09 - Markov chains - 01

Master in Foundations of Data Science

Bayesian Statistics and Probabilistic Programming

Fall 2018-2019

Josep Fortiana

Wednesday, November 21, 2018

Universitat de Barcelona

09 - Markov chains - 01

09 - Markov chains - 01

Rainfall in Tel Aviv

Occurrence of wet and dry days during the rainy period, December, January, February.

Probability α of a wet day following a dry day is 0.250.

Probability β of a dry day following a wet day is 0.338.

Kuno Reuben Gabriel and J. Neumann (1962), A Markov chain model for daily rainfall occurrence at Tel Aviv Quarterly Journal of the Royal Meteorological Society 88(375):90-95.

Data for a total of 2437 days

		Today		
		Dry	Wet	Total
Yesterday	Dry	1049	350	1399
	Wet	351	687	1038

A system with two possible states

States: A_1 and A_2,

The system evolves along time.

We know:

- State at $t = 0$ is A_1, say.
- The *transition probabilities:*

$$\begin{aligned} \text{P(Stay in state } A_1) &= 1 - \alpha, \\ \text{P(Move to state } A_2) &= \alpha. \end{aligned}$$

What is the state at $t = 1$?

When $t = 1$ we no longer know the exact state, only the probabilities:

$$p_{11} = \mathrm{P}(A_1 \quad [t = 1]) = 1 - \alpha,$$

$$p_{12} = \mathrm{P}(A_2 \quad [t = 1]) = \alpha.$$

In vector notation: $\boldsymbol{p}_1 = (p_{11}, p_{12})$.

What is the state at $t = 2$?

Similarly for $t = 2$.

$$\boldsymbol{p}_2 = (\mathrm{P}(A_1\ [t=2]),\ \mathrm{P}(A_2\ [t=2])),$$

we arrange transition probabilities in a matrix:

$$\boldsymbol{P}_{1\to 2} = \begin{pmatrix} \mathrm{P}(A_1\ [t=2]|A_1\ [t=1]) & \mathrm{P}(A_2\ [t=2]|A_1\ [t=1]) \\ \mathrm{P}(A_1\ [t=2]|A_2\ [t=1]) & \mathrm{P}(A_2\ [t=2]|A_2\ [t=1]) \end{pmatrix}$$

By the total probability formula:

$$\boldsymbol{p}_2 = \boldsymbol{p}_1 \cdot \boldsymbol{P}_{1\to 2}.$$

Notation

In Markov chains literature probability vectors are written as rows.

Rows in transition matrices are probabilities, they add up to 1.

To obtain the next vector of probabilities, we multiply the current one on the left of the transition matrix.

In the Tel Aviv weather example

$$\boldsymbol{P} = \begin{pmatrix} 0.750 & 0.250 \\ 0.338 & 0.662 \end{pmatrix} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}.$$

Successive probabilities

Assuming a dry day at $t = 0$, then $\boldsymbol{p}_0 = (1, 0)$, the following days have probabilities:

$$\boldsymbol{p}_1 = \boldsymbol{p}_0 \cdot \boldsymbol{P} = (0.750, 0.250),$$

$$\boldsymbol{p}_2 = \boldsymbol{p}_1 \cdot \boldsymbol{P} = (0.647, 0.353),$$

$$\boldsymbol{p}_3 = \boldsymbol{p}_2 \cdot \boldsymbol{P} = (0.605, 0.395),$$

$$\boldsymbol{p}_4 = \boldsymbol{p}_3 \cdot \boldsymbol{P} = (0.587, 0.413),$$

Stationary distribution

This sequence converges to the stationary distribution:

$$\boldsymbol{p}_\infty = (0.5748, 0.4252).$$

Which we can check by solving the (eigen-)equation :

$$\boldsymbol{p}_\infty \cdot \boldsymbol{P} = \boldsymbol{p}_\infty,$$

Stationary distribution

Solving, for (x, y), the equation:

