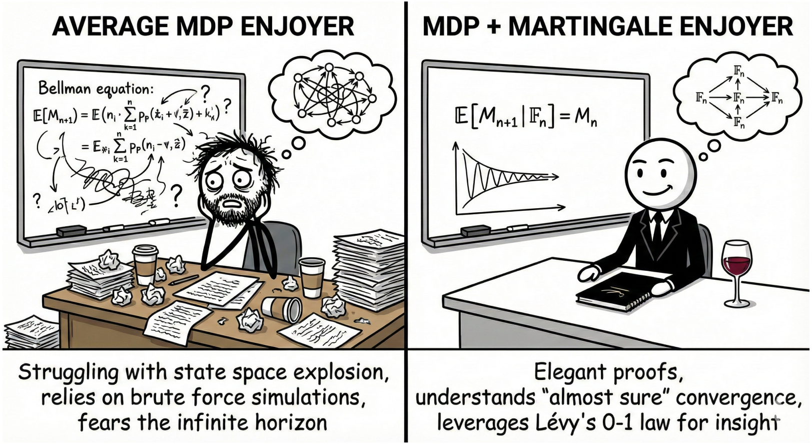
Martingale Theory for the Average MDP Enjoyer

Pierre Vandenhove

December 1, 2025 — UMONS Formal Methods Reading Group

Martingale theory for the average MDP enjoyer



- Left: me in 2019.
- Right: me in 2020, after discovering martingales.

Sample Problem

Let $\mathcal{M} = (S, s_0, P)$ be an (infinite) Markov chain and $\top \in S$.

For some $p > 0$, define the set of states from which $\top$ is reached with probability at least p:

$$\text{Reach}_{\geq p} = \{s \in S \mid \mathbb{P}_s(\Diamond \top) \geq p\}.$$

Sample Problem

Let $\mathcal{M} = (S, s_0, P)$ be an (infinite) Markov chain and $\top \in S$.

For some $p > 0$, define the set of states from which $\top$ is reached with probability at least p:

$$\mathrm{Reach}_{\geq p} = \{s \in S \mid \mathbb{P}_s(\Diamond\top) \geq p\}.$$

Theorem

If you visit $\mathrm{Reach}_{\geq p}$ infinitely often, then you reach $\top$ almost surely, i.e.,

$$\mathbb{P}_{s_0}(\Diamond\top \mid \Box\Diamond\mathrm{Reach}_{\geq p}) = 1.$$

Sample Problem

Let $\mathcal{M} = (S, s_0, P)$ be an (infinite) Markov chain and $\top \in S$.

For some $p > 0$, define the set of states from which $\top$ is reached with probability at least p:

$$\mathrm{Reach}_{\geq p} = \{s \in S \mid \mathbb{P}_s(\Diamond \top) \geq p\}.$$

Theorem

If you visit $\mathrm{Reach}_{\geq p}$ infinitely often, then you reach $\top$ almost surely, i.e.,

$$\mathbb{P}_{s_0}(\Diamond \top \mid \Box \Diamond \mathrm{Reach}_{\geq p}) = 1.$$

Intuitively true.

Sample Problem

Let $\mathcal{M} = (S, s_0, P)$ be an (infinite) Markov chain and $\top \in S$.

For some $p > 0$, define the set of states from which $\top$ is reached with probability at least p:

$$\text{Reach}_{\geq p} = \{s \in S \mid \mathbb{P}_s(\Diamond \top) \geq p\}.$$

Theorem

If you visit $\text{Reach}_{\geq p}$ infinitely often, then you reach $\top$ almost surely, i.e.,

$$\mathbb{P}_{s_0}(\Diamond \top \mid \Box \Diamond \text{Reach}_{\geq p}) = 1.$$

Intuitively true.

$\rightarrow$ Perhaps your intuition follows the second Borel-Cantelli lemma: if events have summed probability $+\infty$, they happen infinitely often.

Sample Problem

Let $\mathcal{M} = (S, s_0, P)$ be an (infinite) Markov chain and $\top \in S$.

