Chapter 6

Stochastic Processes

Let us summarize what we have already accomplished. Our ultimate goal was to learn to make decisions under uncertainty. We introduced a language of *probability* in Chapter 2 and learned how to measure uncertainty. Then, through Chapters 3–5, we studied *random variables*, *random vectors*, and their *distributions*. Have we learned enough to describe a situation involving uncertainty and be able to make good decisions?

Let us look around. If you say "Freeze!" and everything freezes for a moment, the situation will be completely described by random variables surrounding us at a particular moment of time. However, the real world is dynamic. Many variables develop and change in real time: air temperatures, stock prices, interest rates, football scores, popularity of politicians, and also, the CPU usage, the speed of internet connection, the number of concurrent users, the number of running processes, available memory, and so on.

We now start the discussion of *stochastic processes*, which are random variables that evolve and change in time.

6.1 Definitions and classifications

DEFINITION 6.1

A **stochastic process** is a random variable that also depends on time. It is therefore a function of two arguments, $X(t,\omega)$, where:

- $t \in \mathcal{T}$ is time, with $\mathcal{T}$ being a set of possible times, usually $[0,\infty)$, $(-\infty,\infty)$, $\{0,1,2,\ldots\}$, or $\{\ldots,-2,-1,0,1,2,\ldots\}$;
- $\omega \in \Omega$, as before, is an outcome of an experiment, with Ω being the whole sample space.

Values of $X(t,\omega)$ are called *states*.

At any fixed time t, we see a random variable $X_t(\omega)$, a function of a random outcome. On the other hand, if we fix ω, we obtain a function of time $X_\omega(t)$. This function is called a *realization*, a *sample path*, or a *trajectory* of a process $X(t,\omega)$.

Example 6.1 (CPU USAGE). Looking at the past usage of the central processing unit (CPU), we see a realization of this process until the current time (Figure 6.1a). However, the future behavior of the process is unclear. Depending on which outcome ω will actually take place, the process can develop differently. For example, see two different trajectories for $\omega = \omega_1$ and $\omega = \omega_2$, two elements of the sample space Ω, on Figure 6.1b.

Remark: You can observe a similar stochastic process on your personal computer. In the latest versions of Windows, Ctrl-Alt-Del, pressed simultaneously and followed by "Windows Task Manager," will show the real-time sample path of CPU usage under the tab "Performance." ♢

Depending on possible values of T and X, stochastic processes are classified as follows.

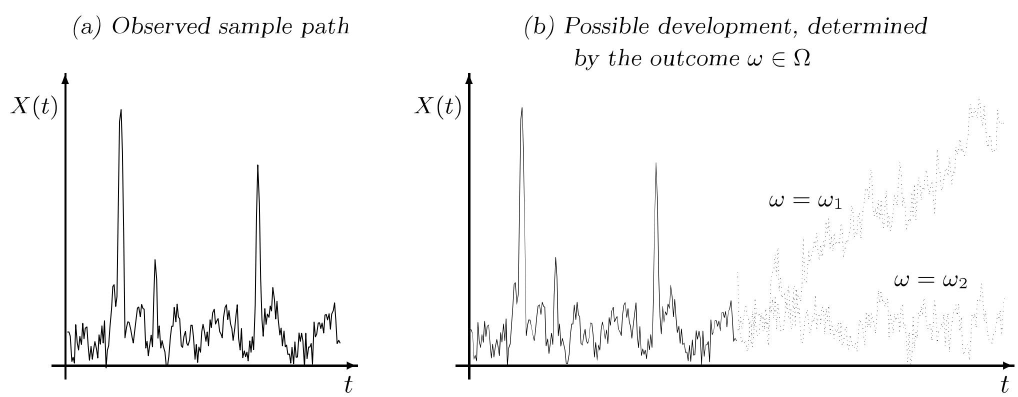

FIGURE 6.1: *Sample paths of CPU usage stochastic process.*

DEFINITION 6.2

Stochastic process $X(t,\omega)$ is **discrete-state** if variable $X_t(\omega)$ is discrete for each time t, and it is a **continuous-state** if $X_t(\omega)$ is continuous.

DEFINITION 6.3

Stochastic process $X(t,\omega)$ is a **discrete-time process** if the set of times $\mathcal{T}$ is discrete, that is, it consists of separate, isolated points. It is a **continuous-time process** if $\mathcal{T}$ is a connected, possibly unbounded interval.

