Markov Process

Kexing Ying

January 20, 2022

Contents

1 Introduction and Review

We will in this course assume the following notation:

- $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space;
- $\mathcal{X}$ is a Polish space, i.e. a separable, completely metrizable, topological space;
- $\mathcal{B}(\mathcal{X})$ is the Borel σ-algebra of $\mathcal{X}$.

Definition 1.1 (Stochastic Process)**.** A stochastic process $(x_n)_{n\in I}$ is a collection of random variables. In the case that $I = \mathbb{N}$ or $\mathbb{Z}$, we say that the stochastic process is discrete time. On the other hand if $I = \mathbb{R}_{\geq 0}$ or $[0,1] \subseteq \mathbb{R}$, then we say the process is continuous time.

We recall some definitions from elementary probability theory.

Definition 1.2 (Random Variable)**.** A random variable $x : \Omega \to \mathcal{X}$ is simply a measurable function.

Definition 1.3 (Probability Distribution)**.** Given a random variable $x : \Omega \to \mathcal{X}$, the probability distribution of x, denoted by $\mathcal{L}(x)$ is the push-forward measure of $\mathbb{P}$ along x, i.e.

$$\mathcal{L}(x) = x_*\mathbb{P} : A \in \mathcal{F} \mapsto \mathbb{P}(x^{-1}(A)).$$

Proposition 1.1. Let $x : \Omega \to \mathcal{X}$ be a random variable where $\mathcal{X}$ is countable, then

$$\mathcal{L}(x) = \sum_{i\in X} \mathbb{P}(x = i)\delta_i := \sum_{i\in X} x_*\mathbb{P}(\{i\})\delta_i$$

where δ_i is the Dirac measure concentrated at i.

Proof. Let $A \subseteq X$, then

$$\mathcal{L}(x)(A) = \sum_{i\in A} \mathcal{L}(x)(\{i\}) = \sum_{i\in X} \mathcal{L}(x)(\{i\})\delta_i(A) = \sum_{i\in X} x_*\mathbb{P}(\{i\})\delta_i(A),$$

as required. $\square$

Definition 1.4 (Independence)**.** Given random variables $x_1, \cdots, x_n$, we say $x_1, \cdots, x_n$ are independent if

$$\mathcal{L}((x_1, \cdots, x_n)) = \bigotimes_{i=1}^{n} \mathcal{L}(x_i),$$

where $\otimes$ denotes the product measure.

As the name suggests, we will in this course mostly focus on a class of stochastic processes known as Markov processes. These are processes in which given information about the process at the present time, its future is independent from its history. In particular, if (x_n) is a Markov process, given its value at x_k, the value of x_j is independent of the values of x_i for all $i < k < j$.

Definition 1.5 (Invariant Probability Measure)**.** A probability measure π is said to be an invariant probability measure or an invariant distribution of a Markov process $(x_n)_{n\in I}$ if for all $n \in I$, we have $\pi = \mathcal{L}(x_n)$.

A Markov chain started from an invariant distribution does is called a stationary Markov process as its distribution do not evolve and we say that the chain is in equilibrium.

In this course we will study the behaviour of the distribution of Markov processes. In particular, we ask

- does there exists an invariant measure? If so, is it unique?
- how does the distribution evolve over time?
- does $\mathcal{L}(x_n)$ converge as $n \to \infty$ (convergence in distribution)?

2 Markov Property

Let us now consider the Markov property in a more formal context.

2.1 Filtration and Simple Markov Property

In formation and filtration is an important notion not only for Markov processes but for stochastic processes in general.

In formally, the information of a random variable x is the collection of all possible events, i.e. the sigma algebra generated by x,

$$\sigma(x) = \sigma(\{x^{-1}(A) \mid A \in \mathcal{B}(\mathcal{X})\}).$$

In the case of a stochastic process (x_n), the information on the process up to time n is the σ-algebra generated by $x_0, \cdots, x_n$, i.e. $\sigma(x_0, \cdots, x_n)$.

With this in mind, we see that the notion of possible events evolving in time is naturally described by a sequence of increasing σ-algebras. We call such a sequence a filtration.

Definition 2.1 (Filtration)**.** A filtration is a sequence $(\mathcal{F}_n)$ of increasing sub-σ-algebras of $\mathcal{F}$.

Definition 2.2 (Adapted)**.** A stochastic process (x_n) is adapted to the filtration $(\mathcal{F}_n)$ if for all n, x_n is $\mathcal{F}_n$-measurable.

Definition 2.3 (Natural Filtration)**.** Given a stochastic process (x_n), the natural filtration $(\mathcal{F}^x_n)$ for (x_n) is

$$\mathcal{F}^x_n := \sigma(x_0, \cdots, x_n).$$

We note that by definition, a stochastic process is always adapted to its natural filtration.

Recalling the definition of conditional expectation, we introduce the following notations.

