probability theory, are needed. Proving them is non-trivial and outside the main focus of this text, but they will be used as stated below:

**Proposition 5.1** (Fatou's Lemma)**.** *If $X_n \geq 0$, then $\mathbb{E}[\liminf X_n] \leq \liminf \mathbb{E}[X_n]$.*

**Proposition 5.2** (Monotone Convergence Theorem)**.** *If $X_n \uparrow X$ a.s. and $X_n \geq 0$ for all $n$, then $\mathbb{E}[X_n] \uparrow \mathbb{E}[X]$.*

**Notation.** $X_n \uparrow X$ means $X_n \subseteq X_{n+1}$, for all positive integers $n$, and $\cup X_n = X$.

**Proposition 5.3** (Dominated Convergence Theorem)**.** *If $X_n \to X$ a.s., $|X_n| < Y$ for all $n$, and $\mathbb{E}[Y] < \infty$, then $\mathbb{E}[X_n] \to \mathbb{E}[X]$.*

One special case of the Dominated Convergence Theorem is where there is some $M \in R$ such that $|X_n| \leq M$ a.s. for all $n$. In this case, sometimes called the Bounded Convergence Theorem, we can just take $Y = M \in R$.

*Proof.* See [1] and [2]. ■

There are several more laws that concern conditional expectation:

**Theorem 5.4.** *Several laws:*

(1) *If* $Y$ *is a version of* $\mathbb{E}[X \mid \mathcal{G}]$ *then* $\mathbb{E}[Y] = \mathbb{E}[X]$.
(2) *If* $X$ *is* $\mathcal{G}$ *measurable then* $\mathbb{E}[X \mid \mathcal{G}] = X$ *a.s.*
(3) $\mathbb{E}[a_1X_1 + a_2X_2] = a_1\mathbb{E}[X_1 \mid \mathcal{G}] + a_2\mathbb{E}[X_2 \mid \mathcal{G}]$ *a.s.*
(4) *If* $X \geq 0$, *then* $\mathbb{E}[X \mid \mathcal{G}] \geq 0$.
(5) *If* $X_n$ *is an increasing sequence of random variables converging to* $X$ *pointwise, then* $\mathbb{E}[X_n \mid \mathcal{G}]$ *converges monotonically upwards to* $\mathbb{E}[X \mid \mathcal{G}]$ *a.s.*
(6) *If* $|X_n(\omega)| < V(\omega)$ *for all* $n$, $\mathbb{E}[V] < \infty$ *and* $X_n$ *converge to* $X$ *pointwise,* $\mathbb{E}[X_n \mid \mathcal{G}]$ *converges to* $\mathbb{E}[X \mid \mathcal{G}]$.
(7) *If* $c$ *is a convex function on* $\mathbb{R}$ *and* $\mathbb{E}(|c(X)|) < \infty$ *then* $\mathbb{E}[c(X) \mid \mathcal{G}] \geq c(\mathbb{E}[X \mid \mathcal{G}])$.
(8) *If* $\mathcal{H} \subset \mathcal{G}$, *then* $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}] = \mathbb{E}[X \mid \mathcal{H}]$.
(9) *If* $Z$ *is a* $\mathcal{G}$ *measurable and bounded random variable,* $\mathbb{E}[ZX \mid \mathcal{G}] = Z\mathbb{E}[X \mid \mathcal{G}]$ *a.s.*
(10) *If* $\mathcal{H}$ *is independent of* $\sigma(\sigma(X), \mathcal{G})$, *then* $\mathbb{E}[X \mid \sigma(\mathcal{G}, \mathcal{H})] = \mathbb{E}[X \mid \mathcal{G}]$ *a.s.*

*Proof.* See [3]. ■

We also define the concept of convergence in $\mathcal{L}^1$ as follows:

**Definition 5.5.** $X_n$ converges to a random variable $X$ (defined on $\Omega$) in $\mathcal{L}^p$ if

$$\lim_{n\to\infty}[|X_n - X \mid p] = 0.$$

We will mainly consider the case where $p = 1$.

