Chapter 1
Stochastic Processes

Compiled and adapted by Jesús Cid-Sueiro, Feb. 2023.

Contents

This document introduces the concept of stochastic process and analyzes some important examples.

1.1 Introduction

A stochastic process (SP) is a collection of random variables $\{X_n\}$ indexed by time, n. Depending on the nature of the time variable, we will consider different types of SP:

- **Discrete-time** SP: when the time variable is a real number ($n \in \mathbb{R}$)
- **Continuous-time** SP: when the time variable is integer ($n \in \mathbb{Z}$).

This chapter discusses discrete-time processes only. Thus, under contrary stated, all SP will be assumed to be discrete.

Depending on the range of the time variable, two types of SP will be considered:

- **One-side** SP: $\{X_n, n \geq 0\}$. Note that a one-side SP is a sequence of random variables.
- **Two-side** SP: $\{X_n, n \in \mathbb{Z}\}$, the SP is defined for all integer times.

Example 1 Consider the one-side stochastic process X_n such that, for all n, $X_n \in \{-n, n\}$ with $P\{X_n = n\} = P\{X_n = -n\}$ and all variables in the collection are mutually independent.

A realization of all random variables in an SP defines a signal. For this reason, we can define stochastic processes in the following alternative but equivalent form:

Definition 1 (Stochastic Process) A discrete-time stochastic process is a probability measure over a space of (one-side or two-side) discrete-time signals.

Example 2 Consider the one-side stochastic process X_n such that

$$X_n = n, \qquad \text{for all } n \geq 0 \tag{1.1}$$

or

$$X_n = -n, \qquad \text{for all } n \geq 0 \tag{1.2}$$

with equal probabilities. Note that, the Universe set of this SP if finite, containing only two signals.

1.1.1 Characterization of a stochastic process

Note that the random variables defining a SP can be continuous or discrete. In general, a SP is completely characterized when we can determine the joint distribution $P(X_{n_0}, X_{n_1}, \ldots, X_{n_{M-1}})$ for any subset of random variables from the process. However, for many SP, the computation of these distributions is unfeasible. We will pay attention to processes satisfying different properties that simplify the characterization.

1.2 Independent processes

Definition 2 (Independent process) A stochastic process X_n is statistically independent (or, simply, independent) if all its random variables are statistically independent, that is, for any $M > 0$ and any subset of M random variables $X_{n_0}, X_{n_1}, \ldots, X_{n_{M-1}}$ with $n_0 < n_1, \ldots < n_{M-1}$ we have

$$P(X_{n_0}, X_{n_1}, \ldots, X_{n_{M-1}}) = \prod_{k=0}^{M-1} P(X_{n_k}) \tag{1.3}$$

Note: A quick explanation on the mathematical notation. Eq. (1.3) should be interpreted as: the joint distribution of the set of M random variables $X_{n_0}, X_{n_1}, \ldots, X_{n_{M-1}}$ is the product of the marginal distribution of each variable. Thus, for instance, the cumulative distributions gets factorized:

$$F_{X_{n_0}, X_{n_1}, \ldots, X_{n_{M-1}}}(x_{n_0}, x_{n_1}, \ldots, x_{n_{M-1}}) = \prod_{k=0}^{M-1} F_{X_{n_k}}(x_{n_k}) \tag{1.4}$$

for any $x_{n_0}, x_{n_1}, \ldots, x_{n_{M-1}}$ in the respective domains of the variables. If all the random variables in the process are continuous, Eq. (1.4) is equivalent to the factorization of the density functions

$$p_{X_{n_0}, X_{n_1}, \ldots, X_{n_{M-1}}}(x_{n_0}, x_{n_1}, \ldots, x_{n_{M-1}}) = \prod_{k=0}^{M-1} p_{X_{n_k}}(x_{n_k}) \tag{1.5}$$

Also, if the random variables are discrete, this is equivalent to the factorization of the probability mass functions

$$P_{X_{n_0}, X_{n_1}, \ldots, X_{n_{M-1}}}(x_{n_0}, x_{n_1}, \ldots, x_{n_{M-1}}) = \prod_{k=0}^{M-1} P_{X_{n_k}}(x_{n_k}) \tag{1.6}$$

In general, we will use expressions like (1.3) instead of (1.4), (1.5), or (1.6) to state some relations between distributions, because it is simpler and encompasses all of them.

A particularly interesting case arises when all variables in the SP follow the same distribution.