Definition 2.4 (Conditional Probability)**.** Given a σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ and a random variable x, we define the conditional probability of x with respect to $\mathcal{G}$ to be

$$\mathbb{P}(x \in A \mid \mathcal{G}) := \mathbb{E}(\mathbf{1}_A(X) \mid \mathcal{G}),$$

for all $A \in \mathcal{B}(\mathcal{X})$ where $\mathbf{1}_A$ is the indicator function of A.

Furthermore, given random variables $x_0, \cdots, x_n$, we denote

$$\mathbb{P}(x \in A \mid x_0, \cdots, x_n) := \mathbb{P}(x \in A \mid \sigma(x_0, \cdots, x_n)).$$

Definition 2.5 (Simple Markov Property)**.** A stochastic process (x_n) with state space $\mathcal{X}$ is said to have the simple Markov property if for any $A \in \mathcal{B}(\mathcal{X})$ and $n \geq 0$, we have

$$\mathbb{P}(x_{n+1} \in A \mid x_0, \cdots, x_n) = \mathbb{P}(x_{n+1} \in A \mid x_n),$$

almost surely.

Unfolding the notation, the simple Markov property states that

$$\mathbb{E}(\mathbf{1}_A(x_{n+1}) \mid \mathcal{F}^x_n) = \mathbb{E}(\mathbf{1}_A(x_{n+1}) \mid \sigma(x_n)).$$

We call a stochastic process which has the simple Markov property a Markov process and we call $\mathcal{L}(x_0)$ the initial distribution. Furthermore, if the Markov process is discrete, we call it a Markov chain.

The definition of the simple Markov property can be generalized to continuous stochastic processes by taking the property to be $\mathbb{E}(\mathbf{1}_A(x_t) \mid \mathcal{F}_s^x) = \mathbb{E}(\mathbf{1}_A(x_t) \mid \sigma(x_s))$ for all $s \leq t$.

In the case that $\mathcal{X} = \mathbb{N}$, the simple Markov property is equivalent to the statement that

$$\mathbb{P}(x_{n+1} = j \mid x_0 = i_0, \cdots, x_n = i_n) = \mathbb{P}(x_{n+1} = j \mid x_n = i_n),$$

almost surely for every n where $i_0, \cdots, i_n \in \mathcal{X} = \mathbb{N}$

$$\mathbb{P}(x_0 = i_0, \cdots, x_n = i_n) > 0.$$

Lemma 2.1. Let $\mathcal{G} \subseteq \mathcal{F}$, $X : \Omega \to \mathcal{X}, Y : \Omega \to \mathcal{Y}$ be random variables such that X is $\mathcal{G}$-measurable, Y is independent of $\mathcal{G}$. Then, if $\phi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is measurable such that $\phi(X, Y) \in L^1$, we have

$$\mathbb{E}(\phi(X, Y) \mid \mathcal{G})(\omega) = \mathbb{E}_Y(\phi(X(\omega), Y))$$

almost surely.

Proof. Exercise. □

Proposition 2.1. Let $\xi_1, \xi_2, \cdots$ be a sequence of independent random variables with state space $\mathcal{Y}$ and is independent with respect to $x_0 : \Omega \to \mathcal{X}$. Then, if $F : \mathcal{X} \times \mathcal{Y} \to \mathcal{X}$ is a measurable function, we may define the stochastic process

$$x_{n+1} = F(x_n, \xi_{n+1}).$$

(x_n) is a Markov process.

Proof. Let $A \in \mathcal{B}(\mathcal{X})$. Then,

$$\begin{aligned}\mathbb{E}(\mathbf{1}_A(x_{n+1}) \mid x_0, \cdots, x_n) &= \mathbb{E}(\mathbf{1}_A(F(x_n, \xi_{n+1}) \mid x_0, \cdots, x_n) \\ &= \omega \mapsto \mathbb{E}(\mathbf{1}_A(F(x_n(\omega), \xi_{n+1})),\end{aligned}$$

where the second equality follows by the above lemma (setting $\phi = \mathbf{1}_A \circ F$ and observing that x_n is $\sigma(x_0, \cdots, x_n)$-measurable and ξ_{n+1} is independent of $\sigma(x_0, \cdots, x_n)$). Similarly,

$$\begin{aligned}\mathbb{E}(\mathbf{1}_A(x_{n+1}) \mid x_n) &= \mathbb{E}(\mathbf{1}_A(F(x_n, \xi_{n+1}) \mid x_n) \\ &= \omega \mapsto \mathbb{E}(\mathbf{1}_A(F(x_n(\omega), \xi_{n+1})),\end{aligned}$$

we have $\mathbb{E}(\mathbf{1}_A(x_{n+1}) \mid \mathcal{F}_n^x) = \mathbb{E}(\mathbf{1}_A(x_{n+1}) \mid \sigma(x_n))$ as required. □

2.2 Markov Property

So far we have looked at the simple Markov property in which we have taken the filtration to be the natural filtration of the process. However, in the case that we are looking at multiple processes, we would like to consider a larger filtration such that each process is adapted. This motivates the general definition for the Markov property.