## 6. Upcrossing and Integrability

Another important lemma is this, involving the interesting concept of an *upcrossing*:

**Lemma 6.1** (Doob Upcrossing Lemma). *For an interval* $[a, b]$, $\omega$ *fixed in a probability space, and* $X$ *a supermartingale, we define the number of upcrossings by time* $N$ *as the largest* $k$ *such that* $0 \leq s_1 < t_1 < s_2 < t_2 < ... < s_k < t_k \leq N$ *and* $X_{s_i}(\omega) < a$, $X_{t_i}(\omega) > b$. *Therefore, an upcrossing is a pair of times between which the value of the sequence* $X_n(\omega)$ *goes from below* $a$ *to above* $b$. *Let* $U_N[a, b]$ *be the number of upcrossings by time* $N$. *Then, we have*

$$(b - a)\mathbb{E}[U_N[a, b]] \leq \mathbb{E}[\max(a - X_N, 0)].$$

*Proof.* We use a martingale transform argument to get information about upcrossings. If our gambler bets 1 starting when $X_n$ is below $a$, and quitting when $X_n$ is above $b$ they will make $b - a$ for each upcrossing. If they continue this betting strategy until time $N$ they will only lose $a - X_N$ after making $b - a$ for each upcrossing. The total amount they have won until time $n$ is a supermartingale by the proposition on martingale transforms. Therefore the Optional Stopping Theorem applies, giving the result. ■

Lastly, we define the concept of *uniform integrability* of a martingale, which will be used in a few proofs. The motivation is essentially that when we take the limit of a sequence, we want the integrals, such as expectations, to converge properly, allowing for the limit and integral to be swapped. Conveniently, this condition is usually easy to verify for martingales.

**Definition 6.2.** A family of random variables $\mathcal{C}$ is *uniformly integrable* if for each $\epsilon > 0$ we can pick a real $K$ such that for all $X \in \mathcal{C}, \mathbb{E}[|X| \mid |X| > K] < \epsilon$.

Several lemmas can be shown concerning this condition.

**Theorem 6.3.** *Let $X \in \mathcal{L}^1$. Then the class $\{\mathbb{E}[X \mid G] \mid \mathcal{G}$ sub-$\sigma$-algebra of $F\}$ is uniformly integrable.*

*Proof.* Let $\epsilon > 0$. Then choose $\delta > 0$ such that for $F \in F$, $\mu(F) < \delta$ implies $\mathbb{E}[|X| \mid F] < \epsilon$. Then choose $K$ such that $\mathbb{E}[|X|] < \delta K$. Now let $\mathcal{G}$ be a sub-$\sigma$-algebra of $F$, and $Y = \mathbb{E}[X \mid \mathcal{G}]$. By Jensen's inequality $|Y| \leq \mathbb{E}[|X| \mid G]$, so $\mathbb{E}[|Y|] \leq \mathbb{E}[|X|]$, and $K\mu(|Y| > K) \leq \mathbb{E}[|Y|] \leq \mathbb{E}[|X|]$ so $\mu(|Y| > K) < \delta$. This implies

$$E(|Y| \mid |Y| \geq K) \leq E(|X| \mid |Y| \geq K) < \epsilon.$$

■

Uniform integrability makes a.s. convergence into $\mathcal{L}^1$ convergence.

## 7. Convergence

Another key aspect of martingales is concerning their convergence, where several interesting conclusions can be made.

**Theorem 7.1** (Martingale Convergence Theorem)**.** *Let $\{X_n, \mathcal{F}_n\}$ be a submartingale, and assume $\mathbb{E}[|X_n|]$ is bounded. Then with probability 1, there exists a finite integrable random variable $X_\infty$ such that*

$$\lim_{n\to\infty} X_n = X_\infty \text{ almost everywhere.}$$

*Proof.* First, we show that $\liminf X_n(\omega) = \limsup X_n(\omega)$, using the upcrossing inequality.

Suppose the sequence does not converge. Then there are rational numbers $a$ and $b$ such that $\liminf X_n(\omega) < a < b < \limsup X_n(\omega)$. It follows that there exists a subsequence $n_k$ such that $X_{n_k}(\omega) \to \liminf X_n(\omega)$ and another subsequence $n_j$ such that $X_{n_j}(\omega) \to \limsup X_n(\omega)$, and specifically, $X_{n_k}(\omega) < a$ for infinitely many $k$, and $X_{n_j}(\omega) > b$ for infinitely many $j$. This implies that there are infinitely many upcrossings of $[a, b]$. Thus, if $\limsup X_n(\omega) > \liminf X_n(\omega)$, there exist rational $a < b$ such that the number of upcrossings of $[a, b]$ is infinite. Now for every pair $r_1 < r_2$ of rationals,

$$\mathbb{E}[\nu_\infty(r_1, r_2)] \le \sup_n \mathbb{E}[(X_N - a)^+]/(b - a) \le \sup_N (\mathbb{E}[|X_N|] + a)/(b - a) < \infty.$$