$$(x, y) \cdot \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} = (x, y),$$

gives:

$$y = \frac{\alpha}{\beta} \cdot x.$$

Together with the probability condition, $x + y = 1$:

$$x = \frac{\beta}{\alpha + \beta}, \qquad y = \frac{\alpha}{\alpha + \beta}.$$

Computing the whole eigen-decomposition

Sometimes it will be possible to compute the whole eigen-decomposition:

$$\boldsymbol{P} = \boldsymbol{U} \cdot \boldsymbol{D} \cdot \boldsymbol{U}^{-1},$$

where:

$$\boldsymbol{U} = \begin{pmatrix} 0.707 & -0.595 \\ 0.707 & 0.804 \end{pmatrix}, \qquad \boldsymbol{D} = \begin{pmatrix} 1 & 0 \\ 0 & 0.412 \end{pmatrix}.$$

Remark on notation

Functions for computing eigenvectors obtain right eigenvectors, here we need left eigenvectors.

Left eigenvectors of $\boldsymbol{P}$ are the *rows* in $\boldsymbol{U}^{-1}$, since:

$$\boldsymbol{U}^{-1} \cdot \boldsymbol{P} = \boldsymbol{D} \cdot \boldsymbol{U}^{-1}.$$

Remark on normalization

Usually the (right) eigenvectors, the columns in $\boldsymbol{U}$, are normalized so that their ℓ^2 norm is 1.

We need, instead, to normalize the first left eigenvector, with eigenvalue 1, the first row in $\boldsymbol{U}^{-1}$, so that the sum of its entries is 1, to be a probability.

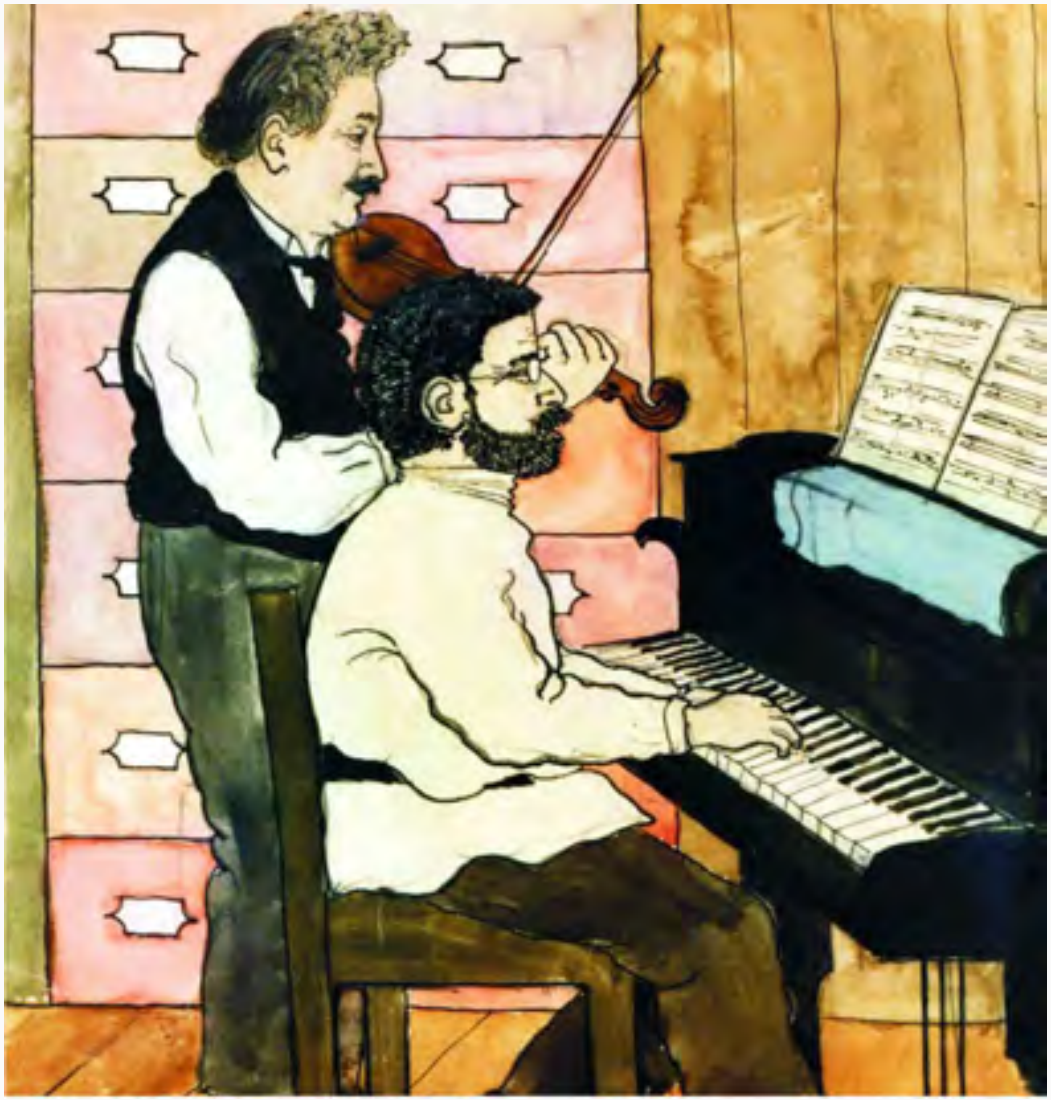
Time to attain the stationary distribution