Definition 2.6. Let $(\mathcal{F}_t)_{t\in I}$ be a filtration indexed by the set I on the measurable space $(\Omega, \mathcal{F})$. A stochastic process $(x_t)_{t\in I}$ on $\mathcal{X}$ is a Markov process with respect to $\mathcal{F}_t$ if it is adapted to $\mathcal{F}_t$ and

$$\mathbb{P}(x_t \in A \mid \mathcal{F}_s) = \mathbb{P}(x_t \in A \mid x_s)$$

almost surely for all $s, t \in I$, $t > s$ and $A \in \mathcal{B}(\mathcal{X})$.

Again, unfolding the notation, the above statement says

$$\mathbb{E}(\mathbf{1}_A(x_t) \mid \mathcal{F}_s) = \mathbb{E}(\mathbf{1}_A(x_t) \mid \sigma(x_s))$$

almost surely.

Proposition 2.2. If (x_t) is a Markov process with respect to the filtration $(\mathcal{F}_t)$, then it is also a Markov process with respect to its natural filtration $(\mathcal{F}^x_t)$.

Proof. Recalling that $\mathcal{F}^x_t \subseteq \mathcal{F}_t$ for all t, by the tower property of the conditional expectation, we have

$$\begin{aligned}\mathbb{P}(x_{t+s} \in A \mid \mathcal{F}^x_s) &= \mathbb{E}(\mathbf{1}_A(x_{t+s}) \mid \mathcal{F}^x_s)\\ &= \mathbb{E}(\mathbb{E}(\mathbf{1}_A(x_{t+s}) \mid \mathcal{F}_s) \mid \mathcal{F}^x_s)\\ &= \mathbb{E}(\mathbb{E}(\mathbf{1}_A(x_{t+s}) \mid \sigma(x_s)) \mid \mathcal{F}^x_s),\end{aligned}$$

where the equalities denotes equal a.e. Thus, as $\mathbb{E}(\mathbf{1}_A(x_{t+s}) \mid \sigma(x_s))$ is $\sigma(x_s)$-measurable, and thus $\mathcal{F}^x_s$-measurable (since $\sigma(x_s) \subseteq \sigma(x_r \mid r \leq s) = \mathcal{F}^x_s$)), we have

$$\mathbb{E}(\mathbb{E}(\mathbf{1}_A(x_{t+s}) \mid \sigma(x_s)) \mid \mathcal{F}^x_s) = \mathbb{E}(\mathbf{1}_A(x_{t+s}) \mid \sigma(x_s))$$

implying that the Markov property is satisfied. □

Theorem 1. If (x_t) is a Markov process with respect to the filtration $(\mathcal{F}_t)$, then

$$\mathbb{E}(f(x_t) \mid \mathcal{F}_s) = \mathbb{E}(f(x_t) \mid \sigma(x_s))$$

almost surely for any $f : \mathcal{X} \to \mathbb{R}$ bounded and measurable. In particular, this property is equivalent to the Markov property by choosing $f = \mathbf{1}_A$ for all $A \in \mathcal{B}(\mathcal{X})$.

Proof. By linearity, the property holds for simple functions. Furthermore, by the conditional monotone convergence theorem, the property holds for any non-negative bounded measurable functions. Finally, for arbitrary bounded measurable functions f, the result follows by taking $f = f^+ - f^-$ and applying the non-negative case. □

Proposition 2.3. Let $C \in \mathcal{F}_s$ and suppose $\mathcal{B}(\mathcal{X}) = \sigma(\mathcal{D})$ where $\mathcal{D}$ is a π-system (i.e. non-empty and closed under finite intersections), then, if

$$\mathbb{E}(\mathbf{1}_A(x_{t+s})\mathbf{1}_C) = \mathbb{E}(\mathbb{P}(x_{t+s} \in A \mid x_s)\mathbf{1}_C)$$

holds for any $A \in \mathcal{D}$, it holds for any $A \in \mathcal{B}(\mathcal{X})$.

Proof. Let $\mathcal{A}$ be the set of Borel sets which the equation holds. Then, by definition $\mathcal{D} \subseteq \mathcal{A}$ and so, it suffices to show $\mathcal{A}$ is a λ-system (i.e. $\mathcal{A}$ contains $\mathcal{X}$, closed under complements and closed under countable unions of increasing sets). Indeed, Dynkin's $\pi - \lambda$ theorem states that if $\mathcal{D}$ is a π-system, $\mathcal{A}$ is a λ-system and $\mathcal{D} \subseteq \mathcal{A}$, then $\sigma(\mathcal{D}) \subseteq \mathcal{A}$.