Hence, $\nu_\infty(r_1, r_2) \ne \infty$ a.s., for each of the countable number of pairs $(r_1, r_2)$ of rationals. Thus for a.e. $\omega$, $\limsup X_n(\omega) = \liminf X_n(\omega)$. Thus the limit exists a.e. but so far we haven't shown it is finite. To see it is finite, note that $\lim |X_n|$ exists a.s. and by Fatou's Lemma, $\mathbb{E}[\lim |X_n|] \le \liminf \mathbb{E}[|X_n|] < \infty$. In particular, $\lim |X_n| < \infty$ a.e. ■

Thus, to show convergence of a martingale, submartingale, or supermartingale, it is enough to show that its absolute expectation is bounded.

## 8. Applications

Martingales can be used to prove a variety of key results in probability theory. We will look at one example, the proof of Kolmogorov's Zero-One Law. We begin with an important theorem necessary for the proof, the Lévy Upwards Theorem.

**Theorem 8.1** (Lévy Upwards Theorem)**.** *Let $X$ be a $\mathcal{L}^1$ random variable. Let $\mathcal{F}_n$ be a filtration. Define $M_n = \mathbb{E}[X \mid \mathcal{F}_n]$. Then $M_n$ is a martingale and $M_n$ converges to $\eta = \mathbb{E}[X \mid \mathcal{F}_\infty]$ a.s. and in $\mathcal{L}^1$.*

*Proof.* We start by looking at $Y = \mathbb{E}[M_{n+1} \mid \mathcal{F}] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{F}_{n+1}] \mid \mathcal{F}_n]$. This is a random variable that is $F_n$ measurable and $\mathbb{E}[Y \mid F] = \mathbb{E}[X \mid F]$ for $F \in \mathcal{F}_n$, so it is a.s. $M_n$. Each $M_n$ is bounded in $\mathcal{L}^1$ by 5.4, so $M_n$ is a martingale. It's also uniformly integrable by our earlier lemma, so $M_\infty$ exists and is in $\mathcal{L}^1$. Now we only need to show $M_\infty = \mathbb{E}[X \mid \mathcal{F}_\infty]$. Let $F \in \mathcal{F}_\infty$. Then $F \in \mathcal{F}_n$ for some $n$. Therefore,

$$\mathbb{E}[M_\infty \mid F] = \mathbb{E}[M_n \mid F] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{F}_n] \mid F] = \mathbb{E}[X \mid F].$$

Thus, $M_\infty = \mathbb{E}[X \mid \mathcal{F}_\infty]$ a.s. ■

**Definition 8.2.** Let $\mathcal{F}_1, \mathcal{F}_2, \ldots$ be a sequence of $\sigma$-algebras such that $\mathcal{F}_i \subset \mathcal{F}_j$ whenever $i < j$. Then the *tail algebra* of the $\mathcal{F}_i$ is $\cap_j \mathcal{F}_j$. The tail algebra of a sequence of random variables $X_1, X_2, \ldots$ is the tail algebra of $\mathcal{F}_i = \sigma(X_i, X_{i+1}, \ldots)$.

In other words, a *tail event* is an event defined in terms of an infinite sequence of random variables, where even without having an arbitrary large (but finite) chunk of the initial variables the occurrence can still be determined.

**Theorem 8.3** (Kolmogorov's Zero-One Law)**.** *Let $X_1, X_2, \ldots$ be a sequence of independent random variables and $\mathcal{T}$ be the corresponding tail algebra. If $F \in \mathcal{T}$, then $\mathbb{P}[F] = 0$ or $1$.*

*Proof.* Define $\mathcal{F}_n := \sigma(X_1, ..., X_n)$, then $\mathcal{F}_n$ is independent of $\mathcal{T}_n$. Let $F \in \mathcal{T} \subset \mathcal{F}_\infty$ and $\eta = \mathbb{1}_F$. (This means it is an indicator function, which is 1 when an event happens and 0 when it does not.) By Levy's Upward Theorem,

$$\eta = \mathbb{E}[\eta \mid F_\infty] = \lim_{n\to\infty} \mathbb{E}[\eta \mid \mathcal{F}_n].$$

