Martingales and Stochastic Calculus

Alex Rutar[*]
University of Waterloo

Winter 2020[†]

[*] *arutar@uwaterloo.ca*
[†]Last updated: January 20, 2020

Contents

I. Stochastic Calculus

1   MEASURE THEORY FOR PROBABILITY

Definition. Given a measure space $(\Omega, \mathcal{F}, \mathbb{P})$, a measurable function $X : \Omega \to \mathbb{R}$ is called a **random variable**. If X is a random variable, then we define the **distribution** of X to be the measure Borel measure μ on $\mathbb{R}$ given by $\mu(E) = \mathbb{P}(X^{-1}(E))$.

1.1   CONDITIONAL EXPECTATION

1.1 Theorem. (Kolmogorov) *Suppose $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} \subset \mathcal{F}$ is a sub-σ-algebra. Then there exists some $Z \in L^1(\Omega, \mathcal{G}, \mathbb{P})$ such that for each $A \in \mathcal{G}$*

$$\int_A X \, d\mathbb{P} = \int_A Z \, d\mathbb{P}.$$

Moreover, if $\tilde{Z} \in L^1(\Omega, \mathcal{G}, \mathbb{P})$ satisfies the above constraint, then $\mathbb{P}\{x \in \Omega : \tilde{Z}(x) = Z(x)\} = 1$.

Definition. In the context of the above theorem, we write $Z = \mathbb{E}[X|\mathcal{G}]$ and call $\mathbb{E}[X|\mathcal{G}]$ a **conditional expectation** with respect to $\mathcal{G}$. Certainly $\mathbb{E}[X|\mathcal{G}]$ nee not be pointwise unique.

1.2 Theorem. (Properties of Conditional Expectation) *Suppose X, Y are random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ with $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra. Then*
1. *$\mathbb{E}[X|\mathcal{G}] \geq 0$ a.s. whenever $X \geq 0$ a.s.*
2. *For $\alpha, \beta \in \mathbb{R}$ and $Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, $\mathbb{E}[\alpha X + \beta Y|\mathcal{G}] = \alpha \mathbb{E}[X|\mathcal{G}] + \beta \mathbb{E}[Y|\mathcal{G}]$ a.s.*
3. *If $XY \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and Y is $\mathcal{G}$−measurable, then $\mathbb{E}[XY|\mathcal{G}] = Y \mathbb{E}[X|\mathcal{G}]$*
4. *If $\mathcal{H} \subset \mathcal{G}$ is a σ-algebra, then $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$ a.s.*
5. *$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]$*
6. *If $\mathcal{H}$ is a σ-algebra which is independent of the σ−algebra $\sigma\{\sigma\{X\}, \mathcal{G}\}$, then $\mathbb{E}[X|\sigma\{\mathcal{G}, \mathcal{H}\}] = \mathbb{E}[X|\mathcal{G}]$. In particular, $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$ a.s. whenever $\sigma\{X\}$ and $\mathcal{H}$ are independent.*

1.2   STOCHASTIC PROCESSES

Definition. A **stochastic process** $X = \{X_t\}_{t \in \mathcal{T}}$ is a collection of random variables defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Typically, we assume that $\mathcal{T} = \mathbb{Z}_{\geq 0}$ or $\mathcal{T} = \mathbb{R}_{\geq 0}$ and equip $\mathcal{T}$ with the order topology Intuitively, we expect t to be a discrete or continuous time parameter. Given some $\omega \in \Omega$, the map $t \mapsto X_t(\omega)$ is called a **realization** or **path** of this process. One of the goals of this section is to treat $\{X_t\}_{t \geq 0}$ as a random element in some path space, equipped with a proper σ−algebra and probability.

Definition. We say that a stochastic process is L^p if X_t is L^p for each $t \in \mathcal{T}$. Then we say that a stochastic process is **uniformly L^p** if there exists some $B \in \mathbb{R}$ so that $\int_{\Omega} |X_t|^p \, d\mathbb{P} \leq B$ for each $t \in \mathcal{T}$.

Consider $X_t(\omega)$ as a function $X : \mathcal{T} \times \Omega \to \mathbb{R}$ equipped with the product σ−algebra.

Definition. The **distribution** of a stochastic process is the collection of all its finite-dimensional distributions, i.e. the collection of all the distributions of $\{X_{t_1}, \ldots, X_{t_k}\}$ for and $k \in \mathbb{N}$ and $t_i \in \mathcal{T}$.