For some $p > 0$, define the set of states from which $\top$ is reached with probability at least p:

$$\text{Reach}_{\geq p} = \{s \in S \mid \mathbb{P}_s(\Diamond \top) \geq p\}.$$

Theorem

If you visit $\text{Reach}_{\geq p}$ infinitely often, then you reach $\top$ almost surely, i.e.,

$$\mathbb{P}_{s_0}(\Diamond \top \mid \Box \Diamond \text{Reach}_{\geq p}) = 1.$$

Intuitively true.
$\rightarrow$ Perhaps your intuition follows the second Borel-Cantelli lemma: if **independent** events have summed probability $+\infty$, they happen infinitely often.

Why it is not trivial

Here, the events "$\Diamond\top$ from various states" are not independent!

Perhaps it behaves like this counter-example:

- Let X_i be the outcome of a die roll (same die, rolled once).
- Let A be the event "Obtaining 6".
- We define $A_1 = A_2 = \cdots = A$ (perfect dependence).

The sum of probabilities is infinite, but probability of "eventually" occurring is $\frac{1}{6} \neq 1$.

Why it is not trivial

Here, the events "$\Diamond \top$ from various states" are not independent!

Perhaps it behaves like this counter-example:

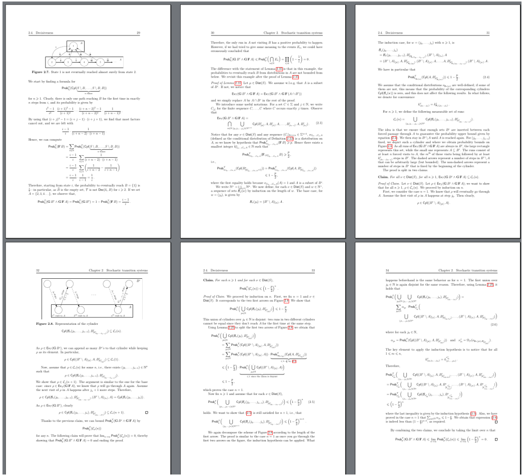
- Let X_i be the outcome of a die roll (same die, rolled once).
- Let A be the event "Obtaining 6".
- We define $A_1 = A_2 = \cdots = A$ (perfect dependence).

The sum of probabilities is infinite, but probability of "eventually" occurring is $\frac{1}{6} \neq 1$.

How to prove it, then? MARTINGALE THEORY

Before martingales

In 2019, unaware of martingales, we wrote an explicit proof for (a version of) this problem.



After martingales

In 2020, we received a comment from a reviewer "*I think this is a trivial application of martingale theory*"...

After martingales

In 2020, we received a comment from a reviewer "*I think this is a trivial application of martingale theory*"... AND IT WAS!

Proof. In order not to obfuscate the interesting ideas of the proof with technical considerations, we first prove the lemma for $n=0$ (with $\mathcal{A}=A\in\Sigma$), and explain afterwards how to extend the proof to obtain the general statement. We want to prove that for all $\mu\in\mathsf{Dist}(S)$,

$$\mathsf{Prob}_{\mu}^{\mathcal{T}}(\mathbf{G}\,B^c\wedge\mathbf{G}\,\mathbf{F}\,A)=0.$$

Let $\mu\in\mathsf{Dist}(S)$ be an initial distribution. We assume w.l.o.g. that $A\cap B=\emptyset$—indeed, if that is not the case, we simply notice that $\mathsf{Prob}_{\mu}^{\mathcal{T}}(\mathbf{G}\,B^c\wedge\mathbf{G}\,\mathbf{F}\,A)=\mathsf{Prob}_{\mu}^{\mathcal{T}}(\mathbf{G}\,B^c\wedge\mathbf{G}\,\mathbf{F}(A\cap B^c))$ and we replace A by $A\cap B^c$ in the rest of the proof.

Let us consider a modified STS $\mathcal{T}_B$ which is equal to $\mathcal{T}$, except that B is made absorbing (we assume that for $s\in B$, $\kappa(s,\cdot)$ is the Dirac distribution δ_s). Notice that $\mathsf{Prob}_{\mu}^{\mathcal{T}}(\mathbf{F}\,B)=\mathsf{Prob}_{\mu}^{\mathcal{T}_B}(\mathbf{F}\,\mathbf{G}\,B)$, and $\mathsf{Prob}_{\mu}^{\mathcal{T}}(\mathbf{G}\,B^c\wedge\mathbf{G}\,\mathbf{F}\,A)\le\mathsf{Prob}_{\mu}^{\mathcal{T}_B}(\mathbf{G}\,\mathbf{F}\,A)$ (as $A\cap B=\emptyset$, runs that see A infinitely often without seeing B in $\mathcal{T}$ are just as likely in $\mathcal{T}_B$). Notice also that the event $\mathbf{F}\,\mathbf{G}\,B$ is shift-invariant. We have