Since $\boldsymbol{p}_n = \boldsymbol{p}_0 \cdot \boldsymbol{P}^n$, we can obtain an explicit expression for $\boldsymbol{p}_n$ from:

$$\boldsymbol{p}_n = \boldsymbol{p}_0 \cdot (\boldsymbol{U} \cdot \boldsymbol{D} \cdot \boldsymbol{U}^{-1})^n = \boldsymbol{p}_0 \cdot \boldsymbol{U} \cdot \boldsymbol{D}^n \cdot \boldsymbol{U}^{-1}.$$

Then compare $\boldsymbol{p}_n$ and $\boldsymbol{p}_\infty$.

Paul Ehrenfest



Origins of the model

How can it be that, whereas Newton's laws are time-reversible, macroscopic world is irreversible?

Modelling microscopical phenomena underlying the concept of entropy and the Second Law of Thermodynamics (entropy of a closed system in equilibrium increases).

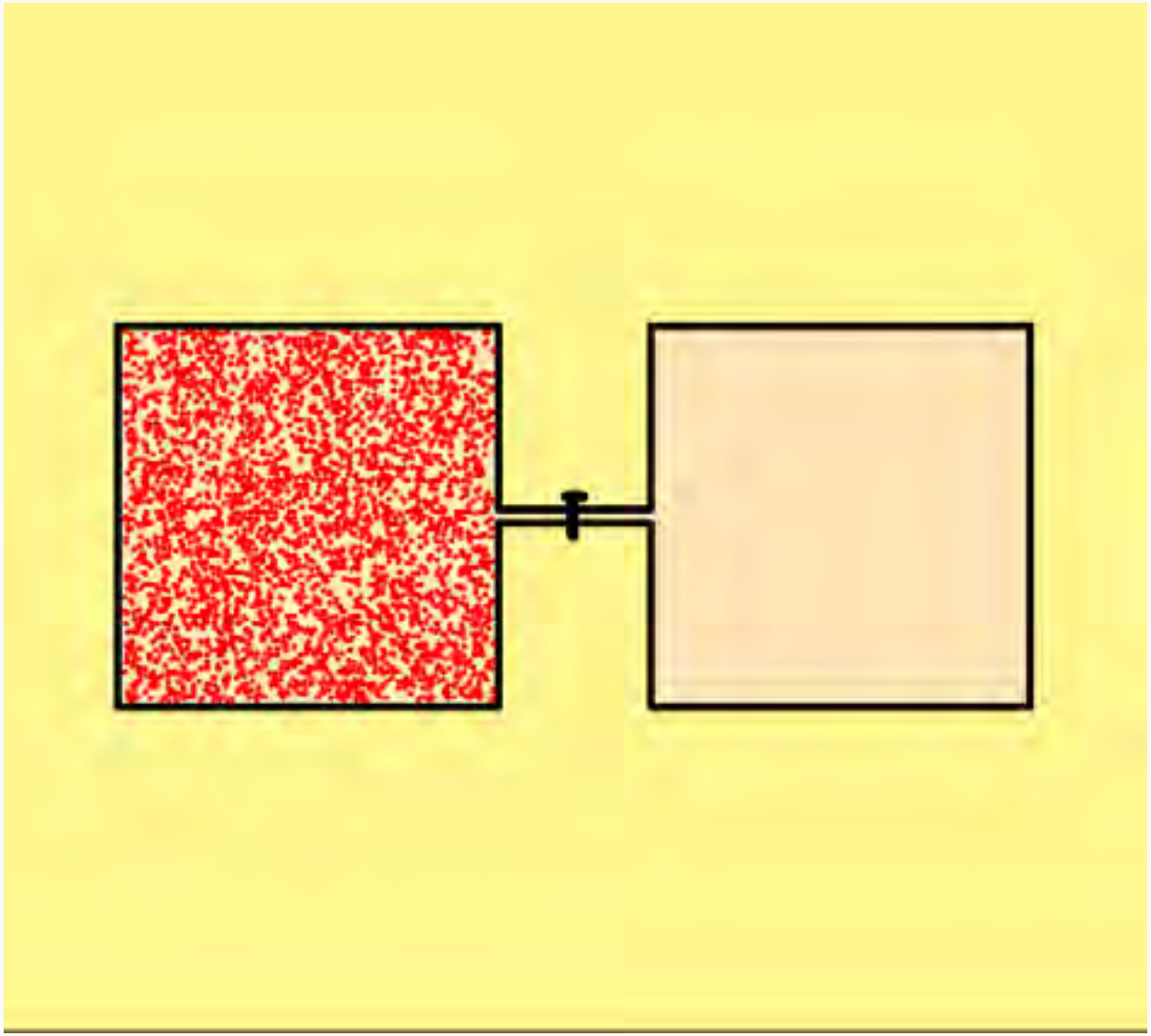
> *Paul Ehrenfest and Tatiana Ehrenfest-Afanasyeva, Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem, Physikalische Zeitschrift, vol. 8 (1907), pp. 311-314.*
> *The Conceptual Foundations of the Statistical Approach in Mechanics, article for the German Encyclopedia of Mathematical Sciences (1912).*