There are a number of ways to say that two processes X and Y are equivalent, which we organize in decreasing order of strength.

Definition. Let $X = \{X_t\}_{t \in \mathcal{T}}$ and $Y = \{Y_t\}_{t \in \mathcal{T}}$ be stochastic processes.

• We say that X and Y are called **indistinguishable** if almost all their sample paths agree; in other words,

$$\mathbb{P}(\omega \in \Omega : X_t(\omega) = Y_t(\omega) \text{ for all } t \in \mathcal{T}) = 1.$$

• We say that Y is a **modification** of X if for each $t \in \mathcal{T}$ we have $\mathbb{P}(\omega \in \Omega : X_t(\omega) = Y_t(\omega)) = 1$.
• Finally, X and Y are said to have the **same distribution** if all the finite dimensional distributions agree. In other words, if for all $n \in \mathbb{N}$ and $t_1 < \cdots < t_n$, with $t_i \in \mathcal{T}$, we have $(X_{t_1}, \ldots, X_{t_n}) \overset{d}{=} (Y_{t_1}, \ldots, Y_{t_n})$.

Example. Let X be a continuous stochastic process and N a Poisson point process on $[0, \infty)$. Then define

$$Y_t := \begin{cases} X_t & : t \notin N \\ X_t + 1 & : t \in N \end{cases}$$

Thus $\mathbb{P}(X_t = Y_t) = 1$ for all t, so X is a modification of Y. However, $\mathbb{P}(X_t = Y_t, t \geq 0) = 0$, so that X and Y are not indistinguishable.

A filtration formalizes the idea of "information acquired over time".

Definition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A **filtration** is a non-decreasing family $\{\mathcal{F}_t\}_{t \in \mathcal{T}}$ of sub-σ-algebras of $\mathcal{F}$ so that $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ for any $s < t$ in $\mathcal{T}$. We write $F_\infty = \sigma(\bigcup_{t \in \mathcal{T}} \mathcal{F}_t)$.

Let $\{X_t\}_{t \in \mathcal{T}}$ be a stochastic process.

Definition. The **filtration generated by** $\{X_t\}_{t \in \mathcal{T}}$ is $\{\sigma(\{X_s : s \leq t\})\}_{t \in \mathcal{T}}$. In other words $\mathcal{F}_t$ is the smallest σ-algebra which makes X_s measurable for all $s \leq t$ in $\mathcal{T}$. A stochastic process $\{X_t\}_{t \in \mathcal{T}}$ is called **adapted** to a filtration $\{\mathcal{F}_t\}_{t \in \mathcal{T}}$ if X_t is $\mathcal{F}_t$-measurable for every $t \in \mathcal{T}$.

Clearly, the filtration generated by $\{X_t\}_{t \in \mathcal{T}}$ is the smallest filtration which makes $(X_t)_{t \in \mathcal{T}}$ adapted.

Definition. A filtration $\{\mathcal{F}_t\}_{t \in \mathcal{T}}$ is said to satisfy the "usual condition" if
  (i) it is right-continuous: $\lim_{s \to t^+} := \bigcap_{s > t} \mathcal{F}_s = \mathcal{F}_t$, and
  (ii) $\mathcal{F}_0$ contains all the $\mathbb{P}$−null events in $\mathcal{F}$.

2 Martingale Theory

2.1 Stopping Times

Consider a filtered space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in \mathcal{T}}, \mathbb{P})$, i.e. a probability space equipped with a filtration.

Definition. A random time $N : \Omega \to \mathcal{T}$ is called a **stopping time** if $N^{-1}([0, t]) \in \mathcal{F}_t$ for each $t \in \mathcal{T}$.

The philosophy here is that we know that a stopping time happens when it happens.

Example. (i) Constants are trivial stopping times.
  (ii) Last hit a constant before some fixed bound is not a stopping time