$$\begin{aligned}
&\mathsf{Ev}_{\mathcal{T}_B}(\mathbf{G}\,\mathbf{F}\,A)\\
&\quad=\{\rho=s_0s_1\ldots\in S^\omega\mid\forall i,\exists j\ge i,s_j\in A\}\\
&\quad\subseteq\{\rho=s_0s_1\ldots\in S^\omega\mid\forall i,\exists j\ge i,\mathsf{Prob}_{\delta_{s_j}}^{\mathcal{T}_B}(\mathbf{F}\,B)\ge p\} && \text{by hypothesis on } A\\
&\quad=\{\rho=s_0s_1\ldots\in S^\omega\mid\forall i,\exists j\ge i,\mathsf{Prob}_{\delta_{s_j}}^{\mathcal{T}_B}(\mathbf{F}\,\mathbf{G}\,B)\ge p\} && \text{by construction of } \mathcal{T}_B\\
&\quad=\{\rho\in S^\omega\mid\forall i,\exists j\ge i,\mathsf{E}_{\mu}^{\mathcal{T}_B}[\mathbf{1}_{\mathbf{F}\mathbf{G}B}\mid\mathcal{F}_{j+1}](\rho)\ge p\} && \text{by Lemma 17, as } \mathbf{F}\,\mathbf{G}\,B \text{ is shift-invariant}\\
&\quad\subseteq\{\rho\in S^\omega\mid\lim_{i\to\infty}\mathsf{E}_{\mu}^{\mathcal{T}_B}[\mathbf{1}_{\mathbf{F}\mathbf{G}B}\mid\mathcal{F}_i](\rho)\text{ is not 0 if it exists}\}\\
&\quad=\{\rho\in S^\omega\mid\mathbf{1}_{\mathbf{F}\mathbf{G}B}(\rho)\ne 0\} && \text{by Lévy's zero-one law (Proposition 16)}\\
&\quad=\{\rho\in S^\omega\mid\mathbf{1}_{\mathbf{F}\mathbf{G}B}(\rho)=1\}\\
&\quad=\mathsf{Ev}_{\mathcal{T}_B}(\mathbf{F}\,\mathbf{G}\,B).
\end{aligned}$$

All inclusions and equalities are almost sure. In $\mathcal{T}_B$, as $A\cap B=\emptyset$ and B is absorbing, we have that $\mathsf{Prob}_{\mu}^{\mathcal{T}_B}(\mathbf{G}\,\mathbf{F}\,A\wedge\mathbf{F}\,\mathbf{G}\,B)=0$. As $\mathsf{Ev}_{\mathcal{T}_B}(\mathbf{G}\,\mathbf{F}\,A)\subseteq\mathsf{Ev}_{\mathcal{T}_B}(\mathbf{F}\,\mathbf{G}\,B)$, this implies that $\mathsf{Prob}_{\mu}^{\mathcal{T}_B}(\mathbf{G}\,\mathbf{F}\,A)=0$.

We conclude

$$\mathsf{Prob}_{\mu}^{\mathcal{T}}(\mathbf{G}\,B^c\wedge\mathbf{G}\,\mathbf{F}\,A)\le\mathsf{Prob}_{\mu}^{\mathcal{T}_B}(\mathbf{G}\,\mathbf{F}\,A)=0.$$

After martingales

In 2020, we received a comment from a reviewer "*I think this is a trivial application of martingale theory*"... AND IT WAS!

