13.3 Stochastic processes

So far we have talked about random variables which can be pretty much any object but are only observed at one time. A stochastic process is a collection of random variables that describes the evolution of a system that is subject to randomness. A stochastic process could, for example, describe the position of a particle that is being buffeted by other particles, the state of a genetic sequence that is subject to copying with mutation, or the shape and size of a land mass that is subject to geological forces.

Mathematically, we consider a random process X as the set of random variables $\{X_t : t \in T\}$ where T is some index set, such as discrete time ($T = 0, 1, 2, \ldots$) or continuous time ($T = [0, \infty)$).

We will try to get an understanding of these process by studying a few examples.

13.3.1 Random walk

One of the simplest stochastic processes is known as the simple symmetric random walk, or drunkard's walk. Imagine a person leaves the pub so drunk that their method of getting home consists of taking random steps, with probability 0.5 the step is in the direction of home with probability 0.5 it is in the other opposite direction. We can model the drunk's position after the ith step as a random variable X_i where $X_0 = 0$ (that is, the pub is the origin). Then $X_{i+1} = X_i + S_i$ where

$$S_i = \begin{cases} +1 & \text{with probability } 1/2 \\ -1 & \text{with probability } 1/2. \end{cases}$$

with is the direction of the ith step. Equivalently, $X_i = X_0 + \sum_{j=0}^{i-1} S_j$. The process, while amenable to analytic techniques, is extremely simple to simulate: we just need to be able to simulate Bernoulli random variables.

The random walk has many variations: instead of looking at a symmetric walk, consider $S_i = 1$ with probability p; we can consider the random walk in higher dimensions, choosing from 2^d possible directions in d dimensions; and choosing a different step size ($S_i = \pm c$, say).

The process has some very nice, and often surprising, properties. For example, the simple symmetric random walk crosses every point an infinite number of times (this is known as Gambler's ruin, as if X models the amount a gambler is winning when betting \$1 on toss of a coin, the gambler will certainly eventually lose all their money if they play for long enough against a casino with infinite resources).

Secondly, the random walk in d dimensions returns to the origin with probability 1 for $d = 1, 2$, but for $d \geq 3$, that probability is below 1 (about 0.34 in for $d = 3$, 0.19 for $d = 4$ etc.).

13.3.2 Poisson process

The Poisson process is a simple yet incredibly useful model for events that occur independently of each other and randomly in time (or space). It is commonly used to model events such as:

- Genetic mutations
- arrival times of customers
- radioactive decay
- occurrences of earthquakes
- industrial accidents
- errors in manufacturing processes (e.g. silicon wafers or cloth)

We will consider processes in time although the concepts extend readily to space (the last of the examples above could be spatial).

A Poisson process is sometimes called a *point process*. It is counts the number of events in the time interval $[0, t]$. Let $N(t)$ be the number of points in the interval, so that $N(t)$ is a counting process.

Define a *Poisson process with intensity* λ, where $\lambda > 0$ (also called the *rate*) to be a process $\mathbf{N} = \{N(t), t \geq 0\}$ that takes values in $S = \{0, 1, 2, \ldots\}$ that satisfies the following properties:

1. $N(0) = 0$ and if $s < t$ then $N(s) < N(t)$.
2. If $s < t$ then $N(t) - N(s)$ is the number of arrivals in $(s, t]$ which is independent of the number (and times) of the arrivals in $(0, s]$.
3. $$\Pr(N(t+h) = n+m|N(t) = n) = \begin{cases} \lambda h + o(h) & \text{if } m = 1 \\ o(h) & \text{if } m > 1 \\ 1 - \lambda h + o(h) & \text{if } m = 0. \end{cases}$$

Here the notation $o(h)$ indicates that, as h gets small the bit of the expression that is $o(h)$ disappears. A strict definition is that function f is $o(h)$ ('of order little oh of h') if

$$\lim_{h \to 0} \frac{f(h)}{h} = 0.$$

Examples: Check that $f(h) = h^2$ is $o(h)$ while $f(h) = h^{-\frac{1}{2}}$ is not. □

The Poisson process is related to the Poisson distribution by the fact that $N(t)$ has a Poisson distribution with parameter λt so that

$$\Pr(N(t) = k) = \frac{(\lambda t)^k}{k!} \exp(-\lambda t)$$

for $k \in \{0, 1, 2, \ldots\}$.

Now look at the times between events in a Poisson process. Let T_i denote the time of the ith event of the process and $T_0 = 0$. Then the ith *inter-arrival time*, $X_i = T_{i+1} - T_i$, is exponentially distributed with parameter λ, that is $X_i \sim \text{Exp}(\lambda)$, $i = 1, 2, 3, \ldots$.

Poisson processes have some very nice properties.

Splitting: Let $\{N(t), t \geq 0\}$ be a Poisson process with rate λ. Suppose each event is of type i, for $i \in \{1, \ldots, k\}$ with probability p_i and suppose that this is independent of other events.

If we observe just the events of type i, they form a Poisson process with rate λp_i independently of the remaining types of events.

For example, if we look at the request for different types of data in a network or arrivals of different types of customer, we get the large Poission process separated out as multiple smaller (lower rate) Poisson processes.

Merging: The converse of splitting is merging: Let $N = \{N(t), t \geq 0\}$ be a Poisson process with rate λ and $M = \{M(t), t \geq 0\}$ be a Poisson process with rate μ independent of N. Then $L = \{L(t) = N(t) + M(t), t \geq 0\}$ is a Poisson process of rate $\lambda + \mu$.

