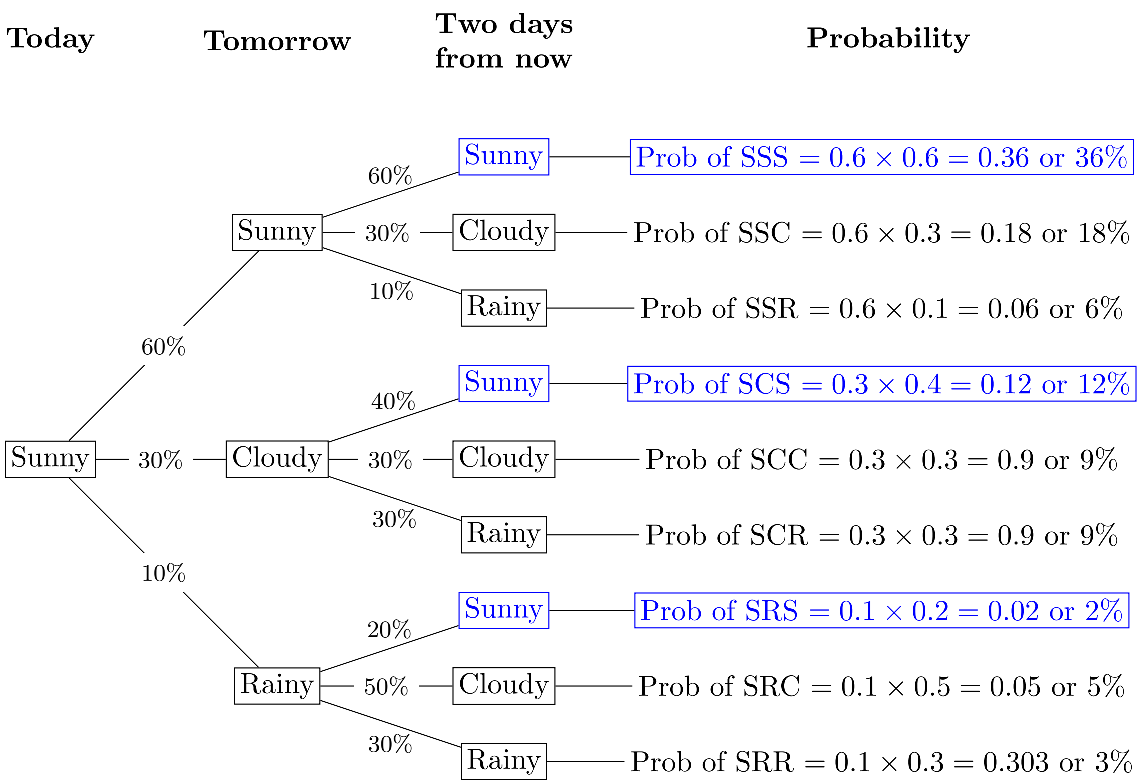
## Markov Chains

A Markov Chain is a random process describing a sequence of possible events in which the probability of each event depends **only** on the state attained in the previous event. To understand what this means, let's consider an example.

Ex 1: Suppose that on any given sunny day, the next day's weather has a 60% chance of being sunny, a 30% chance of being cloudy, and a 10% chance of being rainy. On any given cloudy day, the next day's weather has a 40% chance of being sunny, a 30% chance of being cloudy, and a 30% chance of being rainy. Lastly, On any given rainy day, there is a 20% chance of being sunny, a 50% chance of being cloudy, and a 30% chance of being rainy the following day.

So assuming it's sunny today, here's how we determine the odds that it will be sunny two days from now. We consider the possibilities for the weather from today to two days from now. It will either be SSS (three days of sunshine), SCS (sunny today, cloudy tomorrow, sunny the next day), OR SRS (sunny today, rainy tomorrow, sunny the day after).




We calculate the odds of each of these cases by just multiplying probabilities.

$SSS = 0.6 \times 0.6 = 0.36, \quad SCS = 0.3 \times 0.4 = 0.12, \quad SRS = 0.1 \times 0.2 = 0.02$

Summing these up, we get the total probability that it's sunny two days from now given that it's sunny today. $P = 0.36 + 0.12 + 0.02 = 0.5$, or 50%.

Note that this would be a different answer if it was rainy today (38% if you really want to know) because we'd have three different chains of events we'd have to calculate probabilities for (RRS, RCS, and RSS). Also, if we wanted to predict what would happen even 12 days from now, we'd actually need to compute and add up the probabilities for thousands of possible chains of events. Fortunately, linear algebra offers a way to streamline this procedure.

We begin by arranging the probabilities in columns and obtain a **transition matrix** $P$.

$$P = \begin{bmatrix} 0.6 & 0.4 & 0.2 \\ 0.3 & 0.3 & 0.5 \\ 0.1 & 0.3 & 0.3 \end{bmatrix}$$

Note: A matrix where each column sums to 1 is called a **stochastic, probability,** or **Markov** matrix.