Proof. In order not to obfuscate the interesting ideas of the proof with technical considerations, we first prove the lemma for $n = 0$ (with $\mathcal{A} = A \in \Sigma$), and explain afterwards how to extend the proof to obtain the general statement. We want to prove that for all $\mu \in \mathsf{Dist}(S)$,

$$\mathsf{Prob}_\mu^{\mathcal{T}}(\mathbf{G}\,B^c \wedge \mathbf{G}\mathbf{F}\,A) = 0.$$

Let $\mu \in \mathsf{Dist}(S)$ be an initial distribution. We assume w.l.o.g. that $A \cap B = \emptyset$—indeed, if that is not the case, we simply notice that $\mathsf{Prob}_\mu^{\mathcal{T}}(\mathbf{G}\,B^c \wedge \mathbf{G}\mathbf{F}\,A) = \mathsf{Prob}_\mu^{\mathcal{T}}(\mathbf{G}\,B^c \wedge \mathbf{G}\mathbf{F}(A \cap B^c))$ and we replace A by $A \cap B^c$ in the rest of the proof.

Let us consider a modified STS $\mathcal{T}_B$ which is equal to $\mathcal{T}$, except that B is made absorbing (we assume that for $s \in B$, $\kappa(s, \cdot)$ is the Dirac distribution δ_s). Notice that $\mathsf{Prob}_\mu^{\mathcal{T}}(\mathbf{F}\,B) = \mathsf{Prob}_\mu^{\mathcal{T}_B}(\mathbf{F}\mathbf{G}\,B)$, and $\mathsf{Prob}_\mu^{\mathcal{T}}(\mathbf{G}\,B^c \wedge \mathbf{G}\mathbf{F}\,A) \le \mathsf{Prob}_\mu^{\mathcal{T}_B}(\mathbf{G}\mathbf{F}\,A)$ (as $A \cap B = \emptyset$, runs that see A infinitely often without seeing B in $\mathcal{T}$ are just as likely in $\mathcal{T}_B$). Notice also that the event $\mathbf{F}\mathbf{G}\,B$ is shift-invariant. We have

$$\begin{aligned}
&\mathsf{Ev}_{\mathcal{T}_B}(\mathbf{G}\mathbf{F}\,A)\\
&= \{\rho = s_0 s_1 \ldots \in S^\omega \mid \forall i, \exists j \ge i, s_j \in A\}\\
&\subseteq \{\rho = s_0 s_1 \ldots \in S^\omega \mid \forall i, \exists j \ge i, \mathsf{Prob}_{\delta_{s_j}}^{\mathcal{T}_B}(\mathbf{F}\,B) \ge p\} && \text{by hypothesis on } A\\
&= \{\rho = s_0 s_1 \ldots \in S^\omega \mid \forall i, \exists j \ge i, \mathsf{Prob}_{\delta_{s_j}}^{\mathcal{T}_B}(\mathbf{F}\mathbf{G}\,B) \ge p\} && \text{by construction of } \mathcal{T}_B\\
&= \{\rho \in S^\omega \mid \forall i, \exists j \ge i, \mathsf{E}_\mu^{\mathcal{T}_B}[\mathbf{1}_{\mathbf{F}\mathbf{G}\,B} \mid \mathcal{F}_{j+1}](\rho) \ge p\} && \text{by Lemma 17, as } \mathbf{F}\mathbf{G}\,B \text{ is shift-invariant}\\
&\subseteq \{\rho \in S^\omega \mid \lim_{i\to\infty} \mathsf{E}_\mu^{\mathcal{T}_B}[\mathbf{1}_{\mathbf{F}\mathbf{G}\,B} \mid \mathcal{F}_i](\rho) \text{ is not 0 if it exists}\}\\
&= \{\rho \in S^\omega \mid \mathbf{1}_{\mathbf{F}\mathbf{G}\,B}(\rho) \neq 0\} && \text{by Lévy's zero-one law (Proposition 16)}\\
&= \{\rho \in S^\omega \mid \mathbf{1}_{\mathbf{F}\mathbf{G}\,B}(\rho) = 1\}\\
&= \mathsf{Ev}_{\mathcal{T}_B}(\mathbf{F}\mathbf{G}\,B).
\end{aligned}$$

All inclusions and equalities are almost sure. In $\mathcal{T}_B$, as $A \cap B = \emptyset$ and B is absorbing, we have that $\mathsf{Prob}_\mu^{\mathcal{T}_B}(\mathbf{G}\mathbf{F}\,A \wedge \mathbf{F}\mathbf{G}\,B) = 0$. As $\mathsf{Ev}_{\mathcal{T}_B}(\mathbf{G}\mathbf{F}\,A) \subseteq \mathsf{Ev}_{\mathcal{T}_B}(\mathbf{F}\mathbf{G}\,B)$, this implies that $\mathsf{Prob}_\mu^{\mathcal{T}_B}(\mathbf{G}\mathbf{F}\,A) = 0$.