Setting

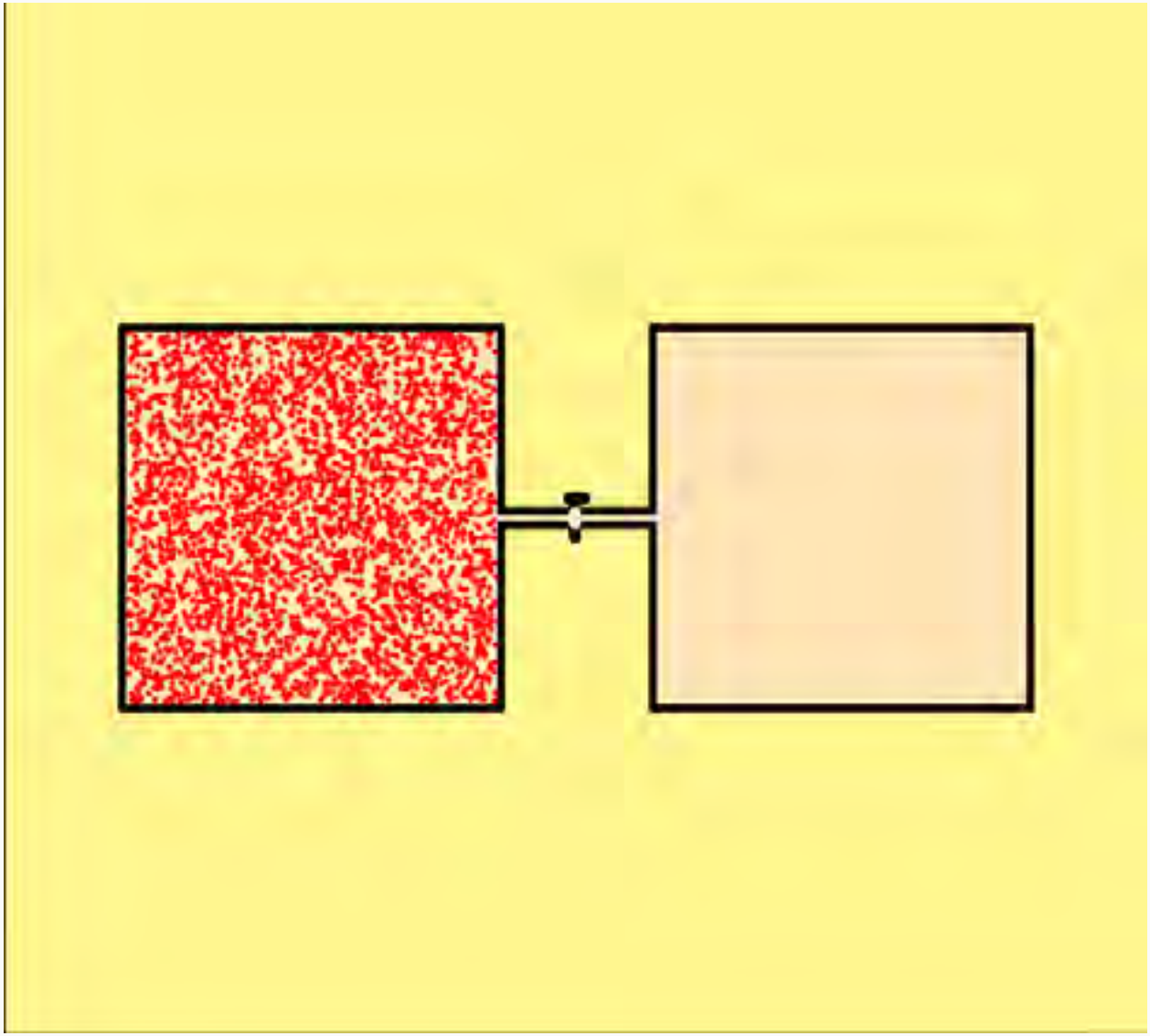
A hermetically closed container with n non-interacting gas molecules is partitioned into two separated parts of equal volume, A and B, communicated by a pipe which can be closed with a valve.

Initially A, the left-hand half container, contains all n molecules, and the valve is closed. Then we open the valve.

Initial configuration



Initial configuration, open valve



System evolution

The gas molecules in the system undergo a *difussion*, which can be modelled as follows:

On each time unit one single molecule switches sides, in such a way that probabilities of both possible transitions are proportional to the number of molecules in each half container.