Clearly $\mathcal{X} \in \mathcal{A}$ since

$$\mathbb{E}(\mathbf{1}_{\mathcal{X}}(x_{t+s})\mathbf{1}_C) = \mathbb{E}(\mathbf{1}_C) = \mathbb{E}(\mathbb{P}(x_{t+s} \in \mathcal{X} \mid x_s)\mathbf{1}_C).$$

Suppose now $A \in \mathcal{A}$. Then, the property holds as $\mathbf{1}_{A^c} = 1 - \mathbf{1}_A$ and so, the result follows by linearity. Finally, if $(A_n) \subseteq \mathcal{A}$ is increasing. Then by the monotone convergence theorem for conditional expectations, it follows that $\bigcup A_n \in \mathcal{A}$ and hence, $\mathcal{A}$ is a λ-system as required. □

Proposition 2.4. Suppose $\mathbb{E}(f(x_{n+1}) \mid x_0, \cdots, x_n) = \mathbb{E}(f(x_{n+1}) \mid x_n)$ for any bounded measurable f. Then, if we have a sequence

$$0 \le t_1 < t_2 < \cdots < t_{m-1} < t_m = n - 1,$$

where $n > 1, t_i \in \mathbb{N}$, for any bounded measurable functions f, h, we have

$$\mathbb{E}(f(x_{n+1})h(x_n) \mid x_{t_1}, \cdots, x_{t_m}) = \mathbb{E}(f(x_{n+1})h(x_n) \mid x_{n-1}).$$

Proof. Exercise. □

As we will often use the bounded measurable functions, let us denote the set of bounded measurable functions $f : \mathcal{X} \to \mathbb{R}$ by $\mathcal{B}_b(\mathcal{X})$.

Lemma 2.2. Let $X, Y \in L_1$ and $\mathcal{G} \subseteq \mathcal{F}$. Then, if X is $\mathcal{G}$-measurable and $XY \in L_1$, we have

$$\mathbb{E}(XY \mid \mathcal{G}) = X\mathbb{E}(Y \mid \mathcal{G}).$$

We call this property "taking out what is known".

Proof. See problem sheet 1. □

Theorem 2. Given a stochastic process (x_n) and indices $l < m < n$, TFAE.

- For any $f \in \mathcal{B}_b(\mathcal{X})$,
$$\mathbb{E}(f(x_n) \mid x_l, x_m) = \mathbb{E}(f(x_n) \mid x_m).$$
- For any $g \in \mathcal{B}_b(\mathcal{X})$,
$$\mathbb{E}(g(x_l) \mid x_m, x_n) = \mathbb{E}(g(x_l) \mid x_m).$$
- For any $f, g \in \mathcal{B}_b(\mathcal{X})$,
$$\mathbb{E}(f(x_n)g(x_l) \mid x_m) = \mathbb{E}(f(x_n) \mid x_m)\mathbb{E}(g(x_l) \mid x_m).$$

  That is to say, given now, the past is independent of the future.

Proof. Suppose the first statement holds, we will prove the third property. Let $f, g \in \mathcal{B}_b(\mathcal{X})$, then by the tower law and the above lemma, we have

$$\begin{aligned}
\mathbb{E}(f(x_n)g(x_l) \mid x_m) &= \mathbb{E}(\mathbb{E}(f(x_n)g(x_l) \mid x_m, x_l) \mid x_m) \\
&= \mathbb{E}(g(x_l)\mathbb{E}(f(x_n) \mid x_m, x_l) \mid x_m) \\
&= \mathbb{E}(g(x_l)\mathbb{E}(f(x_n) \mid x_m) \mid x_m) \\
&= \mathbb{E}(f(x_n) \mid x_m)\mathbb{E}(g(x_l) \mid x_m)
\end{aligned}$$

which is exactly the third property.

On the other hand, if the third property holds, for any $g, h \in \mathcal{B}_b(\mathcal{X})$, we have

$$\begin{aligned}
\mathbb{E}(f(x_n)h(x_m)g(x_l)) &= \mathbb{E}(\mathbb{E}(f(x_n)g(x_l) \mid x_m))h(x_m)) \\
&= \mathbb{E}(\mathbb{E}(f(x_n) \mid x_m))\mathbb{E}(g(x_l) \mid x_m)h(x_m)) \\
&= \mathbb{E}(\mathbb{E}(\mathbb{E}(f(x_n) \mid x_m)g(x_l)h(x_m) \mid x_m)) \\
&= \mathbb{E}(\mathbb{E}(f(x_n) \mid x_m)g(x_l)h(x_m))
\end{aligned}$$

where the last equality is due to the law of total expectation. Now, by considering this equality implies that, for all $A = A_1 \cap A_2$ where $A_1 \in \sigma(x_l), A_2 \in \sigma(x_m)$, by choosing $g = \mathbf{1}_{x_l^{-1}(A_1)}$ and $h = \mathbf{1}_{x_m^{-1}(A_2)}$, we have

$$\int_A f(x_n) \mathrm{d}\mathbb{P} = \int_A \mathbb{E}(f(x_n) \mid x_m) \mathrm{d}\mathbb{P},$$

and so, $\mathbb{E}(f(x_n) \mid x_m) = \mathbb{E}(f(x_n) \mid x_l, x_m)$ almost surely. Hence, the first and third property are equivalent. Similarly, one can should that the second property is equivalent to the third property and hence the equivalence. $\square$

Proposition 2.5. A stochastic process (x_n) is a Markov process if and only if one of the following conditions holds:

