Probability Theory

Daniele Zago

November 25, 2021

Contents

Lecture 6: Martingales and Markov processes

2021-11-18

We introduce two important classes of stochastic processes which can be extended to the continuous time case.

6.1 Martingales

References Bass (2011, §3)

Martingales were well-known stochastic processes in economics which over the last decades became crucial in the theory of stochastic integration, from which we can construct continuous Markov processes (*diffusions*).

Def. (Discrete-time martingale)
A discrete stochastic process $X = (X_n)_n$ is called a ***martingale*** w.r. to a given filtration $(\mathcal{F}_n)_n$ if

i. $X_n \in L^1(\Omega, \mathbb{P})$ for all n.

ii. $\mathbb{E}[X_N|\mathcal{F}_n] = X_n$ for all $n \leq N$.

Adaptability There is no need to specify that X has to be adapted to $(\mathcal{F}_n)_n$, since $X_n = \mathbb{E}[X_N|\mathcal{F}_n]$ implies measurability w.r. to $\mathcal{F}_n$.

Expected value The second equality is a very strong property which tells us that if we condition the future process on the information at time n, then the expected value is equal to the value that we have observed. Using the tower property, we have that $\mathbb{E}[X_N] = \mathbb{E}[\mathbb{E}[X_N|\mathcal{F}_n]] = \mathbb{E}[X_n]$, therefore the expectation is *a priori* constant in time.

Example (Just $\mathbb{E}(X) = \mu$ is not enough)
Let $(X_n)_n$ be a family of independent random variables with $\mathbb{E}[X_n] = \mu$ for all n, and consider the natural filtration $(\mathcal{F}_n^X)_n$. The process $X = (X_n)_n$ is not a martingale for all possible distributions of X_n, since

$$\mathbb{E}[X_N|\mathcal{F}_n^X] \overset{\perp\!\!\!\perp}{=} \mathbb{E}[X_N] = \mu.$$

Therefore, this process is a martingale $\iff$ $\mathbb{E}[X_N] = \mu = X_n$ for all $n \leq N$, which is satisfied $\iff$ $X_n \equiv \mu$ almost surely.

Remark From the example above, independence is *orthogonal* to martingality, unless we choose a degenerate distribution $X_n \equiv \mu$.

Example (Martingale from independent variables)
Let us consider the process defined in the previous example, and define the stochastic process $Y_n = \sum_{k=1}^n X_k$. Clearly, Y_{n+1} and Y_n are marginally not independent, therefore the process

could be a martingale. Indeed, we have that

$$\mathbb{E}[Y_{n+1}|\mathcal{F}_n^X] = \mathbb{E}[Y_n + X_{n+1}|\mathcal{F}_n^X] = \underbrace{\mathbb{E}[Y_n|\mathcal{F}_n^X]}_{=Y_n} + \underbrace{\mathbb{E}[X_{n+1}|\mathcal{F}_n^X]}_{=\mathbb{E}[X_{n+1}]} = Y_n + \mu.$$

Therefore, we have that Y_n is a martingale $\iff \mu = 0$.

What can we say now about a martingale which is not defined w.r. to the filtration $\mathcal{F}_n^X$ but to a different filtration? For instance, what happens to the martingale property when enlarging to a bigger filtration?

Example (Adding events breaks martingality)
Let $X = (X_n)_n$ be a martingale w.r. to a filtration $(\mathcal{F}_n)_n$, and consider now a new filtration equal to all possible events $\mathcal{F}$ at all times, $(\mathcal{G}_n)_n = \mathcal{F}$. We now have that X is a martingale w.r. to $\mathcal{G}_n$ if

$$X_n = \mathbb{E}[X_N|\mathcal{G}_n] = \mathbb{E}[X_N|\mathcal{F}] = X_N,$$

therefore this means that X can again only be a constant process $X_n = \mu$ for all n.

In general When adding events we can't immediately conclude that the process is still a martingale.

Prop. 1 (Removing events does not break martingality)
Let $(X_n)_n$ be a martingale w.r. to a filtration $(\mathcal{F}_n)_n$. Let now $(\mathcal{G}_n)_n$ be another filtration such that

a) X is adapted to $(\mathcal{G}_n)_n$.

b) $\mathcal{G}_n \subset \mathcal{F}_n$ is a sub-filtration at all times.

Then, X is a martingale w.r. to $(\mathcal{G}_n)_n$.

