Markov chains

A **Markov chain** is specified by a **state space** S and by a **transition matrix** P_{xy}, $x, y \in S$ such that

$$P_{xy} \geq 0, \qquad \sum_{y \in S} P_{xy} = 1$$

The second condition ensures that, for each fixed x, P_{xy} is a probability in y.

Two basic features:
1) the probability to be in y at time $t+1$ depends only on the position x at time t;
2) the probabilities are time independent.

In practice, the dynamics proceeds as follows:

(a) at time $t = 0$ the system is in point x_0.

(b) $P_{x_0 y}$ at fixed x_0 is a probability p_y. The two conditions above imply $p_y \geq 0$ and $\sum_{y \in S} p_y = 1$ (they should be satisfied by a probability).
Choose a point y with probability p_y and set $x_1 = y$. x_1 is the state of the system at time $t = 1$.

(c) $P_{x_1 y}$ at fixed x_1 is a probability p_y. Choose a point y with probability p_y and set $x_2 = y$. x_2 is the state of the system at time $t = 2$.
We go on analogously generating x_3, ...

Example

Let us consider two Ising spins s_1, s_2. An Ising spin is a variable which can assume the values ± 1.
The state space has dimension 4. There are four different states or configurations (s_1, s_2):

$$\begin{aligned} S_1 &= (+1,+1) \\ S_2 &= (+1,-1) \\ S_3 &= (-1,+1) \\ S_4 &= (-1,-1) \end{aligned}$$

Dynamics. The dynamics consists in choosing randomly one spin and flipping it with probability p.

We should now define the transition matrix.
Let us first compute P_{12}, i.e. the, probability of going from $(+1,+1)$ to $(+1,-1)$. We obtain $(+1,-1)$ if: (i) we choose the second spin (this occurs with probability $1/2$), (ii) we flip it (this occurs with probability p). It follows

$$P_{12} = 1/2 \times p = p/2.$$

Let us compute P_{13}, i.e. the, probability of going from $(+1,+1)$ to $(-1,+1)$. We obtain $(-1,+1)$ if: (i) we choose the first spin (this occurs with probability $1/2$), (ii) we flip it (this occurs with probability p). It follows

$$P_{13} = 1/2 \times p = p/2.$$

Let us compute P_{14}, i.e. the, probability of going from $(+1,+1)$ to $(-1,-1)$. This can never occur as we only change the direction of one spin at each time.

$$P_{14} = 0.$$

Finally, we compute P_{11}, which is defined by the condition that $\sum_y P_{1y} = 1$. Therefore:

$$P_{11} + p/2 + p/2 + 0 = 1 \qquad P_{11} = 1 - p$$

We do the same calculation for all other three points. We obtain

$$P = \begin{pmatrix} 1-p & p/2 & p/2 & 0 \\ p/2 & 1-p & 0 & p/2 \\ p/2 & 0 & 1-p & p/2 \\ 0 & p/2 & p/2 & 1-p \end{pmatrix}$$

A numerical implementation.

The configurations are stored in the vectors $s1[N+1], s2[N+1]$, where N is the number of iterations

```
s1[0] = 1; s2[0] = 1;           /* starting point */
REPEAT N times (i in [1,N])
    x = RAN()
    y = RAN()
    IF (x < 0.5)                /* first spin selected */
       if (y < p) s1[i] = -s1[i-1] /* spin 1 changes with prob p */
       else       s1[i] = s1[i-1]
       fi
       s2[i] = s2[i-1]          /* spin 2 does not change */
    ELSE                        /* second spin selected */
       if (y < p) s2[i] = -s2[i-1] /* spin 2 changes with prob p */
       else       s2[i] = s2[i-1]
       fi
       s1[i] = s1[i-1]          /* spin 1 does not change */
    FI                          /* end of IF */
END REPEAT
```

Given P_{xy}, we can define the probability of a given **history**.

The probability that the system is in x_0 at time 0, in x_1 at time 1, in x_2 at time 2, ..., in x_N at time N is simply the product of the probabilities of each transition

$$P_{x_0,x_1} P_{x_1,x_2} \ldots P_{x_{N-1},x_N}$$

We can also determine the probability of being in x_N at time N, **irrespective of the position at intermediate times**.
To compute this probability we should sum over all histories that start in x_0 and end in x_N.

