Martingales

Lee De Zhang

December 19, 2020

1 Introduction

In this document, we discuss and prove several properties related to martingales.

A martingale, X_n, can be seen as the fortune of a gambler after the n-th fair game (i.e., equal odds of wining or losing). A sub (super)-martingale is when he bets on a favourable (unfavourable) game. It is of interest to study some properties of X_n, such as $\mathbb{E}X_n$, and sufficient conditions for convergence in L^1 and subsequently L^p, $p > 1$.

To recap, we will first state the definition of a random variable.

Definition 1.1 (*Random Variable*)**.** *Given a probability space* $(\Omega, \mathcal{F}, P)$*, then a real valued random variable* X *is a measurable map* $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$*, where* $\mathcal{B}$ *is the Borel* σ*-algebra on* $\mathbb{R}$*, with distribution* $P \cdot X^{-1}$*, that is,*

$$\forall A \in \mathcal{B},\ P(X \in A) = P(w : X(w) \in A) = P(X^{-1}(A)).$$

In a more familiar context, $X \in A$ simply means that the realization of the random event X is in A.

2 Conditional Expectation

We first state the definition of conditional expectation in the measure theoretic setting.

Definition 2.1 (*Conditional Expectation*)**.** *Given a probability space* $(\Omega, \mathcal{F}_0, P)$*, and a* σ*-field* $\mathcal{F} \subset \mathcal{F}_0$*, and some random variable* $X \in \mathcal{F}_0$ *with* $\mathbb{E}|X| < \infty$*. Then the conditional expectation* $X|\mathcal{F}$ *is a random variable* Y*, such that,*

1. $Y \in \mathcal{F}$*, i.e.,* Y *is measurable in* $\mathcal{F}$*.*
2. $\forall A \in \mathcal{F}$*,* $\int_A X\, dP = \int_A Y\, dP$*.*

The first order of business is to prove that the conditional expectation Y exists and is unique.

Proposition 2.2 (*Uniqueness*)**.** *If* Y *satisfies the above conditions, then it is integrable and unique.*

Proof. We first show that Y is integrable, $\mathbb{E}|Y| < \infty$. Recall the definition of integration,

$$\int f\, d\mu = \int f^+\, d\mu - \int f^-\, d\mu,$$

where $f^+(x) = \max(f(x), 0)$, and $f^-(x) = \min(f(x), 0)$. Therefore,

$$\int Y\, dP = \int Y^+\, dP - \int Y^-\, dP.$$

Let $A = \{Y > 0\} \subset \mathcal{F}$. Then,

$$\int Y^+\, dP = \int_A Y\, dP = \int_A X\, dP \leq \mathbb{E}|X| < \infty,$$

and similarly,

$$\int Y^-\, dP = \int_{A^c} Y\, dP = \int_{A^c} X\, dP \leq \mathbb{E}|X| < \infty,$$

Therefore $\mathbb{E}Y = \int Y\, dP < \infty$.

To prove uniqueness, suppose there exists random variable Z which satisfies the definition, and $Z \neq Y$. Then define $A = \{|Y - Z| > \epsilon\}$ for some arbitrary $\epsilon > 0$. Then,

$$0 = \int_A X - X\, dP = \int_A Z - Y\, dP \geq \epsilon P(A).$$

Since $\epsilon > 0$ is arbitrary, $P(A) = 0$. □

The next part would be to prove the existence of conditional probabilities. To do so, we first state the following definitions.

Definition 2.3 (*Absolute Continuity*)**.** *Let μ and υ be measures on a common probability space. Then υ is absolutely continuous to μ (abbreviated as $\upsilon << \mu$ if*

$$\mu(A) = 0 \Rightarrow \upsilon(A) = 0.$$

Definition 2.4 (*Radon-Nikodym derivative*)**.** *Suppose $\upsilon << \mu$ be σ-finite measures on $(\omega, \mathcal{F})$, then the function f such that for all $A \in \mathcal{F}$,*

$$\upsilon(A) = \int_A f\, d\mu,$$

is the Radon-Nikodym derivative, written as, $f = d\upsilon/d\mu$.