Proof.
We use the tower property to prove the result, indeed since $\mathcal{G}_n \subset \mathcal{F}_n$ we can write

$$\begin{aligned}\mathbb{E}[X_N|\mathcal{G}_n] &\overset{(b)}{=} \mathbb{E}[\overbrace{\mathbb{E}[X_N|\mathcal{F}_n]}^{=X_n}|\mathcal{G}_n] \\ &\overset{(a)}{=} X_n\end{aligned}$$

$\square$

Corollary 1
If X is a martingale w.r. to any given filtration $(\mathcal{F}_n)_n$, then X is also a martingale w.r. to the natural filtration $(\mathcal{F}_n^X)_n$.

Proof.
Since $\sigma(X_n) \subseteq \mathcal{F}_n^X$ we can apply the tower property in order to show that

$$\mathbb{E}[X_N|X_n] = \mathbb{E}[\mathbb{E}[X_N|\mathcal{F}_n^X]|X_n] \overset{(a)}{=} \mathbb{E}[X_n|X_n] = X_n.$$

□

To sum up, the above properties show that if X is a martingale then for all $N \geq n$ we have that $\mathbb{E}[X_N|X_n] = X_n$.

Finally, we introduce two broader classes of stochastic processes whose intersection gives exactly the set of martingale processes.

Def. (Submartingale and supermartingale)
A process $X = (X_n)_n$ is called a ***submartingale*** (***supermartingale***) w.r. to a given filtration $(\mathcal{F}_n)_n$ if

i. $X_n \in L^1(\Omega, \mathbb{P})$ for all n.

ii. X is adapted to $(\mathcal{F}_n)_n$

iii. $X_n \overset{(\geq)}{\leq} \mathbb{E}[X_N|\mathcal{F}_n]$.

Expected value It's straightforward to check that, for a supermartingale (submartingale), the expected value is always increasing (decreasing), since

$$\mathbb{E}[X_N] = \mathbb{E}[\mathbb{E}[X_N|\mathcal{F}_n]] \underset{(\leq)}{\overset{\text{a.s.}}{\geq}} \mathbb{E}[X_n].$$

6.2 Stopping times

We now introduce a class of events which is extremely relevant to the analysis of stochastic process. Broadly speaking, this class of events is comprised by all events such that at time n we can tell whether they have occurred or not.

Def. (Stopping time)
Let $(\mathcal{F}_n)_n$ be a filtration. We say that a random variable $\tau : \Omega \longrightarrow [0, +\infty]$ is a ***stopping time*** if the event $\{\tau \leq n\}$ is such that

$$\{\tau \leq n\} \in \mathcal{F}_n \quad \text{for all } n.$$

Observability This is an observability condition for the random variable τ, i.e. at time n we must be able to tell whether the above event occurred or not based on the available information $\mathcal{F}_n$.

Remark Let τ be a stopping time and consider the event $\{\tau > n\}$. Then, the following events are also observable

$$\{\tau > n\} = \{\tau \leq n\}^c \in \mathcal{F}_n$$

$$\{\tau = n\} = \{\tau \leq n\} \setminus \underbrace{\{\tau \leq n-1\}}_{\in \mathcal{F}_{n-1} \subset \mathcal{F}_n} \in \mathcal{F}_n.$$

Example (Exit – or hitting – time)

Let X be a discrete-time stochastic process and consider a Borel set H. Let now I_H be the set of times at which X exits from H, i.e.

$$I_H := \{n : X_n \notin H\}.$$

Let now τ be the random variable that describes the time of first exit,

$$\tau := \begin{cases} \inf I_H & \text{if } I_n \neq \emptyset \\ +\infty & \text{if } I_H = \emptyset \end{cases}$$

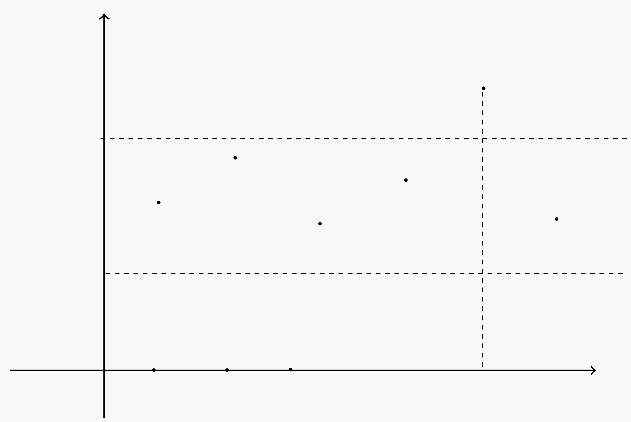

Figure 1: Example of a hitting time for a given set H.