Since $\eta$ is independent of every $\mathcal{F}_n$, $\mathbb{E}[\eta \mid \mathcal{F}_n] = E[\eta] = \mathbb{P}[F]$. Hence $\eta = \mathbb{P}[F]$, a.s. Since $\eta$ can only be 0 or 1, the result follows. ■

*Example.* One example of a tail event is the probability that a random walk on the integers, starting at 0 and biased with a probability $p > \frac{1}{2}$ of going one step to the right and $1 - p$ of going to the left, will reach (1) all positive $x$, and (2) all negative $x$. By the Zero-One Law, the probability must be 0 or 1, and indeed it is 1 in the former and 0 in the latter.

For our finale, we shall prove the strong form of the famous Law of Large Numbers through applying martingale theory. We begin with one more needed theorem, the Lévy Downwards Theorem.

**Theorem 8.4** (Lévy Downwards Theorem)**.** *Let $\mathcal{G}_{-n} \subset \cdots \subset \mathcal{G}_{-1} \subset \mathcal{F}$. Let $\mathcal{G}_{-\infty} = \cap_k \mathcal{G}_{-k}$. Let $X$ be a $\mathcal{L}^1$ random variable. Then*

$$\mathbb{E}[X \mid \mathcal{G}_{-\infty}] = \lim_{n\to\infty} \mathbb{E}[X \mid \mathcal{G}_{-n}].$$

*Proof.* Let $M_{-n} = \mathbb{E}[X \mid \mathcal{G}_{-n}]$. From 5.4 we find that $M_{-n}$ is a martingale starting at $-N$ and ending at $M_{-1}$. Using the Upcrossing Lemma, we find that this converges to $M_\infty$ and then proceed as in the Upwards Theorem. ■

Now the main proof:

**Theorem 8.5** (Strong Law of Large Numbers)**.** *Let $X_1, X_2, \ldots$ be independent identically distributed variables, with $\mathbb{E}[|X_1|] < \infty$ and $m = \mathbb{E}[X_1]$. Define $S_n = \sum_{i=1}^n X_i$. Then, $\frac{S_n}{n} \to m$ a.s. and $\mathbb{E}[\frac{S_n}{n} - m] \to 0$.*

*Proof.* Let $\mathcal{G}_{-n} = \sigma(S_n, S_{n+1}, \ldots)$. By 5.4, we then have $\mathbb{E}[X_1 \mid \mathcal{G}_{-n}] = \mathbb{E}[X_1 \mid \sigma(S_n)]$. However, as $\mathbb{E}[X_1 \mid \sigma(S_n)] = \mathbb{E}[X_i \mid \sigma(S_n)]$ for $i < n$ from the independent and identically distributed condition, we have

$$\mathbb{E}[X_1 \mid \sigma(S_n)] = \frac{\mathbb{E}[X_1 + X_2 + \cdots + X_n \mid \sigma(S_n)]}{n} = \frac{S_n}{n}.$$

Now $\mathbb{E}[X_1 \mid \mathcal{G}_{-n}]$ satisfies the conditions for the Lévy Downward Theorem, so we see that $L = \lim \frac{S_n}{n}$ exists and converges in $\mathcal{L}^1$. $L = \limsup \frac{X_{k+1} + \cdots + X_{k+n}}{n}$ for each $k$, so we can apply Kolmogorov's Zero-One Law to find that $L$ is a constant, since it does not depend on the first $k$ $S_k$ for any $k$. However, $L = \mathbb{E}[L] = m$, so we are done. ■

The Law of Large Numbers is key in statistics. If $X_n$ are independent observations of a process at different times or the data of a limited sample of a population, then this theorem formally shows that the mean of the variables will approach the full average of the distribution as the sample size grows.

## References


[1] Lennard Bakker. *Fatou's Lemma and the Dominated Convergence Theorem*. URL: `https://math.byu.edu/~bakker/Math346/Lectures/M346Lec16.pdf`. Last visited on 2020/11/21. Brigham Young University.

[2] Lennard Bakker. *Fatou's Lemma and the Monotone Convergence Theorem*. URL: `https://math.byu.edu/~bakker/Math541/Lectures/M541Lec29.pdf`. Last visited on 2020/11/21. Brigham Young University.

[3] David Williams. *Probability with Martingales*. Cambridge University Press, 1991.