To prepare for the proof, we also first state the dominated convergence theorem.

Theorem 2.5. *Suppose $\{f_n\}$ is a sequence of measurable functions on some measurable space with measure μ, and f_n converges pointwise to f. If there exists some integrable function g (i.e. $\int |g|\, d\mu < \infty$), such that,*

$$\forall n\, |f_n(x)| < g(x),$$

then f is integrable,

$$\int_\Omega |f_n - f|\, d\mu \to 0,$$

which implies,

$$\int_\Omega f_n\, d\mu \to \int_\omega f\, d\mu.$$

Corollary 2.6. *Let $\mu = P$ be a measure of random variable $X \geq 0$, and $\upsilon(A) = \int_A X \, dP$. Then by the definition of the integral, $\upsilon << \mu$, X is the Radon-Nikodym derivative, and the dominated convergence theorem implies that υ exists and hence, is a measure.*

To remove the restriction of $X \geq 0$, write $X = X^+ + X^-$, then $\int_A X^+ \, dP$ and $\int_A X^- \, dP$ exists for $A \subset \mathcal{F}$. Therefore $Y = \mathbb{E}(X|\mathcal{F})$ exists.

Next, we state some properties of conditional expectation.

Proposition 2.7 (*Properties of Conditional Expectation*)**.**

1. *If $\mathbb{E}|X|, \mathbb{E}|Y| < \infty$, then*

$$\mathbb{E}(aX + bY|\mathcal{F}) = a\mathbb{E}(X|\mathcal{F}) + b\mathbb{E}(Y|\mathcal{F}).$$

2. *If $\mathbb{E}|X|, \mathbb{E}|Y| < \infty$ and $X \leq Y$, then*

$$\mathbb{E}(X|\mathcal{F}) \leq \mathbb{E}(Y|\mathcal{F}).$$

3. *If $X_n \geq 0, X_n \uparrow X$, and $\mathbb{E}X < \infty$, then,*

$$\mathbb{E}(X_n|\mathcal{F}) \uparrow \mathbb{E}(X|\mathcal{F}).$$

Proof. Linearity can be verified by checking, for all $A \subset \mathcal{F}$,

$$\int_A a\mathbb{E}(X|\mathcal{F}) + b\mathbb{E}(Y|\mathcal{F}) \, dP = \int_A aX \, dP + \int_A bY \, dP = \int_A aX + bY \, dP$$

The second statement is because for any $A \subset \mathcal{F}$,

$$\int_A \mathbb{E}(X|\mathcal{F}) \, dP = \int_A X \, dP \leq \int_A Y \, dP = \int_A \mathbb{E}(Y|\mathcal{F}) dP,$$

therefore,

$$\int_A \mathbb{E}(X|\mathcal{F}) - \mathbb{E}(Y|\mathcal{F}) \, dP \leq 0.$$

If we define $A_\epsilon = \{\mathbb{E}(X|\mathcal{F}) - \mathbb{E}(Y|\mathcal{F}) > \epsilon\}$ for arbitrary $\epsilon > 0$, then clearly $A_\epsilon \subset \mathcal{F}$, and since,

$$\epsilon P(A_\epsilon) \leq \int_{A_\epsilon} \mathbb{E}(X|\mathcal{F}) - \mathbb{E}(Y|\mathcal{F}) \, dP \leq 0,$$

hence $P(A_\epsilon) = 0$.

For the third claim, we use the dominated convergence theorem and the second result. □

We can derive an equality similar to Jensen's inequality for conditional expectations. We will use $L^1(\Omega, \mathcal{F}, P)$ to denote the set of functions integrable in L^1 for the probability space, i.e., for any $f \in L^1(\Omega, \mathcal{F}, P)$ and $A \subset \mathcal{F}$, $\int_A f \, dP < \infty$.