Together, these results tell us how to model multiple Poisson processes using a single large process: Suppose we have n independent Poisson processes where process i has rate λ_i. Then the merged process has events of n different types. If we observe an event in the merged process, let p_i be the probability of the event being of type i. What is p_i? According to the splitting theorem, the type i process has rate $\lambda_i = \lambda p_i$ where $\sum_i p_i = 1$ so that $\lambda = \sum_i \lambda_i$. So p_i is given by

$$p_i = \frac{\lambda_i}{\lambda} = \frac{\lambda_i}{\lambda_1 + \ldots + \lambda_n} = \frac{\text{rate of type } i}{\text{total rate}}.$$

Example: We model arrivals at a bus stop as a Poisson process. Some people arriving are students and some are office workers. Students arrive at rate λ_1, office workers arrive at rate λ_2. Merging these processes tells us the total rate of arrivals is $\lambda_1 + \lambda_2$. The probability that any given arrival is a student is $\frac{\lambda_1}{\lambda_1+\lambda_2}$ while the probability that they are an office worker is $\frac{\lambda_2}{\lambda_1+\lambda_2}$. $\square$

14 Markov chains

We think of a random process as a sequence of random variables $\{X_t : t \in T\}$ where T is an index set. T can be thought of as time. If T is discrete, the process $X(t)$ is called a *discrete time random process* while if T is continuous, $X(t)$ is called a *continuous time random process*. The random walk example is an example of a discrete time process (each time unit corresponds to a single step in the process) while the Poisson process is a continuous time process (arrivals happen at any time). We will consider only discrete time processes until further notice.

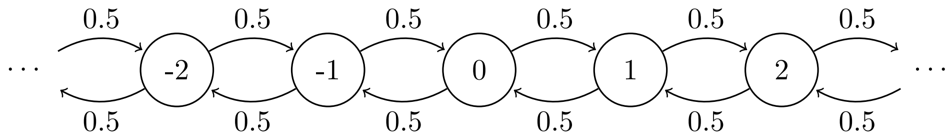
The random walk and Poisson processes described above both share an important property, known as the Markov property. Intuitively, this is the property of memorylessness in that future states depend only on the current state and not any past states. That is, to propagate the process forward, we need only be told the current state to generate the next state.

Formally, the sequence of random variables $X_1, X_2, X_3, \ldots$ is a *Markov chain* if it has the *Markov property*:

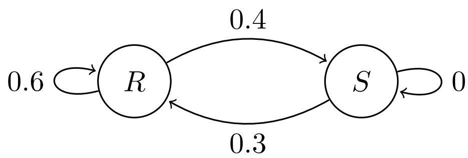
$$P(X_{n+1} = x | X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_2 = x_2, X_1 = x_1) = P(X_{n+1} = x | X_n = x_n).$$

Markov chains are commonly used to model processes that are sequential in nature and where the future state only depends on the current state. This limited dependence property is called the Markov property (after Andrey Andreyevich Markov, a Russian mathematician from the late 19th century).

Example 1: The random walk. Start at $X_0 = 0$. If X_n is the current state, $P(X_{n+1} = X_n + 1 | X_n) = 1/2 = P(X_{n+1} = X_n - 1 | X_n)$. This is a Markov chain on an infinite state space. A realisation of this chain: 0 1 0 -1 0 1 2 3 2 1 2 1 2 1 0 1



Example 2: Weather. The weather tomorrow depends on the weather today. If it is sunny today, tomorrow it is sunny with probability 0.7 and rainy otherwise. Rain clears a bit faster, so if it is rainy today, it is rainy tomorrow with probability 0.6 and sunny otherwise. This is a Markov chain with state space $\{R, S\}$ (for rainy and sunny, respectively). The following is a simulated realisation: S S S S S S R S R R R



Example 3: The following is **not** a Markov chain. Recall our random walk is called a drunkard's walk. Imagine someone occasionally helps the drunkard on the way home by carrying him 10 paces either to the left or the right. This person has limited patience, though, so will help at most 3 times. When the person has not yet reached the limit of their patience, possible transitions include $X_{n+1} = X_n \pm 10$ or $X_{n+1} = X_n \pm 1$. After the person has intervened to help 3 times, the only possible transitions are $X_{n+1} = X_n \pm 1$. So to see if this large movement is possible, we need to look back in history to see how many interventions have occurred. Thus the distribution of the next state depends on more than just the current state and the chain is not a Markov chain. □

The chain is *homogeneous* if $\Pr(X_{n+1} = j|X_n = i) = \Pr(X_1 = j|X_0 = i)$. If a chain is homogeneous, we write $P_{ij} = \Pr(X_1 = j|X_0 = i)$. The transition probabilities are normalised so that $\sum_j P_{ij} = 1$.

The matrix $P = [P_{ij}]$ is called a *stochastic matrix* as all its entries are non-negative and its rows sum to 1, so that $\sum_j P_{ij} = 1$.

Example: The transition matrix for the weather example given above is $\begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$ where rows and columns 1 and 2 are indexed by S and R, respectively. $\square$

A homogeneous Markov chain is completely defined by specifying an initial distribution $\Pr(X_0 = i)$ for X_0 and the transition probabilities X_{n+1} given X_n, P_{ij}.