This random variable is as a stopping time, since the event $\{\tau \leq n\}$ can be written as

$$\{\tau \leq n\} = \bigcup_{i \leq n} \underbrace{\{X_i \notin H\}}_{\mathcal{F}_i \subset \mathcal{F}_n} \in \mathcal{F}_n.$$

Continuous-time The previous example shows why this definition of a stopping time becomes problematic for continuous-time stochastic processes, due to the fact that a countable union of events is not guaranteed to belong to the σ-algebra $\mathcal{F}_n$.

6.3 Markov processes

Def. (Markov property)
A discrete-time stochastic process $X = (X_n)_n$ defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n)_n, \mathbb{P})$ has the ***Markov property*** if it is adapted and the following property holds true for any n:

$$\mathbb{E}[\varphi(X_{n+1})|\mathcal{F}_n] = \mathbb{E}[\varphi(X_{n+1})|X_n], \tag{M}$$

for any φ $\mathcal{B}$-measurable and bounded.

Interpretation Expectation of future values conditional to all cumulated information is equal to the expectation given the value of the process at time n.

Regression function If X has the Markov property, then we can find a function g_n such that

$$\mathbb{E}[\varphi(X_{n+1})|\mathcal{F}_n] = g_n(X_n),$$

where $g_n(x) = \mathbb{E}[\varphi(X_{n+1})|X_n = x]$ is the ***regression function*** (see Doob's **??**).

In practice Assume now that $\mathbb{E}[\varphi(X_{n+1})|\mathcal{F}_n] = f_n(X_n)$ is a deterministic function of X_n, then by the tower property of $\mathbb{E}$ we can write

$$\mathbb{E}[\varphi(X_{n+1})|X_n] = \mathbb{E}[\mathbb{E}[\varphi(X_{n+1})|\mathcal{F}_n]|X_n] = \mathbb{E}[f_n(X_n)|X_n] = f_n(X_n).$$

Therefore, if we can find that the expectation of X_{n+1} is a deterministic function of X_n, we can conclude that X has the Markov property.

Example (Independent r.v.'s form a Markov proces)
Let $X = (X_n)_n$ be a sequence of independent r.v.'s, then $\mathcal{F}_n = \mathcal{F}_n^X$ and

$$\mathbb{E}[\varphi(X_{n+1})|\mathcal{F}_n^X] \stackrel{\perp}{=} \mathbb{E}[\varphi(X_{n+1})].$$

Lemma 1 (Freezing)
If X, Y are random variables and $\mathcal{G}$ a σ-algebra such that Y and $\mathcal{G}$ are independent and X is $\mathcal{G}$-measurable, then we have that

$$\mathbb{E}[f(X,Y)|\mathcal{G}] = \mathbb{E}[f(x,Y)]\Big|_{x=X}$$

Proof.
No.

□

Interpretation Since Y is independent of the information, the randomness in X goes out of the conditioning operation.

Example (Cumulative sum is a Markov process)
Consider now the stochastic process $Y_n := \sum_{i=1}^n X_i$ for the process X defined in the previous example. Then, we have that

$$\begin{aligned}\mathbb{E}[\varphi(Y_{n+1})|\mathcal{F}_n^X] &\overset{\text{def}}{=} \mathbb{E}[\varphi(Y_n + X_{n+1})|\mathcal{F}_n^X] \\ &= \mathbb{E}[\varphi(y + X_{n+1})]\Big|_{y=Y_n} \qquad \text{(Freezing Lemma 1)},\end{aligned}$$

which is a deterministic function of X_{n+1} and therefore makes Y a Markov process.

Prop. 2 (Characterization of Markov's property)
The Markov property (M) for a process X is equivalent to satisfying, for any $A \in \mathcal{B}$,

$$\underbrace{\mathbb{E}[\mathbb{1}_{X_{n+1}\in A}|\mathcal{F}_n]}_{\mathbb{P}(X_{n+1}\in A|\mathcal{F}_n)} = \underbrace{\mathbb{E}[\mathbb{1}_{X_{n+1}\in A}|X_n]}_{\mathbb{P}(X_{n+1}\in A|X_n)} \tag{M'}$$

Proof.
$(M) \implies (M')$: Since $\mathbb{1}_A$ is a bounded and $\mathcal{B}$-measurable function, it is valid by choosing $\varphi = \mathbb{1}_A$.