We also need to define **state vectors** of the form $\mathbf{x} = \begin{bmatrix} x_S \\ x_C \\ x_R \end{bmatrix}$ that give the probability that it will be sunny ($x_S$), cloudy ($x_C$), or rainy ($x_R$) on a particular day. For the given example,

$$\mathbf{x^{(0)}} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \mathbf{x^{(1)}} = P\mathbf{x^{(0)}} \begin{bmatrix} 0.6 & 0.4 & 0.2 \\ 0.3 & 0.3 & 0.5 \\ 0.1 & 0.3 & 0.3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.6 \\ 0.3 \\ 0.1 \end{bmatrix}$$

$$\mathbf{x^{(2)}} = P^2\mathbf{x^{(0)}} = P\mathbf{x^{(1)}} \begin{bmatrix} 0.6 & 0.4 & 0.2 \\ 0.3 & 0.3 & 0.5 \\ 0.1 & 0.3 & 0.3 \end{bmatrix} \begin{bmatrix} 0.6 \\ 0.3 \\ 0.1 \end{bmatrix} = \begin{bmatrix} 0.50 \\ 0.32 \\ 0.18 \end{bmatrix}$$

Here, each $\mathbf{x^{(n)}}$ represents the state vector on day $n$. Note that if it's sunny today (day 0), then the probability that it is sunny today is obviously 1, and the probability that it's cloudy or rainy is 0, corresponding to the three entries in the vector. Similarly, given that it's sunny today, the probability that it's sunny tomorrow (day 1) is 0.6, the probability that it's cloudy is 0.3, and the probability that it's rainy is 0.1 accordingly. Similarly the entries $\mathbf{x^{(2)}}$ tell us the likelihood that it will be sunny, cloudy or rainy on two days from today.

Thus to determine the probabilities of the weather conditions $n$ days from today, we get the general formula

$$\mathbf{x^{(n)}} = P\mathbf{x^{(n-1)}} = P^n\mathbf{x^{(0)}}$$

So by multiplying the initial vector with powers of $P$, we can calculate $\mathbf{x^{(n)}}$. Looking at some powers of $P$, you should see a pattern of emerging. Note that the as the exponent gets bigger, the entries barely change, and the columns look more and more alike.

$$P = \begin{bmatrix} 0.6 & 0.4 & 0.2 \\ 0.3 & 0.3 & 0.5 \\ 0.1 & 0.3 & 0.3 \end{bmatrix} \quad P^2 = \begin{bmatrix} 0.5 & 0.42 & 0.38 \\ 0.32 & 0.36 & 0.36 \\ 0.18 & 0.22 & 0.26 \end{bmatrix} \quad P^5 = \begin{bmatrix} 0.44912 & 0.44656 & 0.44496 \\ 0.34144 & 0.3424 & 0.34304 \\ 0.20944 & 0.21104 & 0.212 \end{bmatrix}$$

$$P^8 = \begin{bmatrix} 0.447428 & 0.447341 & 0.447287 \\ 0.342083 & 0.342116 & 0.342136 \\ 0.21049 & 0.210543 & 0.210577 \end{bmatrix} \qquad P^{10} = \begin{bmatrix} 0.447375 & 0.447366 & 0.44736 \\ 0.342103 & 0.342106 & 0.342109 \\ 0.210522 & 0.210528 & 0.210532 \end{bmatrix}$$

**Definition**: A matrix $P$ is called **regular** if there exists some power of it with entries **strictly** greater than zero, i.e. $P^k > 0$

**Theorem**: Let $P$ is a regular transition $k \times k$ matrix. Then

$$\lim_{n\to\infty} P^n = \begin{bmatrix} q_1 & q_1 & \cdots & q_1 \\ q_2 & q_2 & \cdots & q_2 \\ \vdots & \vdots & \ddots & \vdots \\ q_k & q_k & \cdots & q_k \end{bmatrix}, \text{ where } q_1 + q_2 + \cdots + q_k = 1.$$

Furthermore, for any initial probability vector $\mathbf{x^{(0)}}$ and regular transition matrix $P$, we have

$$\lim_{n\to\infty} P^n \mathbf{x^{(0)}} = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_k \end{bmatrix}$$

This vector which we'll call $\mathbf{q}$ is called the **steady-state vector** (or stable vector) of the transition matrix $P$, and it is the **only** probability vector that satisfies the equation $P\mathbf{q} = \mathbf{q}$. We can rewrite the equation into $(I - P)\mathbf{q} = 0$ so that we're now solving for $Nul(I - P)$. We can then find our unique $\mathbf{q} \in Nul(I-P)$ since it is a probability vector and consequently its entries must add up to 1.

Let's return to our example and solve for the steady-state vector.

$$P = \begin{bmatrix} 0.6 & 0.4 & 0.2 \\ 0.3 & 0.3 & 0.5 \\ 0.1 & 0.3 & 0.3 \end{bmatrix}, \qquad \text{so } I - P = \begin{bmatrix} 0.4 & -0.4 & -0.2 \\ -0.3 & 0.7 & -0.5 \\ -0.1 & -0.3 & 0.7 \end{bmatrix}$$

Solving for the nullspace, we get that the vectors in $Nul(I - P)$ are of the form $\begin{bmatrix} 17t/8 \\ 13t/8 \\ t \end{bmatrix}$

So to obtain $\mathbf{q}$, we solve $\frac{17t}{8} + \frac{13t}{8} + t = 1$, giving us $t = \frac{4}{19}$.

Thus $\mathbf{q} = \begin{bmatrix} 17/38 \\ 13/38 \\ 4/19 \end{bmatrix} \approx \begin{bmatrix} 0.4474 \\ 0.3421 \\ 0.2105 \end{bmatrix}$

This means that looking far enough into the future, the probability that it will be sunny, cloudy, or rainy on a given day will be very close to 44.74%, 34.21%, and 21.05% respectively. In fact, if you look at the calculations of $P^n$ we did earlier, by the eighth day, these estimates are already accurate within $\frac{1}{100}th$ of a percent, which is a good enough approximation already for most practical purposes.