Example

Let us go back to the example of the dynamics of two spins. In this system there are four states S_i. At time 0 the system is S_1 and want to determine the probability of being in S_4 at time $t = 2$.
There are four possible histories:

$$\begin{aligned}
S_1 \to S_1 \to S_4 \qquad & \text{probability } = P_{11}P_{14} \\
S_1 \to S_2 \to S_4 \qquad & \text{probability } = P_{12}P_{24} \\
S_1 \to S_3 \to S_4 \qquad & \text{probability } = P_{13}P_{34} \\
S_1 \to S_4 \to S_4 \qquad & \text{probability } = P_{14}P_{44}
\end{aligned} \tag{1}$$

The required probability is obtained by summing over all hystories:

$$P_{11}P_{14} + P_{12}P_{24} + P_{13}P_{34} + P_{14}P_{44} = \sum_x P_{1x}P_{x4}$$

We sum over the all state points. If we consider P as a matrix, the probability is $(P \cdot P)_{14} = P^2_{1,4}$, i.e., it is the element $(1,4)$ of the matrix P^2.

It is thus clear how to generalize the expression in the generic case:

$$\sum_{x_1, x_2, \ldots x_{N-1}} P_{x_0,x_1} P_{x_1,x_2} \ldots P_{x_{N-1},x_N} = P^N_{x_0,x_N}$$

The relevant family of Markov chains

Among all Markov chains we will consider processes which satisfy two conditions:

1) **Ergodicity** (in the mathematical literature it is called irreducibility). We can go from any state to any state:
For any x, y, there is $n > 0$ such that $P^n_{xy} > 0$

2) **Aperiodicity** (technical, not very interesting condition): the greatest common divisor of the set of integers n such that $P^n_{xx} > 0$ is 1.

The fundamental theorem

If P is irreducible and aperiodic:

a)

$$\lim_{n\to\infty} P^n_{xy} = \pi_y$$

with π_y nonnegative.

b) If π_x is not identically zero, it satisfies

$$\sum_x \pi_x = 1.$$

Hence π_x is a probability distribution on the state space.

c) If π_x is not identically zero, it satisfies the **stationarity condition**

$$\sum_x \pi_x P_{xy} = \pi_y.$$

Hence π_x is the **equilibrium distribution**.

The fundamental theorem

d) (**Uniqueness**) π_x is the unique probability distribution satisfying the stationarity condition.

e) (**Ergodic theorem**) Consider the process $X_0 \to X_1 \to X_2 \ldots \to X_N$ generated by P. Then

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(X_n) = \sum_x f(x)\pi_x$$

for any function $f(x)$ defined on the state space, irrespective of X_0.

Example

Again, we consider our example wth two spins, in which the transition matrix is (we set $p = 1/2$)

$$P = \begin{pmatrix} 0.5 & 0.25 & 0.25 & 0 \\ 0.25 & 0.5 & 0 & 0.25 \\ 0.25 & 0 & 0.5 & 0.25 \\ 0 & 0.25 & 0.25 & 0.5 \end{pmatrix}$$

Now, we take products

$$P^2 = P \cdot P = \begin{pmatrix} 0.375 & 0.25 & 0.25 & 0.125 \\ 0.25 & 0.375 & 0.125 & 0.25 \\ 0.25 & 0.125 & 0.375 & 0.25 \\ 0.125 & 0.25 & 0.25 & 0.375 \end{pmatrix} P^4 = P^2 \cdot P^2 = \begin{pmatrix} 0.281 & 0.25 & 0.25 & 0.219 \\ 0.25 & 0.281 & 0.219 & 0.25 \\ 0.25 & 0.219 & 0.281 & 0.25 \\ 0.219 & 0.25 & 0.25 & 0.281 \end{pmatrix}$$

$$P^8 = P^4 \cdot P^4 = \begin{pmatrix} 0.252 & 0.25 & 0.25 & 0.248 \\ 0.25 & 0.252 & 0.248 & 0.25 \\ 0.25 & 0.248 & 0.252 & 0.25 \\ 0.248 & 0.25 & 0.25 & 0.252 \end{pmatrix}$$

Example

Iterating the procedure we find that P^n converges to

$$\begin{pmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{pmatrix}$$

As expected, each row is the same we can estimate $\pi_1 = \pi_2 = \pi_3 = \pi_4 = 0.25$.

π_x is the equiibrium distribution of the Markov chain.

The stationarity condition states that π_x is a left-eigenvalue of P_{xy}. We find

$$\sum_x \pi_x P_{xy} = 1/4(P_{1x} + P_{2x} + P_{3x} + P_{4x}).$$

The quantity in parentheses is the sum of the elements of each column, which is 1:

$$\sum_x \pi_x P_{xy} = 1/4 = \pi_y$$

Static algorithms

The algorithms that generate uncorrelated data can be considered as trivial Markov chains. The transition matrix is defined as

$$P_{xy} = \pi_y$$

Note that

$$P^2_{xy} = \sum_z P_{xz} P_{zy} = \sum_z \pi_z \pi_y = (\sum_z \pi_z) \pi_y = \pi_y$$

and in general $P^N_{xy} = \pi_y$.
All Markov chain theorems are trivial for these distributions.