2.1 Proposition. *If T, S are stopping times, $\min\{T, S\}$, $\max\{T, S\}$, and $T + S$ are stopping times.*

Proof   • We have $\min\{T, S\}^{-1}([0, t]) = T^{-1}([0, t]) \cap S^{-1}([0, t]) \in \mathcal{F}_t$.
  • Similarly, we have $\max\{T, S\}^{-1}([0, t]) = T^{-1}([0, t]) \cup S^{-1}([0, t]) \in \mathcal{F}_t$.
  • For $T + S$, we have that

$$(T + S)^{-1}((t, \infty)) = \left(T^{-1}(\{0\}) \cap S^{-1}((t, \infty))\right) \cup S^{-1}((t, \infty))$$
$$\cup \left(T^{-1}((0, t]) \cap (T + S)^{-1}((t, \infty))\right)$$

where $T^{-1}(\{0\}) \cap S^{-1}((t, \infty)) \in \mathcal{F}_t$ and $S^{-1}((t, \infty)) \in \mathcal{F}_t$. To finish the proof, since T has a countable dense subset Q, we have

$$T^{-1}((0, t]) \cap (T + S)^{-1}((t, \infty)) = \bigcup_{r \in Q \cap (0, t)} S^{-1}((r, t]) \cap T^{-1}((t - r, \infty))$$

where the expressions in the union are certainly $\mathcal{F}_t$-measurable. ∎

Definition. The **σ–algebra generated by a stopping time T** is the set of all events A for which $A \cap T^{-1}([0, t]) \in \mathcal{F}_t$ for every $t \in \mathcal{T}$.

Intuitively, this is the information you collect until the stopping time. Note that this σ-algebra is not the same as the σ-algebra generated by the random variable T. It is left as an exercise to the reader to verify that the above collection is indeed a σ-algebra.

  We write $X_{T \wedge t}$ is a random variable evaluated at time $T \wedge t$ (or T); in other words, $(X_{T \wedge t})(\omega) = X_{T \wedge t}(\omega)$. Then $\{X_{T \wedge t}\}_{t \in \mathcal{T}}$, or X^T, is a stochastic process stopped at time t.

2.2 Doob's Upcrossing Inequality

Definition. Consider a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in \mathcal{T}}, \mathbb{P})$. A $\{\mathcal{F}_t\}_{t \in \mathcal{T}}$–adapted proces $\{X_t\}_{t \in \mathcal{T}}$ is said to be a **submartingale** if
  (i) for all $t \in \mathcal{T}$, $X_t \in L^1(\Omega, \mathcal{F}, \mathcal{P})$, and
  (ii) for all $s < t$ where $s, t \in \mathcal{T}$,
$$\mathbb{E}(X_t | \mathcal{F}_s) \geq X_s$$

  almost surely.
and is said to be a **supermartingale** if condition (ii) above is replaced with
 (ii') for all $s < t$ where $s, t \in \mathcal{T}$,
$$\mathbb{E}(X_t | \mathcal{F}_s) \leq X_s$$

  almost surely.
Then $\{X_t\}_{t \in \mathcal{T}}$ is a **martingale** if it is both a submartingale and supermartingale.

If X is a submartingale and $0 \leq s < t$ are fixed times, then $\mathbb{E}(X_0) \leq \mathbb{E}(X_s) \leq \mathbb{E}(X_t)$. Similar statements hold in the (super)martingale cases as well. One of the goals of martingale theory is to extend these results with respect to stopping times, rather than with respect to fixed times.

Definition. Let $X = \{X_n\}_{n \in \mathbb{Z}^+}$ be a discrete time process and fix levels $a < b$ with $a, b \in \mathbb{R}$. Then the **number of upcrossings** of $[a, b]$ by X before time N with respect to the event $\omega \in \Omega$, denoted by $U_N^X([a, b], \omega)$, is the maximal $k \in \mathbb{Z}^+$ such that there exists times $0 \le s_1 < t_1 < s_2 < t_2 < \cdots < s_k < t_k \le N$ such that for all i, $X_{s_i}(\omega) \le a$ and $X_{t_i}(\omega) \ge b$.

Note that $U_N^X([a, b]) : \Omega \to \mathbb{Z}^+$ given by $\omega \mapsto U_N^X([a, b], \omega)$ is a random variable.

Definition. A process $C = \{C_n\}_{n \in \mathbb{Z}^+}$ is called **previsible** if C_n is $\mathcal{F}_{n-1}$−measurable for all $n \ge 1$. Suppose in addition that $\{X_n\}_{n \in \mathbb{Z}^+}$ is a discrete time process. Then the **martingale transform** of X by C, denoted by $C \cdot X$, is defined by

$$(C \cdot X)_n = \begin{cases} \sum_{k=1}^n C_k(X_k - X_{k-1}) & : n > 0 \\ 0 & : n = 0 \end{cases}$$

2.2 Lemma. *Suppose C is a bounded or L^2 previsible process. Then*
  (i) Let C be non-negative and X a supermartingale. Then $C \cdot X$ is a supermartingale which is null at 0.
  (ii) Let X be a martingale. Then $C \cdot X$ is a martingale which is null at 0.