Example 6.2. The CPU usage process, in percents, is continuous-state and continuous-time, as we can see in Figure 6.1a. $\Diamond$

Example 6.3. The *actual* air temperature $X(t,\omega)$ at time t is a continuous-time, continuous-state stochastic process. Indeed, it changes smoothly and never jumps from one value to another. However, the temperature $Y(t,\omega)$ reported on a radio every 10 minutes is a discrete-time process. Moreover, since the reported temperature is usually rounded to the nearest degree, it is also a discrete-state process. $\Diamond$

Example 6.4. In a printer shop, let $X(n,\omega)$ be the amount of time required to print the n-th job. This is a discrete-time, continuous-state stochastic process, because $n = 1,2,3,\ldots$, and $X \in (0,\infty)$.

Let $Y(n,\omega)$ be the number of pages of the n-th printing job. Now, $Y = 1,2,3,\ldots$ is discrete; therefore, this process is discrete-time and discrete-state.

$\Diamond$

From now on, we shall not write ω as an argument of $X(t,\omega)$. Just keep in mind that behavior of a stochastic process depends on chance, just as we did with random variables and random vectors.

Another important class of stochastic processes is defined in the next section.

6.2 Markov processes and Markov chains

DEFINITION 6.4

Stochastic process $X(t)$ is **Markov** if for any $t_1 < ... < t_n < t$ and any sets $A; A_1,\ldots,A_n$

$$\begin{aligned}\boldsymbol{P}\left\{X(t)\in A \mid X(t_1)\in A_1,\ldots,X(t_n)\in A_n\right\} \\ = \boldsymbol{P}\left\{X(t)\in A \mid X(t_n)\in A_n\right\}. \end{aligned} \tag{6.1}$$

Let us look at the equation (6.1). It means that the conditional distribution of $X(t)$ is the same under two different conditions,

(1) given observations of the process X at several moments in the past;

(2) given only *the latest* observation of X.

If a process is Markov, then its future behavior is the same under conditions (1) and (2). In other words, knowing the present, we get no information from the past that can be used to predict the future,

$$\boldsymbol{P}\{\text{ future } \mid \text{ past, present }\} = \boldsymbol{P}\{\text{ future } \mid \text{ present }\}$$

Then, for the future development of a Markov process, only its present state is important, and it does not matter *how* the process arrived to this state.

Some processes satisfy the Markov property, and some don't.

Example 6.5 (Internet connections). Let $X(t)$ be the total number of internet connections registered by some internet service provider by the time t. Typically, people connect to the internet at random times, regardless of how many connections have already been made. Therefore, the number of connections in a minute will only depend on the current number. For example, if 999 connections have been registered by 10 o'clock, then their total number will exceed 1000 during the next minute regardless of *when* and *how* these 999 connections were made in the past. This process is *Markov*. $\Diamond$

Example 6.6 (Stock prices). Let $Y(t)$ be the value of some stock or some market index at time t. If we know $Y(t)$, do we also want to know $Y(t-1)$ in order to predict $Y(t+1)$? One may argue that if $Y(t-1) < Y(t)$, then the market is rising, therefore, $Y(t+1)$ is likely (but not certain) to exceed $Y(t)$. On the other hand, if $Y(t-1) > Y(t)$, we may conclude that the market is falling and may expect $Y(t+1) < Y(t)$. It looks like knowing the past *in addition* to the present did help us to predict the future. Then, this process is *not Markov*. $\Diamond$

Due to a well-developed theory and a number of simple techniques available for Markov processes, it is important to know whether the process is Markov or not. The idea of Markov dependence was proposed and developed by *Andrei Markov* (1856–1922) who was a student of P. Chebyshev (p. 54) at St. Petersburg University in Russia.

6.2.1 Markov chains

DEFINITION 6.5

> A **Markov chain** is a discrete-time, discrete-state Markov stochastic process.

Introduce a few convenient simplifications. The time is discrete, so let us define the time set as $\mathcal{T} = \{0, 1, 2, \ldots\}$. We can then look at a Markov chain as a random sequence

$$\{X(0), X(1), X(2), \ldots\}.$$

The state set is also discrete, so let us enumerate the states as $1, 2, \ldots, n$. Sometimes we'll start enumeration from state 0, and sometimes we'll deal with a Markov chain with infinitely many (discrete) states, then we'll have $n = \infty$.