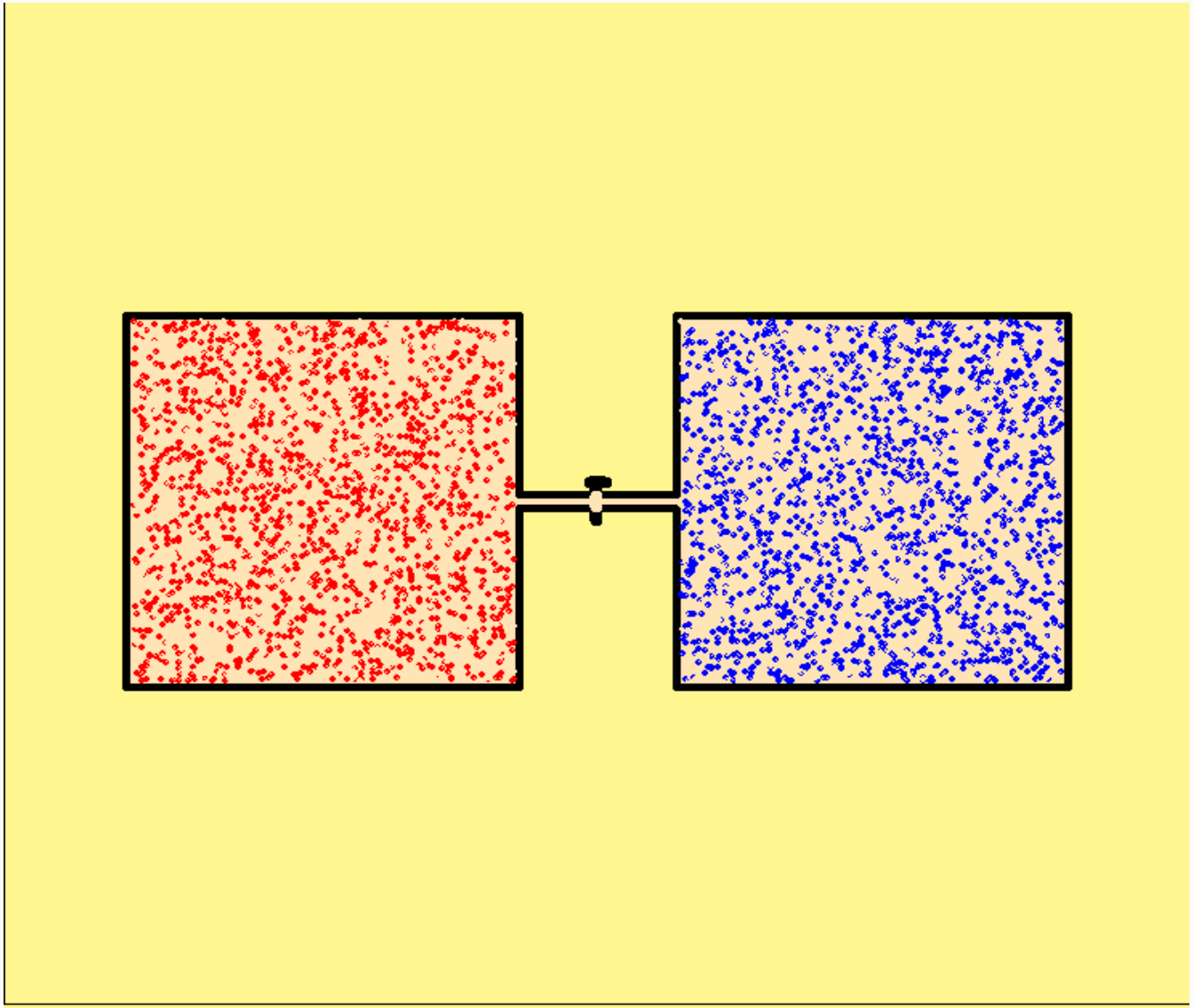
Equivalent description

Equivalently a molecule, from 1 to n, is randomly selected and brought to the opposite side.

In this way initially the number of molecules in B increases steadily, but with a decreasing speed.

When both halves hold approximately the same number of molecules, there are only small changes, fluctuating around the stationary state, half of the molecules in each side.

Long term configuration



Writing down the model

This system has $n+1$ possible states, labelled by the number of molecules in the A side.