PROOF We first treat the case where C is bounded.
 (i) We have by properties of conditional expectation and non-negativity of C

$$\mathbb{E}[(C \cdot X)_n - (C \cdot X)_{n-1} | \mathcal{F}_{n-1}] = \mathbb{E}[C_n(X_n - X_{n-1}) | \mathcal{F}_{n-1}]$$
$$= C_n \cdot \mathbb{E}[X_n - X_{n-1} | \mathcal{F}_{n-1}]$$
$$\le 0$$

where the second line follows since C is previsible. Integrability follows immediately since C is bounded and X_n is L^1 for each $n \in \mathbb{Z}^+$.
 (ii) Consider $C + k$ where k is a constant and $k \ge |C_n(w)|$ for all n and w.
 (iii) Similar, but the integrability is now guaranteed by Hölder's inequality. ∎

We now have the following result:

2.3 Proposition. (Doob's Upcrossing Inequality)   *(i) Let X be a supermartingale. Then*

$$(b - a) \cdot \mathbb{E}[U_N^X([a, b])] \le \mathbb{E}[(X_N - a)^-].$$

 (ii) Let X be a submartingale. Then

$$(b - a) \cdot \mathbb{E}[U_N^X([a, b])] \le \mathbb{E}[(X_N - a)^+].$$

PROOF We prove (i); the proof of (ii) is analgous. Define a process $\{C_n\}_{n \in \mathbb{Z}^+}$ by

$$C_0 := 0$$
$$C_1 := \chi_{X_0^{-1}((-\infty, a))}$$
$$C_n := \chi_{C_{n-1}^{-1}(\{1\})} \cdot \chi_{X_{n-1}^{-1}((-\infty, b])} + \chi_{C_{n-1}^{-1}(\{0\})} \cdot \chi_{X_{n-1}^{-1}((-\infty, a))}.$$

Certainly C is non-negative and bounded; previsibility follows since C_n is a characteristic function depending only on preimages of X_k for $k < n$. Thus set $Y = C \cdot X$, i.e. $Y_n = \sum_{k=1}^n C_k \cdot (X_k - X_{k-1})$ almost surely; by Lemma 2.2, Y is a supermartingale.

Since each finished upcrossing increases the value of Y by at least $b - a$, we have for any $\omega \in \Omega$

$$Y_N(\omega) \geq (b - a)U_N^X([a, b], \omega) - (X_n - a)^-(\omega)$$

where $(X_N - a)^-(\omega)$ is the upper bound for the "loss" due to the last unfinished upcrossing. Since $\mathbb{E}(Y_1) \leq 0$, we have $\mathbb{E}(Y_N) \leq \mathbb{E}(Y_1) \leq 0$. Thus

$$(b - a) \cdot \mathbb{E}[U_N^X([a, b])] \leq \mathbb{E}[(X_n - a)^-]$$

as required. ∎

2.3 Optimal Sampling Theorems

2.4 Theorem. *Suppose $\{X_t\}_{t \in \mathcal{T}}$ is uniformly L^2. Then X_t converge in L^2 to a limit X_∞.*

Proof Note that we have the result for both discrete and continuous time martingales. By discretization, it is clear that it suffices to rove the result for discrete case.

We have the following orthogonality between the increments of a martingale $\{X_n\}_{n=0}^\infty$: if $n_1 < n_2 \leq n_2 < n_4$, then

$$\mathbb{E}[(X_{n_2} - X_{n_1})(X_{n_4} - X_{n_3})] = 0.$$

This result follows by conditioning on $\mathcal{F}_{n_3}$ and applying the law of total expectation.

Now the proof proceeds. Set $Y_n := X_n - X_{n-1}$, so

$$\|X_n\|_2^2 = \mathbb{E}(X_n^2) = \sum_{i=1}^n \|Y_i\|_2^2$$

and $\sum_{i=0}^n \|Y_n\|_2^2 \leq B$ for all n, so that $\sum_{i=0}^\infty \|Y_n\|_2^2 \leq B$. Thus $\{X_n\}_{n=0}^\infty$ is Cauchy in $L^2(\Omega, \mathcal{F}, \mathbb{P})$. ∎

2.5 Theorem. (Optimal Sampling for Bounded Stopping Times in Discrete Times)
Let $\{X_n\}_{n=0}^\infty$ be a $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n=0}^\infty, \mathbb{P})$ supermartingale and S, T be $\{\mathcal{F}_n\}$–stopping times such that $0 \leq S \leq T \leq N$ for some constant $N < \infty$. Then X_T is integrable and $\mathbb{E}(X_T | \mathcal{F}_s) \leq X_s$ almost surely.