Proposition 2.8 (*Jensen-like Inequality*)**.** *Suppose $\phi \in L^1(\Omega, \mathcal{F}, P)$ is convex and $\mathcal{G} \subset \mathcal{F}$, then,*

$$\mathbb{E}(\phi(X)|\mathcal{G}) \geq \phi(\mathbb{E}(X|\mathcal{G})) \quad a.s.$$

Proof. If ϕ is linear, the result is trivial. Otherwise, any convex function ϕ can be written as,

$$\phi(x) = \sup_a (ax - \psi(a)),$$

for some convex ψ. Therefore,

$$\begin{aligned}
\mathbb{E}(\phi(X)|\mathcal{G}) &= \mathbb{E}(\sup_a (aX - \psi(a))|\mathcal{G}) \\
&\geq \sup_a \mathbb{E}((aX - \psi(a))|\mathcal{G}) \\
&= \sup_a (a\mathbb{E}(X|\mathcal{G}) - \psi(a)) \\
&= \phi(\mathbb{E}(X|\mathcal{G}).
\end{aligned}$$

As a technicality, we should restrict our attention to $a \in \mathbb{Q}$, as conditional expectation is uniquely determined up to a set of measure 0, and the union of an uncountable number of sets of measure 0 may have positive measure. Since $\mathbb{Q}$ is countable, it is appropriate. $\square$

Before we conclude this section, we present one more property.

Proposition 2.9. *Suppose $X \in \mathcal{F}$, and $\mathbb{E}|Y|, \mathbb{E}|XY| < \infty$, then,*

$$\mathbb{E}(XY|\mathcal{F}) = X\mathbb{E}(Y|\mathcal{F}).$$

Proof. Suppose $X = I_B$ for some $B \in \mathcal{F}$, then for some $A \in \mathcal{F}$,

$$\int_A I_B \mathbb{E}(Y|\mathcal{F})\, dP = \int_{A\cap B} \mathbb{E}(Y|\mathcal{F})\, dP = \int_{A\cap B} Y\, dP = \int_A I_B Y\, dP = \int_A XY\, dP$$

$\square$

3 Regular Conditional Probabilities

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and random variable $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ be a measurable map (i.e. X is a random variable taking values in a general space S, and a σ-field $\mathcal{S}$, a σ-field $\mathcal{G} \subset \mathcal{F}$. Under this setting, we state the definition of a regular conditional distribution.

Definition 3.1 (*Regular Conditional Distribution*)**.** *Given the above, $\mu : \Omega \times \mathcal{S} \to [0, 1]$ is a regular conditional distribution for X given $\mathcal{G}$ if,*

1. *For each A, $\omega \to \mu(\omega, A)$ is a version of $P(X \in A|\mathcal{G})$.*
2. *For a.e., $\omega, A \to \mu(\omega, A)$ is a probability measure on $(S, \mathcal{S})$.*

Definition 3.2 (*Regular Conditional Probability*)**.** *If $S = \Omega$, and X is the identity map, then μ is a regular conditional probability.*

To facilitate understand, we also present an alternative but equivalent definition of regular conditional probabilities and distributions.

Definition 3.3 (*Regular Conditional Distribution and Probabilities*)**.** *Let* $(\Omega, \mathcal{F}, P)$*,* $\mathcal{G}$*,* $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ *as above. Then the* ***regular conditional distribution*** *of* X *given* $\mathcal{G}$ *is the family of distributions* $(\mu(\omega, .))_{\omega \in \Omega}$ *on* $(S, \mathcal{S})$*, such that for all* $A \in \mathcal{S}$*,* $\mu(., A) = \mathbb{E}(1 + A)X_{|}\mathcal{G}]$ *a.s.*

If $(S, \mathcal{S}) = (\Omega, \mathcal{F})$*, and* $X(\omega) = \omega$*, then* $(\mu(\omega, .))_{\omega \in \Omega}$ *is a regular conditional probability on* $\mathcal{F}$ *given* $\mathcal{G}$*.*

For convenience, we will use the abbreviations rcp and rcd to denote regular conditional probability and distribution. Next, we will present a result which expresses the expectation of X given $\mathcal{G}$ as an integral over the rcd.