$(M') \implies (M)$: Let $(\varphi_k)_k$ be a sequence of simple functions of the type $\varphi_k = \sum_{j=1}^m c_{j,k}\mathbb{1}_{A_{j,k}}$, which are bounded and $\mathcal{B}$-measurable, and such that

$$\varphi_k \xrightarrow{k\to\infty} \varphi.$$

See for instance here for the standard construction of such a sequence of simple functions $(\varphi_k)_k$ when approximating a bounded function φ. With this approximation, we can chain the following equations:

$$\begin{aligned}\mathbb{E}[\varphi(X_{n+1})|\mathcal{F}_n] &\overset{\text{DCT}}{=} \lim_{k\to\infty} \mathbb{E}[\varphi_k(X_{n+1})|\mathcal{F}_n] \\ &= \lim_{k\to\infty}\sum_{j=1}^m c_{j,k}\mathbb{E}[\mathbb{1}_{A_{j,k}}(X_{n+1})|\mathcal{F}_n] \\ &= \lim_{k\to\infty}\sum_{j=1}^m c_{j,k}\mathbb{P}(X_{m+1} \in A_{j,k}|\mathcal{F}_n) \\ &= \lim_{k\to\infty}\sum_{j=1}^m c_{j,k}\mathbb{P}(X_{m+1} \in A_{j,k}|X_n) \qquad (M') \\ &= \text{(Do the steps backwards)} \\ &= \mathbb{E}[\varphi(X_{n+1})|X_n].\end{aligned}$$

□

Example (Enlarging the filtration breaks Markov)
Let $\mathcal{G}_n = \mathcal{F}$ be the maximal filtration for all $n \in \mathbb{N}$, then for a discrete Markov process X we have

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

On the other hand, when we reduce the filtration we have a preservation result analogous to what we have seen with martingales (Prop. 1).

Prop. 3 (Reducing the filtration preserves Markov)
If X has (M) and $(\mathcal{G}_n)_n$ is a filtration such that

a) X is adapted to $(\mathcal{G}_n)_n$

b) $\mathcal{G}_n \subset \mathcal{F}_n$,

then X has (M) w.r. to $(\mathcal{G}_n)_n$.

Proof.
Similarly to Prop. 1, use the tower property of the conditional expected value.

□

Prop. 4 (Equivalent definition of Markov's property)
Property (M) for a process X is equivalent to satisfying, for each $N > n$,

$$\mathbb{P}(X_N \in A|\mathcal{F}_n) = \mathbb{P}(X_N \in A|X_n) \qquad (M'')$$

Proof.
Homework.

□

Validity All the properties we have discussed until now are expressed in their general form and are valid for any type of discrete-time stochastic process, i.e. whether each random variable X_n is characterized either by a continuous or discrete distribution. What we discuss below is a specialization of the properties in the case when X is a discrete-time process for which X_n takes discrete values.

6.4 Markov chains

References Brémaud (2020)

Def. (Discrete-time process)
A discrete-time process X is called a ***discrete process*** if X_n takes values on a countable state space E.

Example Some examples are $E = \mathbb{N}, \mathbb{N}^2, \mathbb{Z}, \mathbb{Z}^2, \dots$

Notation Following the notation of Brémaud (2020), we use i, j, k, h, l and $i_0, i_1, i_n, \dots$ to denote the elements of the countable space E.

Def. (Markov chain)
A discrete process X is called a ***Markov chain*** if it has the Markov property w.r. to the natural filtration $(\mathcal{F}_n^X)_n$.

Prop. 5 (Equivalent definition of Markov chain for discrete processes)
A discrete process X is a Markov chain if and only if for any n and for any $i_0, i_1, \dots, i_n, j \in E$

$$\mathbb{P}(X_{n+1} = j | X_n = i_n, \dots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j | X_n = i_n), \tag{$*$}$$

whenever this probability is valid, i.e. $\mathbb{P}(X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) > 0$.

Proof.
No.

□

Problem This definition works only for processes which are in discrete time and are defined on a countable state space. The more general definition (M) can be used instead for discrete-time continuous processes.

Def. (Homogeneous Markov chain)
We call a Markov chain X ***homogeneous*** (HMC) if the right-hand side of $(*)$ does not depend on n, i.e. if

$$\mathbb{P}(X_{n+1} = j | X_n = i_n, \dots, X_0 = i_0) = \mathbb{P}(X_1 = j | X_0 = i_n).$$

Example (HMC)
If X is a HMC then, for example

$$\mathbb{P}(X_3 = 4 | X_2 = 1, X_1 = 0, X_0 = -1) \overset{(*)}{=} \mathbb{P}(X_3 = 4 | X_2 = 1) \overset{\text{HMC}}{=} \mathbb{P}(X_1 = 4 | X_0 = 1).$$

A HMC is particularly important since we can define a transition matrix that describes the transition from one state to another regardless of the time.