Proof Notice that

$$|X_T| \leq |X_0| + |X_1| + \cdots + |X_N|$$

so that $\mathbb{E}(|X_T|) < \infty$. To prove that $\mathbb{E}(X_t | \mathcal{F}_s) \leq X_s$ a.s., it suffices to prove that $\mathbb{E}(X_T; A) := \int_A X_T d\mathcal{P} \leq \int_A X_s d\mathcal{P} =: \mathbb{E}(X_s; A)$ for all $A \in \mathcal{F}_s$. Assuming this, then

$$\mathbb{E}(\mathbb{E}(X_T | \mathcal{F}_s) - X_s; A) \leq 0$$

for all $A \in \mathcal{F}_s$, so we may take $A = A_0 := \{\mathbb{E}(X_T | \mathcal{F}_s) - X_s > 0\}$, so that $\mathbb{P}(A_0) = 0$.

Let's prove the required statement. First note that

$$\sum_{n=1}^N \chi_{\{S < n \leq T\}}(X_n - X_{n-1}) = X_T - X_S$$

and taking expectation over A on both sides

$$\mathbb{E}(X_T - X_S; A) = \sum_{n=1}^{N} \mathbb{E}(\chi_{\{S < n \leq T\}}(X_n - X_{n-1}); A)$$

$$= \sum_{n=1}^{N} \mathbb{E}(X_n - X_{n-1}; A \cap \{S < n \leq T\})$$

But $A \cap \{S < n \leq T\} = A \cap \{S \leq n-1\} \cap \{n-1 < T\} \in \mathcal{F}_{n-1}$. Thus

$$\mathbb{E}(X_n - X_{n-1}; A \cap \{S < n \leq T\}) = \mathbb{E}(\mathbb{E}(X_n - X_{n-1}|\mathcal{F}_{n-1}); A \cap \{S < n \leq T\}) \leq 0$$

so the required statement holds. ∎

Definition. Let $\{X_n\}_{n=0}^{\infty}$ be a $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n=0}^{\infty}, \mathbb{P})$ a supermartingale. We say that $\{X_n\}_{n=0}^{\infty}$ is **closed** by a random variable X_∞ if X_∞ is $\mathcal{F}_\infty$–measurable and $X_n \geq \mathbb{E}(X_\infty|\mathcal{F}_n)$ almost surely for all $n = 0, 1, \ldots$.

Similar statements hold with $X_n \leq \mathbb{E}(X_\infty|\mathcal{F}_n)$ for a submartingale, or equality with a martingale.

2.6 Proposition. *Suppose that $\{X_n\}_{n=0}^{\infty}$ is a $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n=0}^{\infty}, \mathbb{P})$ a non-negative super-martingale, and $X_\infty = 0$. If $S, T : \Omega \to \overline{\mathbb{Z}^+}$ are $\{\mathcal{F}_n\}$–stopping times, $S \leq T$, then*
1. $\mathbb{E}(X_T) < \infty$
2. $\mathbb{E}(X_T|\mathcal{F}_s) \leq X_s$

PROOF   1. Note that $X_T \leq \liminf_{n \to \infty} X_{T \wedge n}$ where $T \wedge n$ and 0 are two bounded stopping times. Thus $\mathbb{E}(X_{T \wedge n}) \leq \mathbb{E}(X_0)$ for all $n = 0, 1, \ldots$. Thus by Fatou's lemma

$$\mathbb{E}(X_T) \leq \mathbb{E}(\liminf_{n \to \infty} X_{T \wedge n}) \leq \liminf_{n \to \infty} \mathbb{E}(X_{T \wedge n})$$

$$\leq \mathbb{E}(X_0) < \infty.$$

2. Let $A \in \mathcal{F}_s$. For $n = 0, 1, \ldots,,$

$$\mathbb{E}(X_T; A \cap \{T \leq n\}) = \mathbb{E}(X_{T \wedge n}; A \cap \{T \leq n\})$$

$$\leq \mathbb{E}(X_{T \wedge n}; A \cap \{S \leq n\})$$

$$\leq \mathbb{E}(X_{S \wedge n}; A \cap \{S \leq n\})$$

Note that $S \wedge n$ and $T \wedge n$ are two bounded stopping times with $S \wedge n \leq T \wedge n$. Also, $A \cap \{S \leq n\} \in \mathcal{F}_{S \wedge n}$. Then apply the optimal sampling theorem for bounded stopping times. By the monotone convergence theorem,