- for any $f_i \in \mathcal{B}_b(\mathcal{X})$,

$$\mathbb{E}\left(\prod_{i=1}^{n} f_i(x_i)\right) = \mathbb{E}\left(\prod_{i=1}^{n-1} f_i(x_i)\mathbb{E}(f_n(x_n) \mid x_{n-1})\right).$$

- for any $A_i \in \mathcal{B}(\mathcal{X})$,

$$\mathbb{P}(x_0 \in A_0, \cdots, x_n \in A_n) = \int_{\bigcap_{i=0}^{n-1}\{x_i \in A_i\}} \mathbb{P}(x_n \in A_n \mid x_{n-1}) \mathrm{d}\mathbb{P}.$$

Proof. We note that by choosing $f_i = \mathbf{1}_{A_i}$, the first condition implies the second. On the other hand, the reverse implication follows by the standard routine of proving it for simple function and using monotone convergence. Thus, it suffices to establish an equivalence between the first condition and the Markov property. This is left as an exercise. $\square$

2.3 Gaussian Measure and Gaussian Process

As one of the most important distributions in probability theory, let us in this short section introduce the Gaussian measure which we will again encounter later on with this course.

Definition 2.7 (Gaussian Measure)**.** A measure μ on $\mathbb{R}^n$ is Gaussian if there exists a non-negative definite symmetric matrix K and $m \in \mathbb{R}^n$ such that the Fourier transform of μ is

$$\int_{\mathbb{R}^n} e^{i\langle \lambda, x\rangle}\mu(\mathrm{d}x) = e^{i\langle \lambda, m\rangle - \frac{1}{2}\langle K\lambda, \lambda\rangle}$$

for any $\lambda \in \mathbb{R}^n$. We call the matrix K the covariance of μ and m its mean.

We remark that if X is a random variable with distribution μ, then the Fourier transform of μ is simply the characteristic function of X, $\mathbb{E}(e^{i\langle \lambda, X\rangle})$.

Proposition 2.6. If μ is a Gaussian measure with covariance K and mean m is absolutely continuous with respect to the Lebesgue measure if and only if K is non-degenerate. In this case, for all $A \in \mathcal{B}(\mathbb{R}^n)$,

$$\mu(A) = \int_A \frac{1}{\sqrt{(2\pi)^n \det K}} e^{-\frac{1}{2}\langle K^{-1}(x-m), x-m\rangle}\lambda(\mathrm{d}x).$$

We observe that if X is a random variable with Gaussian distribution μ, then as one might expect, $\mathbb{E}(X) = m$ and

$$\mathrm{Cov}(X_i, X_j) := \mathbb{E}(X_i - m_i)(X_j - m_j) = K_{ij}.$$

For this reason, we call K the covariance operator.

Theorem 3. If X is a Gaussian random variable (i.e. its distribution is Gaussian) on $\mathbb{R}^d$ with covariance K. Then, if $A : \mathbb{R}^d \to \mathbb{R}^m$ is a linear transformation, then AX is Gaussian with covariance AKA^T and mean Am.

Proof. Follows since,

$$\mathbb{E}e^{i\langle \lambda, AX\rangle} = \mathbb{E}e^{i\langle A^T\lambda, X\rangle} = e^{i\langle A^T\lambda, m\rangle - \frac{1}{2}\langle KA^T\lambda, A^T\lambda\rangle} = e^{i\langle \lambda, Am\rangle - \frac{1}{2}\langle AKA^T\lambda, \lambda\rangle}.$$

□

Proposition 2.7. Linear combinations of independent Gaussian random variables are also Gaussian.

Proof. Exercise. □

Definition 2.8 (Gaussian Process)**.** A stochastic process is Gaussian if its finite dimensional distributions are Gaussian.

2.4 Kolmogorov's Extension Theorem

Let (x_n) be a stochastic process with state space $\mathcal{X}$. Denote

$$\mathcal{X}^{\mathbb{N}_0} := \prod_{i=0}^{\infty} \mathcal{X} = \{(a_0, a_1, \cdots) \mid a_i \in \mathcal{X}\}.$$

We may consider (x_n) as a map from $\Omega \to \mathcal{X}^{\mathbb{N}_0}$ by defining

$$(x_n) : \Omega \to \mathcal{X}^{\mathbb{N}_0} : \omega \mapsto (x_n(\omega))_{n=0}^{\infty}.$$

We would like (x_n) to be measurable as a map from $\Omega \to \mathcal{X}^{\mathbb{N}_0}$ and so, let us first equip $\mathcal{X}^{\mathbb{N}_0}$ with a σ-algebra.

Definition 2.9. Given $\mathcal{X}_i$ complete separable metric spaces for $i \in \Lambda$, define the projection maps

$$\pi_m : \prod_{i \in \Lambda} \mathcal{X}_i \to \mathcal{X}_m : (a_i)_{i \in \Lambda} \mapsto a_m.$$

Then, we define $\bigotimes_{i \in \Lambda} \mathcal{B}(\mathcal{X}_i) = \sigma(\pi_i \mid i \in \Lambda)$.