We conclude

$$\mathsf{Prob}_\mu^{\mathcal{T}}(\mathbf{G}\,B^c \wedge \mathbf{G}\mathbf{F}\,A) \le \mathsf{Prob}_\mu^{\mathcal{T}_B}(\mathbf{G}\mathbf{F}\,A) = 0.$$

Rest of the talk: **Definition of martingales, key theorems, and two applications to verification.**

Conditional Expectation

Conditional Expectation w.r.t. a σ-algebra (1/2)

Let $(\Omega, \mathcal{B}, \mathbb{P})$ be a probability space, $X \colon \Omega \to \mathbb{R}$ a random variable, and $\mathcal{F} \subseteq \mathcal{B}$ a sub-σ-algebra.

- The definition of martingales requires the notion of **conditional expectation w.r.t. a σ-algebra** (not just w.r.t. an event). It is a **function** $\mathbb{E}[X \mid \mathcal{F}] \colon \Omega \to \mathbb{R}$.
- **Hard definition**: Non-constructive in the general continuous case, requires a hard proof (*Radon–Nikodym theorem*) just to show it exists and is unique.

Conditional Expectation w.r.t. a σ-algebra (1/2)

Let $(\Omega, \mathcal{B}, \mathbb{P})$ be a probability space, $X\colon \Omega \to \mathbb{R}$ a random variable, and $\mathcal{F} \subseteq \mathcal{B}$ a sub-σ-algebra.

- The definition of martingales requires the notion of **conditional expectation w.r.t. a σ-algebra** (not just w.r.t. an event). It is a **function** $\mathbb{E}[X \mid \mathcal{F}]\colon \Omega \to \mathbb{R}$.
- **Hard definition**: Non-constructive in the general continuous case, requires a hard proof (*Radon–Nikodym theorem*) just to show it exists and is unique.
- Easier argument (according to Matthieu): see it as a projection in L^2 space.
  - Still hard for the average computer scientist/MDP enjoyer.

Conditional Expectation w.r.t. a σ-algebra (2/2)

- In our case, we mainly need the definition **for a finite σ-algebra** $\mathcal{F}$ (and thus generated by a finite partition into "atoms" $\{B_i\}$). Easy definition: for $\omega \in \Omega$,

$$\mathbb{E}[X \mid \mathcal{F}](\omega) = \frac{1}{\mathbb{P}(B)} \int_B X \, \mathrm{d}\mathbb{P}$$

where B is the unique element of the partition such that $\omega \in B$.

Conditional Expectation w.r.t. a σ-algebra (2/2)

- In our case, we mainly need the definition **for a finite σ-algebra** $\mathcal{F}$ (and thus generated by a finite partition into "atoms" $\{B_i\}$). Easy definition: for $\omega \in \Omega$,

$$\mathbb{E}[X \mid \mathcal{F}](\omega) = \frac{1}{\mathbb{P}(B)} \int_B X \,\mathrm{d}\mathbb{P}$$

where B is the unique element of the partition such that $\omega \in B$.

Important reminder: $\mathbb{E}[X \mid \mathcal{F}]$ is a **random variable** $\Omega \to \mathbb{R}$, not a real number!

Conditional Expectation w.r.t. a σ-algebra (2/2)

- In our case, we mainly need the definition **for a finite σ-algebra** $\mathcal{F}$ (and thus generated by a finite partition into "atoms" $\{B_i\}$). Easy definition: for $\omega \in \Omega$,

$$\mathbb{E}[X \mid \mathcal{F}](\omega) = \frac{1}{\mathbb{P}(B)} \int_B X \, \mathrm{d}\mathbb{P}$$

where B is the unique element of the partition such that $\omega \in B$.

Important reminder: $\mathbb{E}[X \mid \mathcal{F}]$ is a **random variable** $\Omega \to \mathbb{R}$, not a real number!