$$\lim_{n \to \infty} \mathbb{E}(X_T; A \cap \{T \leq n\})$$

and similarly for S. Thus

$$\mathbb{E}(X_T; A \cap \{T < \infty\}) \leq \mathbb{E}(X_S : A \cap \{S < \infty\})$$

so that

$$\mathbb{E}(X_T; A \cap \{T = \infty\}) = \mathbb{E}(X_S : A \cap \{S = \infty\}) = 0$$

and $\mathbb{E}(X_T; A) \leq \mathbb{E}(X_S; A)$. Since this holds for all $A \in \mathcal{F}_s$, $\mathbb{E}(X_T : \mathcal{F}_s) \leq X_s$. ∎

2.7 Lemma. Let $\{M_n\}_{n=0}^{\infty}$ be a martingale closed by M_∞. Let $T : \Omega \to \overline{\mathbb{Z}^+}$ be a stopping time. Then $M_T = \mathbb{E}(M_\infty | \mathcal{F}_T)$.

Proof First assume $M_\infty \geq 0$. For any $A \in \mathcal{F}_T$,

$$
\begin{aligned}
E(M_T; A) &= \sum_{n=0}^{\infty} \mathbb{E}(M_n; A \cap \{T = n\}) \\
&= \sum_{n=0}^{\infty} \mathbb{E}(M_\infty, A \cap \{T = n\}) \\
&= \mathbb{E}(M_\infty; A).
\end{aligned}
$$

For the general case, decompose M_∞ into positive and negative parts. ∎

2.8 Theorem. (Optimal Sampling for Closed Supermartingales in Discrete Time)
Let $\{X_n\}_{n=0}^{\infty}$ be a $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n=0}^{\infty}, \mathbb{P})$ supermartingale closed by X_∞. Let $S, T : \Omega \to \overline{\mathbb{Z}^+}$ be two $\{\mathcal{F}_n\}_{n=0}^{\infty}$ stopping times with $S \leq T$. Then
   1. $\mathbb{E}(|X_T|) < \infty$
   2. $\mathbb{E}(X_T | \mathcal{F}_s) \leq X_s$ almost surely

Proof   1. Define $M_n = \mathbb{E}(X_\infty | \mathcal{F}_n)$ and $A_n = X_n - M_n$ for $n = 0, 1, \ldots, \infty$. Since $\{X_n\}_{n=0}^{\infty}$ is a supermartingale closed by X_∞, $A_n \geq 0$, $A_\infty = 0$, and for $m \leq n$,

$$
\begin{aligned}
\mathbb{E}(A_n | \mathcal{F}_m) &= \mathbb{E}(X_n - \mathbb{E}(X_\infty | \mathcal{F}_n) | \mathcal{F}_m) \\
&\leq X_m - \mathbb{E}(X_\infty | \mathcal{F}_m) \\
&= A_m
\end{aligned}
$$

Thus $\{A_n\}_{n=0}^{\infty}$ is a non-negative supermartingale with $A_\infty = 0$.
One can prove that $\{M_n\}_{n=0}^{\infty}$ is a uniformly integrable martingale, i.e. $\lim_{c \to \infty} \sup_{n \in \mathbb{N}} \mathbb{E}(|X_n|; |X_n| > c) = 0$.
By the previous proposition, $\mathbb{E}(A_T) < \infty$. On the other hand, $\mathbb{E}(|M_T|) < \infty$ by definition of the $\{M_n\}$. Thus $\mathbb{E}(|X_T|) < \infty$.
   2. Apply the previous lemma so $\mathbb{E}(M_T | \mathcal{F}_s) = \mathbb{E}(\mathbb{E}(M_\infty | \mathcal{F}_T) \mathcal{F}_S) = \mathbb{E}(M_\infty | \mathcal{F}_s) = M_s$ by the optimal sampling theorem for closed martingales. Meanwhile, as $\{A_n\}_{n=0}^{\infty}$ is a non-negative supermartingale with $A_\infty = 0$, $\mathbb{E}(A_T | \mathcal{F}_s) \leq A_s$ almost surely. Thus $\mathbb{E}(X_T | \mathcal{F}_s) \leq X_s$ almost surely. ∎

To prepare for the optimal sampling theorem for closed supermartingales in continuous time, we introduce one more discrete time notion:

Definition. Denote by $\mathbb{Z}^- = \{z \in \mathbb{Z} : z \leq 0\}$. A **negatively indexed supermartingale** $\{X_n\}_{n \in \mathbb{Z}^-}$ and $\{\mathcal{F}_n\}_{n \in \mathbb{Z}^-}$ where $\mathcal{F}_m \subseteq \mathcal{F}_n$ for $m \leq n$. Then $X_n \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, $\{X_n\}$ is adapted and $\mathbb{E}(X_n | \mathcal{F}_m) \leq X_m$.