Definition 3 (IID process) A stochastic process X_n is independent and identically distributed, or IID, if it is independent and all variables in the process follow the same distribution, i.e. $P(X_n) = P(X_m)$ (which implies, for instance, that $F_{X_n}(x) = F_{X_m}(x)$, for all x and any n, m in the time domain of the process).

One of the simplest IID processes is the Bernoulli process.

Example 3 A Bernoulli process X_n with parameter p, denoted as $\mathscr{B}(p)$ is an IID process given by Bernoulli random variables:

$$P_{X_n}(x) = p^x(1-p)^{1-x}, \qquad x \in \{0,1\} \tag{1.7}$$

(Note that Eq. (1.7) is equivalent to claim that $P\{X_n = 1\} = p$ and $P\{X_n = 0\} = 1-p$). Bernoulli and other IID processes are commonly used as building blocks of other, more complex, processes.

Example 4 Figure 1.1 shows 3 realizations of the independent stochastic process X_n, where each sample of the process en follow a exponential distribution with parameter $\lambda_n = 1/(n+1)$.

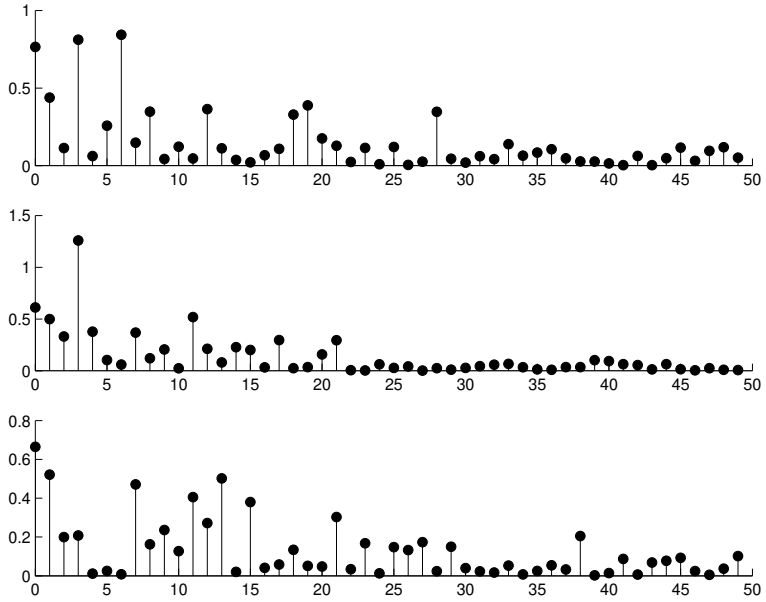

Fig. 1.1 3 realizations of process X_n described in example 4

1.3 Random walks

Definition 4 (Simple random walk (SRW)) Let $Y_n \in \{-1,1\}$ be a one-sided IID process with $P\{Y_n = 1\} = \frac{1}{2}$. A simple random walk is defined by $X_0 = 0$ and

$$X_n = \sum_{k=1}^{n} Y_k, \qquad n > 0 \tag{1.8}$$

Note that we can express the random variables in a SRW recursively as

$$X_n = X_{n-1} + Y_n, \qquad n > 0 \tag{1.9}$$

Thus, a SRW is not an independent process. It is not an identically distributed process, either, since the range of the random variables in the process grows with n.

$$-n \leq X_n \leq n \tag{1.10}$$

Fig. 1.2 shows several realizations of the SRW. Note that the realizations are quite far from the bounds. This is because the probability that X_n is close to n or $-n$ vanishes for large n.

Fig. 1.2 20 realizations of a simple random walk. The straight lines represent the boundaries of the region containing all possible trajectories. The shaded area covers two standard deviations: all trajectories tend to be inside the gray area around 95% of the time.

The mean of the process is given by

$$\mathbb{E}\{X_n\} = \sum_{k=0}^{n} \mathbb{E}\{Y_k\} = 0 \tag{1.11}$$

and, since Y_n is IID, the variance is given by

$$\text{var}\{X_n\} = \sum_{k=1}^{n} \text{var}\{Y_k\} = n \tag{1.12}$$

Therefore, the standard deviation is $\sqrt{n}$ and we can expect from the process to lie between the limits $-\sqrt{n}$ and $\sqrt{n}$ most of the time. This can be expressed a bit more precisely, by noting that, since X_n is the sum of n independent random variable with zero-mean and bounded variance, and as a consequence of the central limit theorem, X_n converges in distribution to a zero-mean Gaussian random variable with variance n. This implies that, after large n, we can expect that $|X_n| < 2\sqrt{n}$ approximately 95% of time.