The *Markov* property means that only the value of $X(t)$ matters for predicting $X(t+1)$, so the conditional probability

$$\begin{aligned} p_{ij}(t) &= \boldsymbol{P}\left\{X(t+1) = j \mid X(t) = i\right\} \\ &= \boldsymbol{P}\left\{X(t+1) = j \mid X(t) = i,\ X(t-1) = h,\ X(t-2) = g, \ldots\right\} \end{aligned} \tag{6.2}$$

depends on i, j, and t only and equals the probability for the Markov chain X to make a *transition* from state i to state j at time t.

DEFINITION 6.6

Probability $p_{ij}(t)$ in (6.2) is called a **transition probability**. Probability

$$p_{ij}^{(h)}(t) = \boldsymbol{P}\left\{X(t+h) = j \mid X(t) = i\right\}$$

of moving from state i to state j by means of h transitions is an **h-step transition probability**.

DEFINITION 6.7

A Markov chain is **homogeneous** if all its transition probabilities are independent of t. Being homogeneous means that transition from i to j has the same probability at any time. Then $p_{ij}(t) = p_{ij}$ and $p_{ij}^{(h)}(t) = p_{ij}^{(h)}$.

Characteristics of a Markov chain

What do we need to know to describe a Markov chain?

By the Markov property, each next state should be predicted from the previous state only. Therefore, it is sufficient to know the distribution of its initial state $X(0)$ and the mechanism of transitions from one state to another.

The distribution of a Markov chain is completely determined by the initial distribution P_0 *and one-step transition probabilities* p_{ij}. Here P_0 is the probability mass function of X_0,

$$P_0(x) = \boldsymbol{P}\left\{X(0) = x\right\} \quad \text{for} \quad x \in \{1, 2, \ldots, n\}$$

Based on this data, we would like to compute:

- *h-step transition probabilities* $p_{ij}^{(h)}$;
- P_h*, the distribution of states at time* h*, which is our forecast for* $X(h)$*;*
- *the limit of* $p_{ij}^{(h)}$ *and* P_h *as* $h \to \infty$*, which is our long-term forecast.*

Indeed, when making forecasts for many transitions ahead, computations will become rather lengthy, and thus, it will be more efficient to take the limit.

Notation			
	p_{ij}	$=$	$P\{X(t+1)=j \mid X(t)=i\}$, transition probability
	$p_{ij}^{(h)}$	$=$	$P\{X(t+h)=j \mid X(t)=i\}$, h-step transition probability
	$P_t(x)$	$=$	$P\{X(t)=x\}$, distribution of $X(t)$, distribution of states at time t
	$P_0(x)$	$=$	$P\{X(0)=x\}$, initial distribution

Example 6.7 (Weather forecasts). In some town, each day is either sunny or rainy. A sunny day is followed by another sunny day with probability 0.7, whereas a rainy day is followed by a sunny day with probability 0.4.

It rains on Monday. Make forecasts for Tuesday, Wednesday, and Thursday.

Solution. Weather conditions in this problem represent a homogeneous Markov chain with 2 states: state 1 = "sunny" and state 2 = "rainy." Transition probabilities are:

$$p_{11}=0.7,\ p_{12}=0.3,\ p_{21}=0.4,\ p_{22}=0.6,$$

where p_{12} and p_{22} were computed by the complement rule.

If it rains on Monday, then Tuesday is sunny with probability $p_{21}=0.4$ (making a transition from a rainy to a sunny day), and Tuesday is rainy with probability $p_{22}=0.6$. We can predict a 60% chance of rain.

Wednesday forecast requires 2-step transition probabilities, making one transition from Monday to Tuesday, $X(0)$ to $X(1)$, and another one from Tuesday to Wednesday, $X(1)$ to $X(2)$. We'll have to condition on the weather situation on Tuesday and use the Law of Total Probability from p. 31,

$$\begin{aligned}
p_{21}^{(2)} &= \boldsymbol{P}\{\text{ Wednesday is sunny } \mid \text{ Monday is rainy }\} \\
&= \sum_{i=1}^{2} \boldsymbol{P}\{X(1)=i \mid X(0)=2\}\,\boldsymbol{P}\{X(2)=1 \mid X(1)=i\} \\
&= \boldsymbol{P}\{X(1)=1 \mid X(0)=2\}\,\boldsymbol{P}\{X(2)=1 \mid X(1)=1\} \\
&\quad +\boldsymbol{P}\{X(1)=2 \mid X(0)=2\}\,\boldsymbol{P}\{X(2)=1 \mid X(1)=2\} \\
&= p_{21}p_{11}+p_{22}p_{21}=(0.4)(0.7)+(0.6)(0.4)=0.52.
\end{aligned}$$

By the Complement Rule, $p_{22}^{(2)}=0.48$, and thus, we predict a 52% chance of sun and a 48% chance of rain on Wednesday.