Information-theoretic intuition

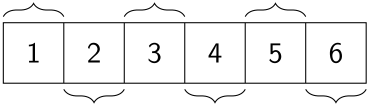
$\mathbb{E}[X \mid \mathcal{F}]$ is the most that we can know about X given information that we can glean from observing $\mathcal{F}$. It is finer than just $\mathbb{E}[X]$ (no information), but coarser than X (full information).

Example: The Die

- $\Omega = \{1, \ldots, 6\}$
- $X(\omega) = \omega$ (The value)
- P = fair die
- $\mathcal{F} = \{\emptyset, \{1,3,5\}, \{2,4,6\}, \Omega\}$ (Information: odd or even)

$$\mathbb{E}[X \mid \mathcal{F}](\omega) = \begin{cases} \frac{1+3+5}{3} = \mathbf{3} & \text{if } \omega \in \{1,3,5\} \\ \frac{2+4+6}{3} = \mathbf{4} & \text{if } \omega \in \{2,4,6\} \end{cases}$$

Odd $\to$ Expectation 3



Even $\to$ Expectation 4

Properties of the conditional expectation

1. If X is $\mathcal{F}$-measurable (i.e., observing $\mathcal{F}$ gives you everything there is to know about X):

$$\mathbb{E}[X \mid \mathcal{F}] = X.$$

2. If $\mathcal{F} = \{\emptyset, \Omega\}$ (no information at all):

$$\mathbb{E}[X \mid \mathcal{F}] = \frac{1}{\mathbb{P}(\Omega)} \int_\Omega X \,\mathrm{d}\mathbb{P} = \int_\Omega X \,\mathrm{d}\mathbb{P} = \mathbb{E}[X] \quad \text{(constant)}.$$

3. If $\mathcal{F}_1 \subseteq \mathcal{F}_2$:

$$\mathbb{E}[\,\mathbb{E}[X \mid \mathcal{F}_2] \mid \mathcal{F}_1\,] = \mathbb{E}[X \mid \mathcal{F}_1].$$

Projecting a projection returns the coarser projection.

Markov Chain Example (1/2)

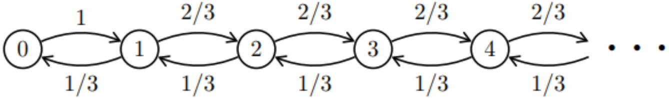
Let $\mathcal{M} = (S, P)$ be a Markov chain (possibly infinite).

- $\Omega = S^\omega$ (infinite paths).
- We define a family of σ-algebras: for $n \in \mathbb{N}$, let

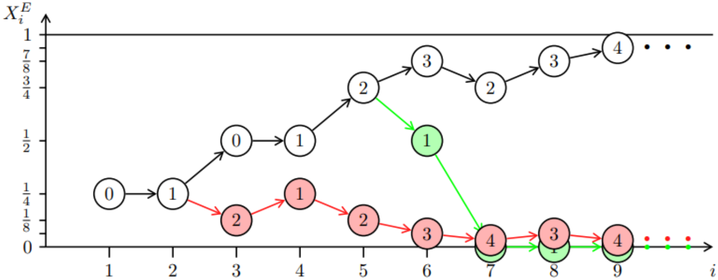
$$\begin{aligned}\mathcal{F}_n &= \text{“exactly the information about the first } n \text{ steps”} \\ &= \sigma\left(\bigcup_{h \in S^n} \mathsf{Cyl}(h)\right).\end{aligned}$$

Markov Chain Example (2/2)[1]

Consider this infinite Markov chain:



Let E be the event "exactly two visits to state 0". Consider the values $X_i^E(\rho) = \mathbb{E}[\mathbb{1}_E \mid \mathcal{F}_i](\rho)$ for a few runs ρ.



[1]From *Kiefer, Mayr, Shirmohammadi, Totzke, Wojtczak: How to Play in Infinite MDPs. ICALP'20.*

Martingales

Definitions

- A (discrete-time) **stochastic process** is a sequence $(X_n)_{n \in \mathbb{N}}$ of random variables.
- A **filtration** is an infinite sequence $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{B}$ of σ-algebras.
- $(X_n)_n$ is **adapted** to $(\mathcal{F}_n)_n$ if for all n, X_n is $\mathcal{F}_n$-measurable.