1.3.1 Properties of random walks

- **Independent increment** For all $0 \leq n_0 \leq n_1 \leq n_2 \leq n_3$, the random variables $X_{n_1} - X_{n_0}$ and $X_{n_3} - X_{n_2}$ are mutually independent
- **Time invariance** For all $h \geq 1$, the SP $Z_n = X_{n+h} - X_h$ is a simple random walk.

1.3.2 Stopping times

The SRW can be used as a model of the total winning and losses of a fair game. Assume that we play a game based on repeated trials of flipping a coin. Every time the result is 'heads', we win 1 point while we loose 1 point otherwise. By taking $Y_n = 1$ if the n-th toss is 'heads', process X_n accounts for the total score after n trials.

Since the SRW is zero mean, we cannot expect a positive gain after an arbitrary number of steps. One may wonder if there is a way to ensure a positive score by choosing the appropriate stopping time. Let's assume we set an upper and lower limits, $U > 0$ and $-L < 0$, respectively, such that, when $X_n = U$ or $X_n = -L$, we stop betting. By choosing L, we set the maximum loss we can tolerate, by choosing U, we set the gains that will make us stop playing.

We can call stopping time, N to the first time where X_N reaches U or $-L$. Note that N is a random variable, because it is a function of the random variables in the SP:

$$N = \min\{n \mid X_n = U \text{ or } X_n = -L\} \tag{1.13}$$

Is there a way to select U and L in order to expect a positive gain? To answer this question, let us compute the probability of stopping at level U, that is $P\{X_N = U | X_0 = 0\}$. Thus, defining function

$$f(i) = P\{X_N = U | X_0 = i\}$$

our goal is to compute $f(0)$.

Applying the total probability theorem,

$$\begin{aligned} f(i) =& P\{X_N = U | X_0 = i\} \\ =& P\{X_1 = i+1 | X_0 = i\} P\{X_N = U | X_0 = i, X_1 = i+1\} \\ &+ P\{X_1 = i-1 | X_0 = i\} P\{X_N = U | X_0 = i, X_1 = i-1\} \\ =& \frac{1}{2} P\{X_N = U | X_1 = i+1\} + \frac{1}{2} P\{X_N = U | X_1 = i-1\} \\ =& \frac{1}{2} f(i+1) + \frac{1}{2} f(i-1) \end{aligned} \tag{1.14}$$

Noting, also, that

$$f(U) = 1,$$
$$f(-L) = 0,$$

if we let $f(-L+1) = \alpha$ then it follows that $f(1-L+r) = \alpha r$ for all $r \leq U+L$. Therefore, $\alpha = U+L$, and it follows that

$$f(0) = P\{X_N = U | X_0 = 0\} = \frac{L}{U+L}$$

Thus, the gambler can increase the probability of scoring U points by increasing L. However, this implies a higher risk to suffer large temporal losses. The expected gain becomes:

$$\mathbb{E}\{X_N\} = Uf(0) - L(1-f(0)) = 0$$

Thus, the is no way to set the bounds U and L in order to expect a positive gain.

1.4 Markov Chains

One important property of the simple random walk is that the effect of the past on the future is summarized only by the current state, rather than the whole history. In other words, the distribution of X_{n+1} depends only on the value of X_n and not on the whole set of values of $X_0, X_1, \ldots, X_n$. A stochastic process with such property is called a **Markov chain**.

Definition 5 (Markov property) Let X_n be a discrete-time SP that, for each n, takes value in some discrete set $\mathscr{S}$, which we call the **state space**. We say that the SP has the Markov property if

$$P\{X_{n+1} = i | X_n, X_{n-1}, \ldots, X_0\} = P\{X_{n+1} = i | X_n\} \tag{1.15}$$

for all n in the time-domain of the process and for all $i \in \mathscr{S}$.

Definition 6 (Markov chain)
Let X_n be a discrete-time SP. We say that the SP is a Markov chain (MC) if it satisfies the Markov property.

We will say that an MC is **finite** if $\mathscr{S}$ is a finite set. In this case, we usually take the elements of $\mathscr{S}$ as consecutive integers. For instance, for a state space with m elements, we will take $\mathscr{S} = \{0, 1, \ldots, m-1\}$ or $\mathscr{S} = \{1, 2, \ldots, m\}$ under our convenience.

Note that the SRW is an MC, but it is not finite.