For the Thursday forecast, we need to compute 3-step transition probabilities $p_{ij}^{(3)}$ because it takes 3 transitions to move from Monday to Thursday. We have to use the Law of Total Probability conditioning on *both* Tuesday and Wednesday. For example, going from rainy Monday to sunny Thursday means going from rainy Monday to either rainy or sunny Tuesday, then to either rainy or sunny Wednesday, and finally, to sunny Thursday,

$$p_{21}^{(3)}=\sum_{i=1}^{2}\sum_{j=1}^{2} p_{2i}p_{ij}p_{j1}.$$

This corresponds to a sequence of states $2 \to i \to j \to 1$. However, we have already computed 2-step transition probabilities $p_{21}^{(2)}$ and $p_{22}^{(2)}$, describing transition from Monday to Wednesday. It remains to add one transition to Thursday, hence,

$$p_{21}^{(3)} = p_{21}^{(2)} p_{11} + p_{22}^{(2)} p_{21} = (0.52)(0.7) + (0.48)(0.4) = 0.556.$$

So, we predict a 55.6% chance of sun on Thursday and a 44.4% chance of rain.

$\Diamond$

The following *transition diagram* (Figure 6.2) reflects the behavior of this Markov chain. Arrows represent all possible one-step transitions, along with the corresponding probabilities. Check this diagram against the transition probabilities stated in Example 6.7. To obtain, say, a 3-step transition probability $p_{21}^{(3)}$, find all 3-arrow paths from state 2 "rainy" to state 1 "sunny." Multiply probabilities along each path and add over all 3-step paths.

FIGURE 6.2: *Transition diagram for the Markov chain in Example 6.7.*

Example 6.8 (Weather, continued). Suppose now that it does not rain yet, but meteorologists predict an 80% chance of rain on Monday. How does this affect our forecasts?

In Example 6.7, we have computed forecasts under the condition of rain on Monday. Now, a sunny Monday (state 1) is also possible. Therefore, in addition to probabilities $p_{2j}^{(h)}$ we also need to compute $p_{1j}^{(h)}$ (say, using the transition diagram, see Figure 6.2),

$$\begin{aligned} p_{11}^{(2)} &= (0.7)(0.7) + (0.3)(0.4) = 0.61, \\ p_{11}^{(3)} &= (0.7)^3 + (0.7)(0.3)(0.4) + (0.3)(0.4)(0.7) + (0.3)(0.6)(0.4) = 0.583. \end{aligned}$$

The initial distribution $P_0(x)$ is given as

$$P_0(1) = \boldsymbol{P}\{\text{ sunny Monday }\} = 0.2, \quad P_0(2) = \boldsymbol{P}\{\text{ rainy Monday }\} = 0.8.$$

Then, for each forecast, we use the Law of Total Probability, conditioning on the weather on Monday,

$$\begin{aligned} P_1(1) &= P\{X(1) = 1\} = P_0(1)p_{11} + P_0(2)p_{21} = 0.46 && \text{for Tuesday} \\ P_2(1) &= P\{X(2) = 1\} = P_0(1)p_{11}^{(2)} + P_0(2)p_{21}^{(2)} = 0.538 && \text{for Wednesday} \\ P_3(1) &= P\{X(3) = 1\} = P_0(1)p_{11}^{(3)} + P_0(2)p_{21}^{(3)} = 0.5614 && \text{for Thursday} \end{aligned}$$

These are probabilities of a sunny day (state 1), respectively, on Tuesday, Wednesday, and Thursday. Then, the chance of rain (state 2) on these days is $P_1(2) = 0.54$, $P_2(2) = 0.462$, and $P_3(2) = 0.4386$. $\Diamond$

Noticeably, more remote forecasts require more lengthy computations. For a t-day ahead forecast, we have to account for all t-step paths on diagram Figure 6.2. Or, we use the Law of Total Probability, conditioning on *all* the intermediate states $X(1), X(2), \ldots, X(t-1)$.

To simplify the task, we shall employ *matrices*. If you are not closely familiar with basic matrix operations, refer to Section 12.4 in the Appendix.