Assume a known initial state, e.g., the number of molecules in A is $n_0 = n$ for $t = 0$.

There is a stationary distribution which we find.

Writing down the model

We compute the transition matrix, $\boldsymbol{P}_n$, an $(n+1) \times (n+1)$ matrix.

$$\boldsymbol{P}_n = (p_{ij}).$$

where

$$p_{ij} = \mathrm{P}[n_{t+1} = j | n_t = i], \quad 0 \leq i, j \leq n.$$

Writing down the model

Two non null entries in each i-th row, $i \neq 0$, $i \neq n$:

$$p_{i,\,i-1} = \mathrm{P}[n_{t+1} = i-1 | n_t = i] = \frac{i}{n},$$

$$p_{i,\,i+1} = \mathrm{P}[n_{t+1} = i+1 | n_t = i] = 1 - \frac{i}{n}.$$

Writing down the model

Boundary cases:

$$p_{01} = \mathrm{P}[n_{t+1} = 1 | n_t = 0] = 1,$$

$$p_{n,n-1} = \mathrm{P}[n_{t+1} = n-1 | n_t = n] = 1.$$

Meaning: if all molecules are in the same half container, the next displacement is a sure event.

Ehrenfest model for $n = 2$

$$\boldsymbol{P}_2 = \begin{pmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{pmatrix}.$$

Ehrenfest model for $n = 3$

$$\boldsymbol{P}_3 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Ehrenfest model for general n

$$\boldsymbol{P}_n = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & & 0 \\ 1/n & 0 & 1-1/n & 0 & & & \\ 0 & 2/n & 0 & 1-2/n & & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & & \\ & & & & 0 & 2/n & 0 \\ 0 & & & & 1-1/n & 0 & 1/n \\ 0 & 0 & \cdots & & 0 & 1 & 0 \end{pmatrix}.$$

Computing the limit probability

The limit probability vector $\boldsymbol{p}$ must satisfy:

$$\boldsymbol{p} \cdot \boldsymbol{P}_n = \boldsymbol{p},$$

that is, a ladder-like system of simultaneous equations:

$$(1/n)\, p_1 = p_0, \quad \Longrightarrow \quad p_1 = n\, p_0,$$

$$p_0 + (2/n)\, p_2 = p_1, \quad \Longrightarrow \quad p_2 = \frac{n \cdot (n-1)}{2}\, p_0,$$

$$(1 - 1/n)\, p_1 + (3/n)\, p_3 = p_2, \quad \Longrightarrow \quad p_3 = \frac{n \cdot (n-1) \cdot (n-2)}{3 \cdot 2}\, p_0$$

Limit distribution pmf

In general, we find that:

$$p_j = \binom{n}{j} p_0, \quad 0 \leq j \leq n.$$

Normalizing, as $\boldsymbol{p} = (p_0, \ldots, p_n)$ must be a probability,

$$1 = \sum_{j=0}^{n} p_j = p_0 \sum_{j=0}^{n} \binom{n}{j} = p_0\, 2^n,$$

Hence, $\boldsymbol{p}$ is a binomial pmf, $\mathrm{B}(n, 1/2)$,

$$p_j = \binom{n}{j} 2^{-n}, \quad 0 \leq j \leq n.$$

Interpreting the limit distribution

For large n, probability in the $\mathrm{B}(n, 1/2)$ is strongly concentrated around $[n/2]$.

For $n = 100$, a 99% of probability is concentrated on the 26% points around $n/2 = 50$.

For $n = 1000$ on an 8.2% of the points.

For $n = 100000$, on a 0.81%.

Thus the essential irreversibility of the experiment.

09 - Markov chains - 01

Stochastic process

A *discrete-time stochastic process* with *state space* $\mathscr{S}$ is a sequence $X = \{X_t\}_{t\in\mathbb{Z}}$ (or $X = \{X_t\}_{t\in\mathbb{Z},t\geq 0}$) of r.v. with values on $\mathscr{S}$, indexed by integers $\mathbb{Z}$ (or by non-negative integers $\mathbb{Z}_+ \equiv \mathbb{N}\cup\{0\}$).