Definitions

- A (discrete-time) **stochastic process** is a sequence $(X_n)_{n\in\mathbb{N}}$ of random variables.
- A **filtration** is an infinite sequence $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{B}$ of σ-algebras.
- $(X_n)_n$ is **adapted** to $(\mathcal{F}_n)_n$ if for all n, X_n is $\mathcal{F}_n$-measurable.

Definition

The sequence X_n is a **martingale** if:

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n.$$

Intuition: Think of a fair sequential game such that the average value at step $n+1$, when you know the first n steps, is your gain after n steps.

Martingale Example: Betting

Let $Y_1, Y_2, \ldots$ be independent bets that win either $+1$ or -1 with probability $\frac{1}{2}$.
Let $X_n = Y_1 + \cdots + Y_n$ (your money after n bets).

Martingale Example: Betting

Let $Y_1, Y_2, \ldots$ be independent bets that win either $+1$ or -1 with probability $\frac{1}{2}$.
Let $X_n = Y_1 + \cdots + Y_n$ (your money after n bets).

Drawing: **blackboard**.

Martingale Example: Betting

Let $Y_1, Y_2, \dots$ be independent bets that win either $+1$ or -1 with probability $\frac{1}{2}$.
Let $X_n = Y_1 + \cdots + Y_n$ (your money after n bets).

Drawing: **blackboard**.

Proof that $(X_n)_n$ is a martingale:

$$\begin{aligned}
\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] &= \mathbb{E}[X_n + Y_{n+1} \mid \mathcal{F}_n] \\
&= \mathbb{E}[X_n \mid \mathcal{F}_n] + \mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] \quad \text{(linearity of expectation)} \\
&= X_n + \mathbb{E}[Y_{n+1}] \quad (X_n \text{ is } \mathcal{F}_n\text{-measurable, } \mathcal{F}_n \text{ is independent from } Y_{n+1}) \\
&= X_n + 0 \\
&= X_n.
\end{aligned}$$

The "Usual" Martingale for Markov Chains

All the uses I have seen of martingales in verification have the following form.
Take a reasonable random variable X about *infinite* runs (e.g., $X = \mathbb{1}_{Büchi(\top)}$).

Doob Martingale

Take $X_n = \mathbb{E}[X \mid \mathcal{F}_n]$.

Lemma. $(X_n)_n$ is a martingale.

The “Usual” Martingale for Markov Chains

All the uses I have seen of martingales in verification have the following form.
Take a reasonable random variable X about *infinite* runs (e.g., $X = \mathbb{1}_{Büchi(\top)}$).

Doob Martingale

Take $X_n = \mathbb{E}[X \mid \mathcal{F}_n]$.

Lemma. $(X_n)_n$ is a martingale.

Proof:

$$\begin{aligned}\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] &= \mathbb{E}[\underbrace{\mathbb{E}[X \mid \mathcal{F}_{n+1}]}_{X_{n+1}} \mid \mathcal{F}_n] \\ &= \mathbb{E}[X \mid \mathcal{F}_n] \\ &= X_n.\end{aligned}$$

Theorems to Know

First Key Theorem: Doob's Convergence Theorem

Doob's Convergence Theorem

If $(X_n)_n$ is a bounded martingale, then there is a random variable X_∞ such that $X_n \to X_\infty$ almost surely.

I.e., for almost all "runs" ρ, $X_n(\rho)$ converges to $X_\infty(\rho)$ as $n \to \infty$.

Second Key Theorem: Lévy's 0-1 Law

Take the Doob martingale $X_n = \mathbb{E}[X \mid \mathcal{F}_n]$. By Doob's: $X_n \to X_\infty$ a.s.
It can be shown that $X_\infty = \mathbb{E}[X \mid \mathcal{F}_\infty]$ where $\mathcal{F}_\infty = \sigma(\bigcup_n \mathcal{F}_n)$.

Lévy's 0-1 Law

If we take $X = \mathbb{1}_A$ for an event $A \in \mathcal{F}_\infty$:

$$\mathbb{E}[\mathbb{1}_A \mid \mathcal{F}_n] \to \mathbb{E}[\mathbb{1}_A \mid \mathcal{F}_\infty] = \mathbb{1}_A.$$

Second Key Theorem: Lévy's 0-1 Law

Take the Doob martingale $X_n = \mathbb{E}[X \mid \mathcal{F}_n]$. By Doob's: $X_n \to X_\infty$ a.s.
It can be shown that $X_\infty = \mathbb{E}[X \mid \mathcal{F}_\infty]$ where $\mathcal{F}_\infty = \sigma(\bigcup_n \mathcal{F}_n)$.