We note that this definition can be easily extended to arbitrary measurable spaces.

Proposition 2.8. If (x_n) is a stochastic process, then as a map from $\Omega \to \mathcal{X}^{\mathbb{N}_0}$, (x_n) is $\bigotimes_n \mathcal{B}(\mathcal{X})$-measurable.

With this in mind, we can push forward the probability measure along a stochastic process, inducing a measure on $\mathcal{X}^{\mathbb{N}_0}$. In particular, we have the measure space $(\mathcal{X}^{\mathbb{N}_0}, \bigoplus_n \mathcal{B}(\mathcal{X}), (x_n)_* \mathbb{P})$.

On the other hand, if we only consider the first n-components of the process (x_i), by the same argument, $(x_i)_{i=1}^n$ forms a measurable map from $\Omega \to \mathcal{X}^n$. Then, in this case, we call the push-forward measure $\mathcal{L}((x_i)_{i=1}^n) = {(x_i)_{i=1}^n}_* \mathbb{P}$ the joint distribution. These are known as the finite dimensional distributions of (x_n).

Definition 2.10. Let $(\mu_n)_{n=1}^{\infty}$ be a sequence of probability measures on $\mathcal{X}^n$ (i.e. μ_n is aprobability measures on $\mathcal{X}^n$). Then $(\mu_n)_{n=0}^{\infty}$ is said to satisfy Kolmogorov's consistency condition of

$$\mu_{n+1}(A_1 \times \cdots A_n \times \mathcal{X}) = \mu_n(A_1 \times \cdots \times A_n)$$

for all $n \geq 0$, $A_i \in \mathcal{B}(\mathcal{X})$.

Theorem 4 (Kolmogorov's Extension Theorem)**.** Let $(\mu_n)_{n=1}^{\infty}$ be a sequence of probability measures which are consistent. Then, there exists a unique probability measure μ on $\mathcal{X}^{\mathbb{N}_0}$ such that for any n, $A \in \bigotimes_{i=1}^n \mathcal{B}(\mathcal{X})$,

$$\mu(A \times \mathcal{X}^{\mathbb{N}_0}) = \mu_n(A).$$

In other words, if we denote pr_n the map

$$\mathrm{pr}_n : \mathcal{X}^{\mathbb{N}_0} \to \mathcal{X}^n : (a_i)_{i=1}^{\infty} \mapsto (a_i)_{i=1}^n,$$

μ is the unique measure which satisfies

$$(\mathrm{pr}_n)_* \mu = \mu_n.$$

Corollary 4.1. The finite dimensional distributions of a stochastic process (x_n) determines uniquely the probability distribution of the process on $\mathcal{X}^{\mathbb{N}_0}$.

Corollary 4.2. Given any consistent family of probability measures (μ_n), there exists a stochastic process with (μ_n) as its finite dimensional distributions.

Proof. Kolmogorov's extension theorem implies that there exists a compatible measure μ on $\mathcal{X}^{\mathbb{N}_0}$. Thus, it suffices to find a $\mathcal{X}^{\mathbb{N}_0}$-valued random variable with distribution μ. We will describe a trivial method for this purpose below. $\square$

Let μ be a probability measure on $\mathcal{X}$. Then, setting $\Omega = \mathcal{X}, \mathcal{F} = \mathcal{B}(\mathcal{X})$ and $\mathbb{P} = \mu$, it is clear that the push-forward of $\mathbb{P}$ along the identity map provides μ. Thus, the identity is a random variable with the distribution μ.

Thus, in the case of the corollary above, we set $\Omega = \mathcal{X}^{\mathbb{N}_0}$, $\mathcal{F} = \bigotimes \mathcal{B}(\mathcal{X})$, and $\mathbb{P} = \mu$. We call this probability space the canonical probability space and call the resulting process (π_n) the canonical process.

Definition 2.11 (Shift Operator)**.** For $n \in \mathbb{N}$, we define the n-th shift operator by

$$\theta_n : \mathcal{X}^{\mathbb{N}_0} \to \mathcal{X}^{\mathbb{N}_0} : (a_0, a_1, \cdots) \mapsto (a_n, a_{n+1}, \cdots).$$

For connivance, we will denote the above property by $\theta_n(a.) = (a_{n+}.)$.

Definition 2.12 (Stationary Process)**.** A stochastic process $(x.)$ is stationary if for any n,

$$\mathcal{L}(\theta_n x.) = \mathcal{L}(x.).$$

Straight away, by Kolmogorov's extension, we see that an equivalent definition for the stationary process is that the finite dimensional distributions of $(\theta_n x.)$ are the same as the finite dimensional distributions of $(x.)$.

In general, a process (x_n) is stationary if

$$\mathcal{L}(x_n, \cdots, x_{n+m}) = \mathcal{L}(x_0, \cdots, x_m)$$

for all $n, m \geq 0$. In the case that (x_n) is a Gaussian process, as $\mathcal{L}(x_{i_1} \cdots x_{i_n})$ is determined by $(\mathbb{E}x_{i_1}, \cdots, \mathbb{E}x_{i_n})$ and $\mathrm{Cov}(x_{i_k}, x_{i_l})$, it is stationary if

$$\mathbb{E}x_n = \mathbb{E}x_0$$

for all n and the covariances are shift invariant.