2.9 Theorem. Let $\{X_n\}_{n \in \mathbb{Z}^-}$ be a negatively indexed supermartingale such that $\sup_{n \in \mathbb{Z}^-} \mathbb{E}(X_n) < \infty$. Then $\{X_n\}_{n \in \mathbb{Z}^-}$ is uniformly L^1.

Proof Fix $a = \sup_{n \in \mathbb{Z}^-} \mathbb{E}(X_n)$. For any $\epsilon > 0$, get $N(\epsilon) \in \mathbb{Z}^-$ such that $\mathbb{E}(X_{N(\epsilon)}) > a - \epsilon$ for $N \leq N(\epsilon)$. Since X is a supermartingale, $0 \leq \mathbb{E}(X_n) - \mathbb{E}(X_{N(\epsilon)}) \leq \epsilon$ for all $n \leq N(\epsilon)$. For $c > 0$, note that

$$|x|\chi_{|X| > c} = -x\chi_{X < -c} - X\chi_{X \leq x} + x.$$

Since $\{X_n\}$ is a supermartingale, for $n \leq N(\epsilon)$, $\mathbb{E}(X_{N(\epsilon)}|\mathcal{F}_n) \leq X_n$ and $\{X_n < -c\} \in \mathcal{F}_n$. Thus $\mathbb{E}(X_n; X_n < -c) \geq \mathbb{E}(X_{N(\epsilon)}; X_n < -c)$. Similarly, $\mathbb{E}(X_n; X_n \leq c) \geq \mathbb{E}(X_{N(\epsilon)}; X_n \leq c)$. Moreover, $\mathbb{E}(X_n) \leq \mathbb{E}(X_{N(\epsilon)}) + \epsilon$, so $\mathbb{E}(|X_n|; X_n \leq c) \leq \mathbb{E}(X_{N(\epsilon)}; X_n \leq c) + \epsilon$ for all $N \leq N(\epsilon)$. For this $N(\epsilon)$, there exists $\delta(\epsilon) > 0$ such that $\mathbb{E}(X_{N(\epsilon)}; A) < \epsilon$ for all $A \in \mathcal{F}$ and $\mathbb{P}(A) < \delta(\epsilon)$.

Since X is a supermartingale, $\{-2X_n^-\}_{n \in \mathbb{Z}^-}$ is also a supermartingale. To see this, define $f(x) = 2\min\{x, 0\}$, so f is increasing and concave. But then for $M \leq n$, by Jensen's inequality,

$$\mathbb{E}(f(X_n)|\mathcal{F}_m) \leq f(\mathbb{E}(X_n|\mathcal{F}_m)) \leq f(X_m)$$

and $-2X_n^- = f(X_n)$ is thus a supermartingale, so $\mathbb{E}(-2X_n^-) \geq \mathbb{E}(-2X_0^-)$. We also have $\mathbb{E}(X_n) \leq a$ where a is defined as above. Thus since $|X_n| = X_n + 2X_n^-$, we have

$$\mathbb{E}(|X_n|) \leq a - \mathbb{E}(-2X_0^-) = a + 2\mathbb{E}(X_0^-) < \infty.$$

Now, choose c so that $c \cdot \delta(\epsilon) > a + 2\mathbb{E}(X_0^-) \geq \mathbb{E}(|X_n|)$. Then by Markov's ineuality, $\mathbb{P}(|X_n| > c) \leq \delta(\epsilon)$. Thus $\mathbb{E}(X_{N(\epsilon)}|; |X_n| > c) < \epsilon$ for $n \in \mathbb{Z}^-$ so that $\mathbb{E}(|X_m|; |X_n| > c) < 2\epsilon$ for all $n < N(\epsilon)$. But there are only finitely many terms with $n > N(\epsilon)$, so we are done. ∎

2.10 Theorem. (Optimal Sampling Theorem for Closed Supermartingales) Let $\{X_t\}_{t \in \mathbb{R}^+}$ be an $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in \mathbb{R}^+}, \mathbb{P})$ supermartingale which is right continuous and closed. Let $S, T : \Omega \to \overline{\mathbb{R}^+}$ be two stopping times with $S \leq T$. Then
  (i) $\mathbb{E}(|X_T|) < \infty$
  (ii) $\mathbb{E}(X_T|\mathcal{F}_s) \leq X_s$ a.s.