Def. (Transition matrix of a HMC)
For a HMC X we define the ***transition matrix*** as the countable family of numbers

$$P = (p_{ij})_{i,j \in E}, \qquad p_{ij} = \mathbb{P}(X_1 = j | X_0 = i).$$

Properties of P For any $i \in E$, every row of P is a probability distribution and therefore P is a *stochastic matrix*, i.e.

$$\sum_{j \in E} p_{ij} = 1 \qquad \text{for all } i \in E.$$

Consider now the process of making two Markov chain transitions. In this case we have to use P two times in order to transition from $X_0 \to X_1$ and then from $X_1 \to X_2$. To compute these probabilities, we introduce a generalization of the matrix multiplication and addition operations in order to define the powers of an infinite-dimensional matrix $P^2, P^3, \text{etc} \ldots$

Algebraic operations Let $A = (a_{ij})_{i,j \in E}$ and $B = (b_{ij})_{i,j \in E}$ be two transition matrices, then we generalize the usual sum and product operations for standard matrices as

$$A + B = (a_{ij} + b_{ij})_{i,j \in E}$$
$$A \cdot B = \Big(\sum_{k \in E} a_{ik} b_{ik} \Big)_{i,j \in E}$$

Let now $\boldsymbol{x} = (x_i)_{i \in E}$ be a column vector, then

$$A\boldsymbol{x} = \Big(\sum_{k \in E} A_{ik} x_k \Big)_{i \in E}$$
$$\boldsymbol{x}^\top A = \Big(\sum_{k \in E} x_k A_{ki} \Big)_{i \in E}$$

Example (1D random walk)
Consider a r.v. X_0 with values in $E = \mathbb{Z}$. Let now $(Z_n)_{n \in \mathbb{N}}$ be i.i.d r.v.'s such that

$$Z_n \sim p\delta_1 + (1-p)\delta_{-1}, \quad p \in (0,1).$$

We set $X_{n+1} = X_n + Z_{n+1}$ and we consider the stochastic process $X = (X_n)_{n \in \mathbb{N}}$. We already know that X is a HMC, and this stochastic process increases by 1 with probability p and decreases by 1 with probability $1-p$. Therefore, its transition matrix is given by

$$p_{ij} = \begin{cases} p & \text{if } j = i+1 \\ 1-p & \text{if } j = i-1 \\ 0 & \text{otherwise} \end{cases}$$

Exercises

1. Proof of the proposition
2. (Brémaud, 2020, p. 88) ex. 2.1.1 - 2.1.6, 2.2.1

Lecture 7: Homogeneous Markov chains

2021-11-25

Initial distribution The distribution of X HMC only depends on

1. The initial law ν_0, $\nu_0(\{i\}) = \mathbb{P}(X_0 = i)$ for all $i \in E$.
2. The transition matrix P.

More precisely, for $i_0, i_1, \dots, i_n \in E$ we have that the probability of the path from i_0 through $i_1, \dots, i_n$ is equal to

$$\mathbb{P}(X_0 = i_0, \dots, X_n = i_n) = \nu_0(\{i_0\}) \cdot p_{i_0 i_1} \cdot \dots \cdot p_{i_{n-1} i_n}.$$

***n* steps ahead** Now we need to compute the conditional probability of transition for multiple time steps, $\mathbb{P}(X_n = i_n | X_0 = i_0)$. We consider the probability distribution at time n,

$$\begin{aligned}
\nu_n(\{j\}) &= \mathbb{P}(X_n = j) \\
&= \sum_{i \in E} \mathbb{P}(X_n = j, X_{n-1} = i) \\
&= \sum_{i \in E} p_{ij} \mathbb{P}(X_{n-1} = i) \\
&= \sum_{i \in E} \nu_{n-1}(\{i\}) \cdot p_{ij} \\
&= (\nu_{n-1} P)_j && (\nu \text{ row vector})
\end{aligned}$$

Therefore, $\nu_n = \nu_{n-1} \cdot P$ and if we repeat this process n times we obtain the following equation

$$\boxed{\nu_n = \nu_0 P^n}$$

Notation We denote by P^n_{ij} the element (i, j) of P^n.