Any finite MC can be described in terms of the transition probabilities

$$p_{ij} = P\{X_{n+1} = j | X_n = i\} \tag{1.16}$$

All the elements of a MC model can be encoded in a transition probability matrix

$$\mathbf{P} = \begin{pmatrix} p_{00} & p_{01} & \cdots & p_{0,m-1} \\ p_{10} & p_{11} & \cdots & p_{1,m-1} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m-1,0} & p_{m-1,1} & \cdots & p_{m-1,m-1} \end{pmatrix} \tag{1.17}$$

Using (1.16), it is easy to see that

$$\sum_{j \in \mathscr{S}} p_{ij} = 1 \tag{1.18}$$

thus, all rows in $\mathbf{P}$ sum up to one. For this reason, we say that $\mathbf{P}$ is a right-stochastic matrix. In matrix form, this property can be expressed as

$$\mathbf{P}\mathbb{1}_m = \mathbb{1}_m \tag{1.19}$$

where $\mathbb{1}_m$ is an all-ones vector with dimension m.

Note that, in general, the transition probabilities p_{ij} (and, thus, the transition matrix) may change with time. However, the case of constant probabilities is particularly interesting.

Definition 7 (Homogeneous Markov Chain) A markov chain X_n is homogeneous if the transition probabilities do not depend on time, that is.

$$P\{X_{n+1} = j | X_n = i\} = P\{X_{m+1} = j | X_m = i\}, \qquad i \in \mathscr{S}, \quad j \in \mathscr{S} \tag{1.20}$$

for any m and n in the time domain of the process.

Example 5 (Two-state machine) A machine can be either working or broken on a given day. If it is working, it will break down in the next day with probability 0.01, and will continue working with probability 0.99. If it breaks down on a given day, it will be repaired and be working in the next day with probability 0.8, and will continue to be broken down with probability 0.2. We can model this machine by a Markov chain with two states: working, and broken down. The transition probability matrix is given by

$$\begin{pmatrix} 0.99 & 0.8 \\ 0.01 & 0.2 \end{pmatrix}$$

Example 6 (SRW) A simple random walk is an example of an MC. However, there is no transition probability matrix associated with the simple random walk since the sample space is infinite.

1.4.1 Transition graph

The transition probability matrix of a given homogeneous MC can be represented by a directed graph, whose nodes are the states, and and edges from node i to node j represents a transition from state i to state j with nonzero probability. The weight of the edge is the transition probability.

Example 7 The transition graph of an MC given by state space $\mathscr{S} = \{0,1,2\}$ and transition probability matrix

$$\mathbf{P} = \begin{pmatrix} 0.4 & 0.2 & 0.4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{1.21}$$

is shown in Figure 1.3.

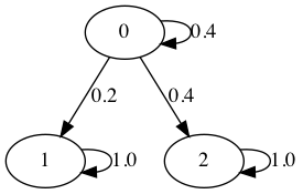

Fig. 1.3 The transition graph for the markov chain given by (1.21)

1.4.2 State probabilities

The transition probability matrix is useful to compute any probabilities about the state of the process at any given time. For instance, the n-th step transition probabilities, defined as $r_{ij}(n) = P\{X_n = j | X_0 = i\}$ can be computed recursively, by applying the total probability theorem, as

$$r_{ij}(n) = \sum_{k \in \mathscr{S}} P\{X_n = j | X_{n-1} = k, X_0 = i\} P\{X_{n-1} = k | X_0 = i\} = \sum_{k \in \mathscr{S}} r_{ik}(n-1) p_{kj}, \qquad n > 1 \tag{1.22}$$

Note that $r_{ij}(1) = p_{ij}$.

Defining the n-step transition probability matrix $\mathbf{R}_n$ as the matrix with components $r_{ij}(n)$, the recurrent relation (1.22) can be written in matrix form as

$$\mathbf{R}_n = \mathbf{R}_{n-1} \mathbf{P} \tag{1.23}$$

and, noting that $\mathbf{R}_1 = \mathbf{P}$, we get

$$\mathbf{R}_n = \mathbf{P}^n \tag{1.24}$$

Using these relations we can compute the state probabilities at any given time, given the initial state or the initial state probabilities.

For instance, let $\mathbf{q}_n$ be the state probability vector at time n, that is

$$q_{n,i} = P\{X_n = i\}, \quad n \geq 0, \quad i \in \mathscr{S} \tag{1.25}$$

then we can express the state probability vector at time n as a function of the initial state probabilities using (1.24) as

$$\begin{aligned} q_{n,j} = P\{X_n = j\} &= \sum_{i \in \mathscr{S}} P\{X_n = j | X_0 = i\} P\{X_0 = i\} \\ &= \sum_{i \in \mathscr{S}} r_{ij}(n) P\{X_0 = i\} \end{aligned} \tag{1.26}$$

which, in vector form, can be computed as

$$\mathbf{q}_n = \mathbf{R}_n^\intercal \mathbf{q}_0 = (\mathbf{P}^\intercal)^n \mathbf{q}_0 \tag{1.27}$$

