Final draft

Introduction

Chapter 1

Markov Processes and State-Space Models

1.1 Notation

Let us fix the basic notation that will be employed henceforth. Upper case letters, e.g. X_t, Y_t, denote random variables, while lower case letters denote the realizations. Finite sequences are made compact using the semi-colon notation, e.g. $x_{0:t} = (x_0, x_1, \dots, x_t)$ for some $t \geq 0$. The sets of outcomes are denoted by upper case calligraphic letters, e.g. $\mathcal{X}, \mathcal{Y}$. The σ-algebra of a set $\mathcal{X}$ reads $\mathcal{B}(\mathcal{X})$. The pair $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is a measurable space. The densities are represented with lower case cursive letters. For example, the density of X_t in x_t is $p_t^\theta(x_t)$, where $\theta \in \Theta$ is some (eventually known) vector, while the density of X_t in x_t, given that X_{t-1} has previously taken value x_{t-1}, is $p_t^\theta(x_t \mid x_{t-1})$. Note that $p_t^\theta(x_t \mid x_{t-1})$ is a member of the parametric family with parameters (θ, x_{t-1}). The probability distribution of a random variable X_t is represented by $\mathbb{P}_t(dx_t)$, while the probability distribution of a sequence $X_{0:t}$ is represented by $\mathbb{P}_t(dx_{0:t})$. The Dirac measure, i.e., the measure that assigns probability 1 to the singleton $\{x\}$, is denoted by $\delta_x(dy)$.The expectation operator is denoted by $\mathbb{E}$, where the section of $\mathbb{E}$ at $\mathbb{Q}$, i.e., $\mathbb{E}_{\mathbb{Q}}$ denotes the operator computed with respect to the probability distribution $\mathbb{Q}$. Let φ be a function defined on the random variable $X \propto \mathbb{Q}$. Then, we denote the expected value of $\varphi(X)$ with

$$\mathbb{E}_{\mathbb{Q}}[\varphi(X)] = \int_{\mathbb{X}} \varphi(x)\mathbb{Q}(dx).$$

1.2 Markov Processes

Let $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ and $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ be two (eventually equal) measurable spaces.

Definition A function $P(x, dy) : (\mathcal{X}, \mathcal{B}(\mathcal{X})) \to [0, 1]$ such that

- for all $x \in \mathcal{X}, P(x, \cdot)$ is a probability measure on $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$, and
- for all subsets $A \in \mathcal{B}(\mathcal{Y})$, the map $x \mapsto P(x, A)$ is measurable in $\mathcal{B}(\mathcal{X})$),

is defined a probability kernel from $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ to $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$.

Notice that for any random variables X_1 and $X_0 = x_0$,

$$\mathbb{P}_1(dx_{0:1}) = \mathbb{P}_0(dx_0)P_1(x_0, dx_1),$$

where $P_1(x_0, dx_1)$ is a probability kernel from $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ to $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. This remark, provides some intuition about the link between probability kernels and conditional probabilities. As a matter of fact, it is possible to show that for given $X_{0:1}$ with probability distribution $\mathbb{P}_1(dx_{0:1})$,

$$\mathbb{P}_1(X_1 \in dx_1 \mid X_0 = x_0) = P_1(x_0, dx_1). \tag{1.1}$$

Probability kernels can be used to define Markov processes. Before moving to the definition, we will introduce some important concepts. Recall that given two measures $\mathbb{M}$ and $\mathbb{Q}$ defined on a measurable space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, $\mathbb{Q}$ is absolutely continuous with respect to $\mathbb{M}$, that is $\mathbb{Q} \ll \mathbb{M}$, if for all sets $A \in \mathcal{B}(\mathcal{X})$ such that $\mathbb{M}(A) = 0$, then it must also hold that $\mathbb{Q}(A) = 0$. If $\mathbb{Q}(A)/\mathbb{M}(A)$ is well-defined for all A, then the Radon-Nikodym theorem guarantees that there exists a measurable function $w(x) \geq 0$ such that

$$w(x) = \frac{\mathbb{Q}(dx)}{\mathbb{M}(dx)},$$

meaning that,

$$\mathbb{Q}(A) = \int_{\mathcal{X}} w(x)\mathbb{M}(dx).$$

Let $P_1, P_2, \ldots, P_T$ be a finite measure of probability kernels from $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ to $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. Fix some $\mathbb{P}_0(dx)$ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.

Definition A discrete-time Markov process is a sequence $X_{0:T}$ of random variables whose joint distribution can be written as

$$\mathbb{P}_T(X_{0:T} \in dx_{0:T}) = \mathbb{P}_0(dx_0) \prod_{s=1}^{T} P_s(x_{s-1}, x_s). \tag{1.2}$$

The set $\mathcal{X}$ is the state-space, $\mathbb{P}_0$ is the initial distribution and the probability kernel P_t is the transition kernel.

It is possible to show that for any $t \in T$,

$$\mathbb{P}_T(X_t \in dx_t \mid X_{0:t-1} \in dx_{0:t-1}) = \mathbb{P}_T(X_t \in dx_t \mid X_{t-1} = x_{t-1}) = \mathbb{P}_t(x_{t-1}, dx_t).$$

The first equality implies conditional independence. The second equality implies that it is possible to identify conditional distributions with transition kernels. Assume $\mathbb{P}_t, t \leq T$, is a sequence of probability measures. Then, the following proposition holds.

Proposition 1

For all $t \leq T$,

$$\mathbb{P}_T(dx_{0:t}) = \mathbb{P}_0(dx_0) \prod_{s=1}^{t} P_s(x_{s-1}, dx_s) = \mathbb{P}_t(dx_{0:t}).$$

This means that the marginal distribution of $X_{0:t}$ with respect to $\mathbb{P}_T$ is $\mathbb{P}_t$.

1.3 State-Space Models

Chapter 2

Linear-Gaussian case

Chapter 3

Particle Filtering