Definition. Define random times $T_n := 2^{-n}(\lfloor 2^n T \rfloor + 1)$. Then the T_n are decreasing stopping times with $T = \lim T_n$ pointwise from above. For each n, $\{X_{m2^{-n}}\}_{n \in \mathbb{Z}^+}$ is a discrete-time closed supermartngale. Thus $\mathbb{E}(X_{T_{n-1}}|\mathcal{F}_{T_n}) \leq X_{T_n}$ a.s. (optimal sampling theorem for closed supermartingale in discrete time). Set $Y_n = X_{T_{n-1}}$, so $\{Y_n\}_{n=-1,-2,\ldots}$ is a negatively indexed supermartingale and $\mathbb{E}(Y_n) = \mathbb{E}(X_{T_{-n-1}}) \leq \mathbb{E}(X_0) < \infty$. Thus by the previous result, $\{X_{T_n}\}_{n \in \mathbb{Z}^+}$ is uniformly L^1. Since $T_n \to T$ from above and $X_{T_n} \to X_T$ a.s., uniform integrability gives that $X_{T_n} \to X_T$ in L^1. Thus $\mathbb{E}(|X_T|) = \lim_{n \to \infty} \mathbb{E}(|X_{T_n}|) < \infty$. Similarly, define $S_n = 2^{-n}(\lfloor 2^n S \rfloor + 1)$. Both S_n and T_n can be regarded as discrete-time stopping times. Thus by the discrete optimal sampling theorem, $\mathbb{E}(X_{T_n}|\mathcal{F}_{S_n}) \leq X_{S_n}$ a.s.
  Now for $A \in \mathcal{F}_S \subseteq \mathcal{F}_{S_n}$, $\mathbb{E}(X_{T_n}; A) \leq \mathbb{E}(X_{S_n}; A)$ so

$$|\mathbb{E}(X_{T_n}; A) - \mathbb{E}(X_T; A)| \, eq \, \mathbb{E}(|X_{T_n} - X_T|) \xrightarrow{n \to \infty} c$$

Also, $|\mathbb{E}(X_{S_n}; A) - \mathbb{E}(X_S; A)| \to 0$. Therefore, as n goes to infinity,

$$\mathbb{E}(X_T; A) \leq \mathbb{E}(X_S; A)$$

for all $A \in \mathcal{F}_S$ so $\mathbb{E}(X_T|\mathcal{F}_S) \leq X_s$.

2.11 Corollary. *Let $\{X_t\}_{t\in\mathbb{R}^+}$ be a $(\Omega,\mathcal{F},\{\mathcal{F}_t\}_{t\in\mathbb{R}^+}$ martingale which is right continuous and closed. If $S,T:\Omega\to\overline{\mathbb{R}^+}$ are two stopping times with $S\le T$, then X_T is integrable and $|E(X_T|\mathcal{F}_S)=X_S$.*

2.12 Corollary. *Same results hold for bounded S,T and general right continuous martingale.*

2.13 Corollary. *A right continuous adapted process X is a martingale if and only if for every bounded stopping time T, X_T is L^1 and $\mathbb{E}(X_T)=\mathbb{E}(X_0)$.*

Proof The forward direction is contained in the above theorem. Conversely, for any $s<t$ and $A\in\mathcal{F}_s$, define $T=t\chi_{A^c}+s\chi_A$ which is a stopping time (exercise). Then $\mathbb{E}(X_0)=\mathbb{E}(X_T)=\mathbb{E}(X_t;A^c)+\mathbb{E}(X_s;A)$. On the other hand, $\mathbb{E}(X_0)=\mathbb{E}(X_t)=\mathbb{E}(X_t;A^c)+\mathbb{E}(X_t;A)$ so $\mathbb{E}(X_s;A)=\mathbb{E}(X_t;A)$. But $A\in\mathcal{F}_s$ is arbitrary, so $\mathbb{E}(X_t|\mathcal{F}_s)=\mathbb{E}(X_s)$.∎