Lévy's 0-1 Law

If we take $X = \mathbb{1}_A$ for an event $A \in \mathcal{F}_\infty$:

$$\mathbb{E}[\mathbb{1}_A \mid \mathcal{F}_n] \to \mathbb{E}[\mathbb{1}_A \mid \mathcal{F}_\infty] = \mathbb{1}_A.$$

Consequences for Markov Chains

- $\mathbb{P}(\{\rho \mid \lim X_n(\rho) \in \{0,1\}\}) = 1.$
- **Almost all runs converge to 0 or 1** as you observe them!
- Moreover, $\lim X_n(\rho) = \mathbb{1}_A$: it converges to 1 if the run ρ is in A, to 0 otherwise!
- All runs "show" at the limit if they are in A or not!

Two Applications

App #1: Back to Motivating Problem

Reminder: Let $\mathcal{M} = (S, s_0, P)$ be an (infinite) Markov chain and $\top \in S$. For $p > 0$, define

$$\mathrm{Reach}_{\geq p} = \{s \in S \mid \mathbb{P}_s(\Diamond \top) \geq p\}.$$

Theorem

$$\mathbb{P}_{s_0}(\Diamond \top \mid \Box\Diamond \mathrm{Reach}_{\geq p}) = 1$$

App #1: Back to Motivating Problem

Reminder: Let $\mathcal{M} = (S, s_0, P)$ be an (infinite) Markov chain and $\top \in S$. For $p > 0$, define

$$\text{Reach}_{\geq p} = \{s \in S \mid \mathbb{P}_s(\Diamond \top) \geq p\}.$$

Theorem

$$\mathbb{P}_{s_0}(\Diamond \top \mid \Box\Diamond \text{Reach}_{\geq p}) = 1$$

Proof:

1. Let $X = \mathbb{1}_{\Diamond \top}$ and $X_n = \mathbb{E}[\mathbb{1}_{\Diamond \top} \mid \mathcal{F}_n]$.
2. If we visit $\text{Reach}_{\geq p}$ infinitely often, then for infinitely many n's, $X_n(\rho) \geq p > 0$.
3. But $X_n(\rho) \to 0$ or 1 (by **Lévy's 0-1 Law**, using that $\Diamond \top \in \mathcal{F}_\infty$).
4. It does not converge to 0 (infinitely often $\geq p$).
5. So it converges to 1.
6. So runs are almost surely in $\Diamond \top$.

App #2: Hypothesis Testing: Tiger POMDP

Tiger POMDP: **blackboard**.

- Assume $\mathcal{F}_n$ is the information after n listens (observations).
- Let X_n^L be the probability to be in L after n listens: $X_n^L = \mathbb{E}[\mathbb{1}_L \mid \mathcal{F}_n]$.
- It is a martingale, so by Doob: X_n^L converges.
- However, $\mathbb{1}_L$ is not $\mathcal{F}_\infty$-measurable (we are never completely sure about the tiger's position). So no Lévy's 0-1 Law directly...

Proof of Convergence

Claim: It still converges to 0 or 1 at the limit! Doob's tells us $X_n^L(\rho)$ converges a.s.
Assume $X_n^L(\rho) \to x \notin \{0, 1\}$.
Then the ratio converges to a constant:

$$\frac{X_n^L(\rho)}{X_n^R(\rho)} \to \frac{x}{1-x}.$$

But, by Bayes rule:

$$\frac{X_{n+1}^L(\rho)}{X_{n+1}^R(\rho)} = \frac{X_n^L(\rho)}{X_n^R(\rho)} \cdot \underbrace{\frac{P(o_{n+1} \mid L)}{P(o_{n+1} \mid R)}}_{\text{Observation Ratio}} .$$

If observations are *distinguishing*, this factor makes **too big of a jump** to converge to anything but 0 or $+\infty$!
So $X_n^L(\rho) \to \{0, 1\}$ a.s.: at the limit, we are almost surely sure about the tiger's position!