1.4.3 Stationary distribution

Definition 8 (Stationary distribution) A stationary distribution of an MC is a probability distribution over the state space $\mathscr{S}$ (where $P\{X_0 = j\} = \pi_j$) such that

$$\mathbf{P}^\intercal \boldsymbol{\pi} = \boldsymbol{\pi} \tag{1.28}$$

Example 8 Let $\mathscr{S} = \mathbb{Z}_m$ and $X_0 = 0$. Consider the MC X_n such that $X_{n+1} = (X_n + 1) \mod m$ with probability $\frac{1}{2}$ and $X_{n+1} = (X_n - 1) \mod m$ with probability $\frac{1}{2}$.

The stationary distribution of this MC is $\pi_i = \frac{1}{m}$.

Note that the vector $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_m)^\intercal$ is an eigenvector of $\mathbf{P}^\intercal$ with eigenvalue 1. Hence the following theorem can be deduced from the Perron-Frobenius theorem.

Theorem 1 *If $p_{ij} > 0$ for all $i, j \in \mathscr{S}$, then there exists a unique stationary distribution of the system.*

Moreover, if $\mathbf{r}_i(n)$ is the state conditional probability vector with components $r_{ij}(n)$, $j \in \mathscr{S}$ (i.e., the values in the i-th row of $\mathbf{R}_n$),

$$\lim_{n\to\infty} \mathbf{r}_i(n) = \boldsymbol{\pi}, \qquad \forall i, j \in \mathscr{S} \tag{1.29}$$

A corresponding theorem is not true if we consider infinite state spaces.

Finite Markov chains can be shown to have at least one stationary distribution. However, if at least one of the transition probabilities is zero, the behavior of the process may depend on the topology of the transition graph:

- The stationary distribution may be not unique.
- The limiting distributions $\lim_{n\to\infty} \mathbf{r}_i(n)$ may depend on i.

Example 9 The MC in example in 7 has at least two limiting distributions: $\boldsymbol{\pi}_A = (0,1,0)^\intercal$ and $\boldsymbol{\pi}_B = (0,0,1)^\intercal$. Also, any convex combination of them $\boldsymbol{\pi} = \alpha\boldsymbol{\pi}_A + (1-\alpha)\boldsymbol{\pi}_B$, with $0 \leq \alpha \leq 1$ is a stationary distribution.

The limiting distributions depend on the initial state: is is easy to verify that

$$\lim_{n\to\infty} \mathbf{r}_i(n) = (0,0,1) \tag{1.30}$$

$$\lim_{n\to\infty} \mathbf{r}_i(n) = (0,1,0) \tag{1.31}$$

$$\lim_{n\to\infty} \mathbf{r}_i(n) = \left(0, \frac{1}{3}, \frac{2}{3}\right) \tag{1.32}$$

1.4.4 Computing the stationary distribution

In general, the stationary distribution are all the solution of the set of equations (1.28) and the probabilistic constraint $\mathbb{1}^\intercal \mathbf{p} = 1$, that is

$$\begin{pmatrix} \mathbf{P}^\intercal - \mathbf{I} \\ \mathbb{1}^\intercal \end{pmatrix} \boldsymbol{\pi} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \tag{1.33}$$

The system of linear equations given by (1.33) is consistent (it has at least one solution) but overdetermined, because the left matrix has $m+1$ rows but only m columns. In fact, using (1.19), it is easy to see that

$$\mathbb{1}_m^\intercal (\mathbf{P}^\intercal - \mathbf{I}) = \mathbf{0} \tag{1.34}$$

so that the rank of $\mathbf{P}^\intercal - \mathbf{I}$ is at most $m-1$. Thus, we can remove one of the rows in the matrix (any one of them) to solve the system.

References

1. Kempthorne, Peter, Choongbum Lee, Vasily Strela, and Jake Xia. "Topics in mathematics with applications in finance." Massachusetts Institute of Technology: MIT OpenCouseWare), https://ocw.mit.edu/courses/18-s096-topics-in-mathematics-with-applications-in-finance-fall-2013/ (Accessed Feb, 2023). License: Creative Commons BY-NC-SA.
2. Artés-Rodríguez, A., Pérez-González, F., Cid-Sueiro, J., López-Valcarce, R., Mosquera-Nartallo, C., Pérez-Cruz, F. (2012). Comunicaciones digitales (http://www.tsc.uc3m.es/~antonio/libro_comunicaciones).