Example

Let $E = \{1, 2, 3, 4\}$ and consider the initial distribution $\nu_0(\{j\}) = \frac{1}{4}$ with the transition graph

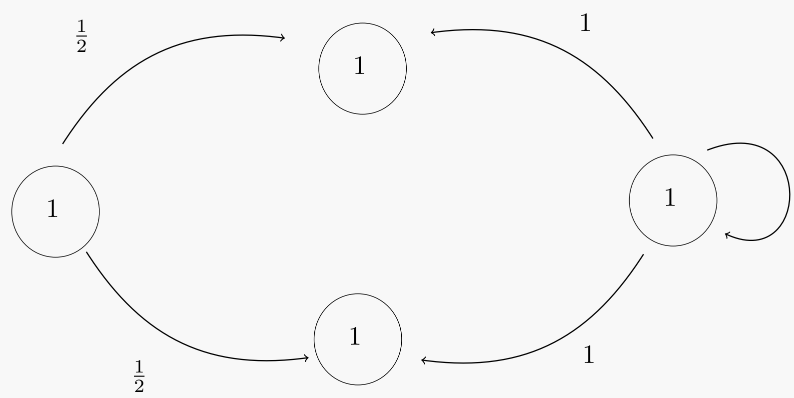

Figure 2: transitionGraph

For the above graph, $\mathbb{P}(X_3 = j|X_2 = 2)$ is <u>not defined</u>, whereas $\mathbb{P}(X_2 = j|X_1 = 2)$ is <u>well-defined</u>.

Remark Consider two sets

$$\text{FUTURE} \quad A = (X_{n+1} = j_1) \cap \ldots \cap (X_{n+k} = j_k)$$

$$\text{PAST} \quad B = (X_{n-1} = j_{n-1}) \cap \ldots \cap (X_0 = j_0)$$

From the Markov property (M) we can prove (long and boring proof) that

$$(M) \iff \mathbb{P}(A|X_n = i_n, B) = \mathbb{P}(A|X_n = i_n). \tag{1}$$

This is a bit stronger than the single value at time $n+1$, since we consider the whole trajectory from time $n+1$ to $n+k$. In (1) we can multiply by $\mathbb{P}(B|X_n = i_n)$ to get

$$\mathbb{P}(A|X_n = i_n, B)\frac{\mathbb{P}(B \cap \{X_n = i_n\})}{\mathbb{P}(\{X_n = i_n\})} = \frac{\mathbb{P}(A \cap B \cap \{X_n = i_n\})}{\cancel{\mathbb{P}(\{X_n = i_n\} \cap B)}} \frac{\cancel{\mathbb{P}(B \cap \{X_n = i_n\})}}{\mathbb{P}(\{X_n = i_n\})}$$

therefore

$$\boxed{\mathbb{P}(A \cap B|X_n = i_n) = \mathbb{P}(A|X_n = i_n)\mathbb{P}(\Big|X_n = i_n)}$$

From this we conclude that $A \perp\!\!\!\perp B$ conditional to $X_n = i_n$.

Thm. 1 (Canonical representation)
Let $(Z_n)_n$ be a sequence of i.i.d r.v.'s with values on a measurable space $(G, \mathcal{G})$. Let $f : E \times G \longrightarrow E$ be a measurable function w.r. to the product σ-algebra $\mathscr{P}(E) \otimes \mathcal{G}$ and $X_0 \perp\!\!\!\perp (Z_n)_n$ be an initial r.v. with values on E. Then, if we define

$$X_{n+1} := f(X_n, Z_{n+1}), \quad \text{for all } n \in \mathbb{N}_0,$$

we have that the process $X = (X_n)_{n \in \mathbb{N}_0}$ is a homogeneous Markov chain with transition matrix

$$P_{ij} = \mathbb{P}\big(f(i, Z_1) = j\big).$$

Non-i.i.d If the Z_n are not identically distributed, then we would have a non-homogeneous Markov chain with transition matrix $P_{ij}(n) = \mathbb{P}\big(f(i, Z_n) = j\big)$

Proof.

□

Example (Random walk)
Setting $G = \{-1, 1\}$ and $f(i, z) = i + z$ yields the 1D random walk

$$X_{n+1} = X_n + Z_{n+1}, \quad Z = p\delta_1 + (1-p)\delta_{-1}.$$

References

Bass, R. F. (2011). *Stochastic Processes*. Cambridge ; New York: Cambridge University Press.

Brémaud, P. (2020). *Markov Chains: Gibbs Fields, Monte Carlo Simulation and Queues*. Second. Springer Nature.