2.5 Transition Probability

Given a Markov process (x_n) on the state space $\mathcal{X}$, for $A \in \mathcal{B}(\mathcal{X})$, the function $\mathbb{P}(x_{n+1} \in A \mid x_n)$ is a Borel function of x_n. This function might depend on A, n and $n-1$. Suppose the case that this function does not depend on time (i.e. time homogeneous), that is there exists some function $\phi(x, A)$ such that

$$\mathbb{P}(x_{n+1} \in A \mid x_n) = \phi(x_n, A)$$

almost surely. Fixing x, under some regularities, its not difficult to show that $\phi(x, \cdot)$ form a probability measure. We shall assume this.

Definition 2.13. The set $P := \{P(x, A) \mid x \in \mathcal{X}, A \in \mathcal{B}(\mathcal{X})\}$ is said to be a family of transition probabilities if

- for all $x \in \mathcal{X}$, $P(x, \cdot)$ is a probability measure;
- for any $A \in \mathcal{B}(\mathcal{X})$, $P(\cdot, A)$ is Borel measurable.

As an example, consider the Markov chain (x_n) on $\mathcal{X}$ where $x_{n+1} := F(x_n, \xi_{n+1})$ for $x_0, \xi_1, \xi_2, \cdots$ independent with $(\xi_i : \Omega \to \mathcal{Y}) \sim \mu$ for all i. Then, for all $\omega \in \Omega$, we have (recall that we denote the event $x_n^{-1}(\{\omega\})$ by $x_n = \omega$

$$\begin{aligned}\mathbb{P}(x_{n+1} \in A \mid x_n = \omega) &= \mathbb{P}(F(x_n(\omega), \xi_{n+1}) \in A) \\ &= \int_{\mathcal{Y}} \mathbf{1}_A(F(x_n(\omega), \xi_{n+1}))\mathrm{d}\mathbb{P} \\ &= \int_{\mathcal{Y}} \mathbf{1}_A(F(x_n(\omega), y))\mu(\mathrm{d}y).\end{aligned}$$

Hence, defining

$$P(x, A) := \int_{\mathcal{Y}} \mathbf{1}_A(F(x, y))\mu(\mathrm{d}y),$$

$\{P(x, A)\}$ are the transition probabilities and

$$\mathbb{P}(x_{n+1} \in A \mid x_n = \omega) = P(x_n(\omega), A).$$

We remark that, in the case that the state space is countable, by σ-additivity, it is sufficient to work with transitional probabilities of singletons. In particular, the transitional probability is simply determined by

$$P(i, j) = P(i, \{j\}), i, j \in \mathcal{X}.$$

Proposition 2.9. Let (x_n) be a Markov chain such that there exists a transition probability P for which

$$\mathbb{P}(x_{n+1} \in A \mid x_n) = P(x_n, A)$$

for all $A \in \mathcal{B}(\mathcal{X})$. Then, for all $f \in \mathcal{B}_b(\mathcal{X})$,

$$\mathbb{E}(f(x_{n+1}) \mid x_n) = \int_{\mathcal{X}} f(y)P(x_n, \mathrm{d}y).$$

Proof. Choosing $f = \mathbf{1}_A$, we see that the property is true for simple functions and so, the results can be extended to all functions by the monotone convergence theorem. $\square$

Proposition 2.10. Let (x_n) as defined above, for all $f \in \mathcal{B}_b(\mathcal{X})$, we have

$$\mathbb{P}(x_{n+2} \in A \mid x_n) = \int_{\mathcal{X}} P(y, A)P(x_n, \mathrm{d}y).$$

Thus, we define the two-step transition probability by

$$P^2(x, A) = \int_{\mathcal{X}} P(y, A)P(x, \mathrm{d}y),$$

such that $\mathbb{P}(x_{n+2} \in A \mid x_n) = P^2(x_n, A)$ almost surely.

Proof. By the tower law, we have

$$\begin{aligned}
\mathbb{P}(x_{n+2} \in A \mid x_n) &= \mathbb{E}(\mathbb{E}(\mathbf{1}_A(x_{n+2}) \mid x_{n+1}, x_n) \mid x_n) \\
&= \mathbb{E}(\mathbb{E}(\mathbf{1}_A(x_{n+2}) \mid x_{n+1}) \mid x_n) \\
&= \mathbb{E}(P(x_{n+1}, A) \mid x_n) \\
&= \int_{\mathcal{X}} P(y, A) P(x_n, \mathrm{d}y),
\end{aligned}$$

where the last equality follows by the above proposition. □

The above process can be extended to k-steps by induction. In particular, for all $k \in \mathbb{N}$, we have

$$\mathbb{P}(x_{n+(k+1)} \in A) \mid x_n) = \int_{\mathcal{X}} P^i(y, A) P^j(x_n, \mathrm{d}y)$$

for any $i, j \in \mathbb{N}, i + j = k$. This is known as the Chapman-Kolmogorov equation.