Proposition 3.4 (*Expectation over rcd*)**.** *Let* $(\mu(\omega, \cdot))_{\omega \in \Omega}$ *be a rcd of* X *given* $\mathcal{G}$*. Then for any Borel measurable function* $f : (S, \mathcal{S}) \to (\mathbb{R}, \mathcal{B})$*, with* $\mathbb{E}|f(X)| < \infty$*, we have,*

$$\mathbb{E}[f(X)|\mathcal{G}] = \int f(x)\mu(\omega, dx), \quad a.s.$$

Proof. By writing f as the sum of its positive and negative parts, we assume wlog that $f \geq 0$. By definition, the above holds when f is an indicator function, and hence when f is the linear combination of linear functions (simple function). Since any non-negative measurable function is the increasing limit of a sequence of simple functions, by the monotone convergence theorem, the integral exists. □

4 Martingales

Martingales capture the notion of fair future returns, given past information. It was initially deveoped as a class of betting strategies popular in the past. We will focus on discrete time martingales. We will start with some definitions to guide our discussion.

Definition 4.1 (*Filtration*)**.** *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space. A filtration* $(\mathcal{F}_n)_{n \in \mathbb{N}}$ *is an increasing sequence of sub* σ*-algebra of* $\mathcal{F}$*, i.e.,*

$$\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}.$$

We can view $\mathcal{F}_n$ *as the information available at time* n*.*

Definition 4.2 (*Adaptation*)**.** *A sequence of random variables* X_n *is adapted to a sequence of* σ*-fields* $\mathcal{F}_n$ *if* $X_n \in \mathcal{F}_n$ *for all* n*.*

Definition 4.3 (*Martingale, Super-Martingale, Sub-Martingale*)**.** *If* X_n *is a sequence such that*

1. $E|X_n| < \infty$*,*
2. X_n *is adapted to* $\mathcal{F}_n$*,*
3. $E(X_{n+1}|\mathcal{F}) = X_n$ *for all* n*,*

then X_n *is a martingale with respect to* $\mathcal{F}_n$*. Replace the* $=$ *in the third condition with* $\leq$ *or* $\geq$ *for super or sub martingale.*

If the filtration is not stated explicitly, we take $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$, i.e. the σ algebra generated by the random variables.

We will now state some examples of martingales.

Example 4.1 (*Mean Zero Random Walk*)**.** *If $X_1, X_2, \cdots$ are iid random variables, $\mathbb{E}X_n = 0$, then $Y_n = \sum_{i=1}^n X_n$ is a martingale adapted to filtration $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$. In this case, Y_n records the position of a random walk on $\mathbb{R}$.*

Proposition 4.4. *If X_n is a martingale wrt $\mathcal{G}_n$, let $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$, then $\mathcal{G}_n \supset \mathcal{F}_n$, and X_n is a martingale wrt $\mathcal{F}_n$.*

Proof. Clearly $\mathcal{F}_n$ is a filtration. Since $\mathcal{F}_n$ is the smallest σ-field containing $X_1, \cdots, X_n$, then $\mathcal{F}_n \subset \mathcal{G}_n$. We verify,

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) = \mathbb{E}(\mathbb{E}(X_{n+1}|\mathcal{G}_n)|\mathcal{F}_n) = \mathbb{E}(X_n|\mathcal{F}_n) = X_n.$$

□

We present one example relating to a betting strategy.

Example 4.2 (*Martingale Transform as Betting Strategy*)**.** *Let the martingale difference $D_n = X_n - X_{n-1}$ as the reward/loss of the n-th game, in a sequence of (possibly dependent games), and $X_0 = 0$ for convenience (although it doesn't matter. Then a martingale corresponds to a fair game, because $\mathbb{E}X_n = \mathbb{E}X_0$.*

A martingale transform is defined by $X'_n = X'_{n-1} + h_{n-1}D_n$, where $h_{n-1} \in \mathcal{F}_{n-1}$ such that $h_{n-1}D_n$ is integrable.

In the context of betting, we interpret h_{n-1} as the size of the bet in the n-th game, and $h_{n-1} \in \mathcal{F}_{n-1}$ means that we can choose our bet based on information right up to just before the n-th game.

We verify that X'_n is a martingale, since

$$\mathbb{E}(X'_{n+1}|\mathcal{F}_n) = \mathbb{E}(X'_n + h_{n-1}D_n|\mathcal{F}_n) = X'_n + \mathbb{E}(h_{n-1}D_n|\mathcal{F}_n) = X'_n + h_{n-1}\mathbb{E}(D_n|\mathcal{F}_n),$$

and $\mathbb{E}(D_n|\mathcal{F}_n) = 0$ since $\mathbb{E}X_n = X_0$ for all n (fair game).

Therefore, this is a martingale, and the game still remains fair ($\mathbb{E}(X'_{n+1}|\mathcal{F}_n) = X'_0$)

4.1 Martingale Decomposition

Let $X \in L_1(\Omega, \mathcal{F}, P)$. A useful technique in bounding the variance of X or establish concentration properties of X is through a martingale decomposition. We introduce a filtration $\mathcal{F}_0 := \{\emptyset, \Omega\} \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n = \mathcal{F}$. Let $X_i = \mathbb{E}(X|\mathcal{F}_i)$, and write,

$$X = \mathbb{E}X + \sum_{i=1}^n (X_i - X_{i-1}).$$

Note that the X_i's are martingales. Furthermore if $X \in L_2(\Omega, \mathcal{F}, P)$, then,

$$Var(X) = \mathbb{E}((X - \mathbb{E}X)^2) = \sum_{i=1}^n \mathbb{E}((X_i - X_{i-1})^2) = \sum_{i=1}^n \mathbb{E}((Var(X_i|\mathcal{F}_{i-1})),$$

and the conditional variance,

$$Var(X_i|\mathcal{F}_{i-1}) := \mathbb{E}(X_i^2|\mathcal{F}_{i-1}) - \mathbb{E}(X_i|\mathcal{F}_{i-1}),$$

can often be bounded using coupling techniques.

To illustrate martingale decomposition, we prove a concentration of measure inequality for martingales with bounded increments.

Theorem 4.5 (*Azuma-Hoeffding Inequality*)**.** *Let* $(X_i)_{1\le i\le n}$ *be martingales adopted to filtration* $(\mathcal{F}_i)_{1\le 1\le n}$ *on* $(\Omega, \mathcal{F}, P)$*. Wlog, assume* $X_0 = \mathbb{E}X_1 = 0$*, and* $|X_i - X_{i-1}| \le K$ *for* $1 \le i \le n$ *a.s.. Then for all* $x \ge 0$*,*

$$P\left(\frac{X_n}{n} \ge x\right) \le \exp\left(-\frac{x^2}{2K^2}\right).$$

Proof. Let $D_i = X_i - X_{i-1}$. By the exponential Markov inequality,

$$P(X_n \ge y) = P(e^{\lambda X_n} \ge e^{\lambda y}) \le e^{\lambda y}\mathbb{E}(e^{\lambda X_n}) = e^{\lambda y}\mathbb{E}(e^{\lambda X_{n-1}}\mathbb{E}(e^{\lambda D_n}|\mathcal{F}_{n-1})).$$

Since $|X_i - X_{i-1}| \le K$, for all $x \in [-K, K]$, we have,

$$e^{\lambda x} \le \frac{e^{\lambda K} + e^{-\lambda K}}{2} + \frac{e^{\lambda K} - e^{-\lambda K}}{2K}x,$$

and since $|D_n| \le K$ a.s and $\mathbb{E}(D_n|\mathcal{F}_{n-1}) = 0$,

$$\mathbb{E}(e^{\lambda D_n}|\mathcal{F}_{n-1}] \le \frac{e^{\lambda K} + e^{-\lambda K}}{2} \le e^{\frac{\lambda^2 K^2}{2}}.$$

Therefore,

$$P(X_n \ge y) \le \exp\left(-\lambda y + \frac{n\lambda^2 K^2}{2}\right)$$

Since λ is arbitrary, optimizing for λ yields,

$$P(X_n \ge y) \le \exp\left(-\frac{y^2}{2nK^2}\right).$$

Substituting $y = x\sqrt{n}$ yields the desired result. □