Definition 2.14. A family

$$\{P^n(x, \cdot) \mid x \in \mathcal{X}, n \in \mathbb{N}_0\}$$

is said to be a transition function if

- $P^n(x, \cdot)$ is a transition probability for any x, n;
- $P^0(x, \cdot) = \delta_x$ for any x;
- (Chapman-Kolmogorov) for all $n, m \geq 0, x \in \mathcal{X}, A \in \mathcal{B}(\mathcal{B})$,

$$P^{n+m}(x, A) = \int_{\mathcal{X}} P^n(y, A) P^m(x, \mathrm{d}y).$$

Proposition 2.11. The Chapman-Kolmogorov equation is satisfied if and only if for all $f \in \mathcal{B}_b(\mathcal{X}), n, m \geq 0, x \in \mathcal{X}, A \in \mathcal{B}(\mathcal{B})$,

$$\int_{\mathcal{X}} f(y) P^{n+m}(x, \mathrm{d}y) = \int_{\mathcal{X}} \left(\int_{\mathcal{X}} f(z) P^n(y, \mathrm{d}z) \right) P^m(x, \mathrm{d}y).$$

Proof. Exercise. □

As alluded to above, the k-step transitional probabilities can be constructed from a 1-step transitional probabilities. In particular, given the 1-step transitional probability P,

1. set $P^0(x, \cdot) = \delta_x$;
2. set $P^1(x, \cdot) = P(x, \cdot)$;
3. for all $n > 1$, $x \in \mathcal{X}$, for all $A \in \mathcal{B}(\mathcal{X})$, set

$$P^{n+1}(x, A) := \int_{\mathcal{X}} P(y, A) P^n(x, \mathrm{d}y).$$

It remains to show that this construction satisfies the Chapman-Kolmogorov equation. Indeed, by induction, suppose that the Chapman-Kolmogorov equation is satisfied for all $k \leq n + m$, then

$$\begin{aligned} P^{n+m+1}(x, A) &= \int_{\mathcal{X}} P(z, A) P^{n+m}(x, \mathrm{d}z) \\ &= \int_{\mathcal{X}} \left(\int_{\mathcal{X}} P(z, A) P^{j}(y, \mathrm{d}y) \right) P^{n+m-j}(x, \mathrm{d}y) \\ &= \int_{\mathcal{X}} P^{j+1}(y, A) P^{n+m-j}(x, \mathrm{d}y) \end{aligned}$$

for all $j = 0, 1, \cdots$, where the second and third equality follows by the inductive hypothesis and Fubini's theorem.

Definition 2.15. A transition probability P is said to be the transitional probability of the Markov chain (x_n) if

$$\mathbb{P}(x_{n+1} \in A \mid x_n) = P(x_n, A)$$

almost surely for all $A \in \mathcal{B}(\mathcal{X})$ and any $n \geq 0$.

If a Markov chain has a transitional probability P, then we say the Markov chain is time homogeneous.

From this point forward, unless otherwise stated, we will assume our Markov chains to be time homogeneous.

Consider again the case where the state space is countable $\mathcal{X} = \mathbb{N}$. As mentioned previously, the transition probability on $\mathcal{X}$ is then determined by $p_{ij} = P(i, \{j\})$. As $P(i, \cdot)$ is a probability measure by definition,

$$1 = P(i, \mathcal{X}) = \sum_{j \in \mathcal{X}} p_{ij}.$$

In the case that $\mathcal{X}$ is finite, these p_{ij} can be represented as a matrix, motivating the definition of a stochastic matrix.

Definition 2.16 (Stochastic Matrix)**.** A matrix $p = (p_{ij})$ with $p_{ij} \geq 0$ is said to be a stochastic matrix if $\sum_{j \in \mathcal{X}} p_{ij} = 1$.

In the discrete case, our construction of the transition probability from the 1-step transition probability is straightforward. In particular, we obtain

$$P^{n+1}(i, A) = \int_{\mathcal{X}} P(y, A) P^{n}(i, \mathrm{d}y) = \sum_{k \in \mathcal{X}} P(k, A) P^{n}(i, k).$$

Thus, if we write $P(i, \{j\}) = p_{ij}$, then

$$P^{2}(i, \{j\}) = \sum_{k \in \mathcal{X}} p_{ik} p_{kj} = ((p_{kl})^{2}_{k,l \in \mathcal{X}})_{ij},$$

where the last term denotes matrix multiplication. Thus, by induction, we obtain that

$$P^{n}(i, \{j\}) = \sum_{k_1 \in \mathcal{X}} \cdots \sum_{k_{n-1} \in \mathcal{X}} p_{ik_1} p_{k_1 k_2} \cdots p_{k_{n-1} j} = ((p_{kl})^{n}_{k,l \in \mathcal{X}})_{ij}.$$