These indexes will be thought of as *time* instants.

The state space $\mathscr{S}$ can be a finite set, or an infinite set, either countable or non-countable.

Markov chain

A *discrete-time Markov chain, with state space* $\mathscr{S}$, is a discrete-time stochastic process $X = \{X_t\}_{t\in\mathbb{Z}, t\geq 0}$ satisfying the *Markov memoryless property:*

$$\begin{aligned} \mathrm{P}(X_{t+1} = x_{t+1} | X_t = x_t, \ldots, X_1 = x_1) \\ = \mathrm{P}(X_{t+1} = x_{t+1} | X_t = x_t), \end{aligned}$$

for any $t \in \mathbb{N}$ and any $(t+1)$-tuple $x_1, \ldots, x_t, x_{t+1} \in \mathscr{S}$.

Meaning: the pdf of X_{t+1} conditional on all the past history, from 1 to t, depends only on X_t.

Homogeneous finite Markov chains

A Markov chain with a finite state space $\mathscr{S}$ is called *finite*.

A Markov chain is *homogeneous* if, for any $t \geq 1$,

$$\mathrm{P}(X_{t+1} = x_{t+1} | X_t = x_t)$$

does not depend on t.

Transition matrix

A finite homogeneous Markov chain, with $\mathscr{S} = \{s_1, \ldots, s_n\}$, has *transition matrix:*

$$\boldsymbol{P} = \big(p_{ij},\ 1 \leq i \leq n,\ 1 \leq j \leq n\big),$$

if, for any $t \geq 1$,

$$\mathrm{P}(X_{t+1} = s_j | X_t = s_i) = p_{ij},$$

the i-th row in $\boldsymbol{P}$ contains the transition probabilities from the state $\{i\}$ at time t to the n states at $t + 1$.

Stochastic matrices

In particular the sum of entries in each row of $\boldsymbol{P}$ is equal to 1. Non-negative matrices with this property are called *stochastic matrices.*

They have interesting algebraic properties. If $\boldsymbol{P}$ is an $n \times n$ stochastic matrix,

$$\boldsymbol{P} \cdot \mathbb{1} = \mathbb{1},$$

where $\mathbb{1}$ is an $n \times 1$ columns of ones.

Infinite matrices - Random walk

When $\mathscr{S}$ is countable infinite the transition matrix concept can be extended to *"infinite matrices"*.

Ordinary random walk is a homogeneous Markov chain with state space $\mathscr{S} = \mathbb{Z}$, initial state $X_0 = 0$ and (infinite) transition matrix:

$$p_{i,j} = \begin{cases} 0.5, & \text{if } j = i - 1, \\ 0.5, & \text{if } j = i + 1, \\ 0, & \text{otherwise.} \end{cases} \qquad i, j \in \mathbb{Z}$$

Initial pmf

Often *the initial state* in a Markov chain $X = \{X_t\}_{t\in\mathbb{N}}$ is specified by giving the distribution of X_1.

If the chain is finite, with $\mathscr{S} = \{s_1, \ldots, s_n\}$, the pmf of X_1 is given by a vector:

$$\boldsymbol{p}_1 \equiv \boldsymbol{p} = (p_1, \ldots, p_n),$$

where $p_i = \mathrm{P}(X_1 = s_i)$, $1 \leq i \leq n$.

pmf of other states

Thus the pmf of X_2, by the total probability formula, is:

$$\boldsymbol{p}_2 = \boldsymbol{p}_1 \cdot \boldsymbol{P}.$$

In general, for $k \geq 1$, the pmf of X_{t+k} is

$$\boldsymbol{p}_{t+k} = \boldsymbol{p}_t \cdot \boldsymbol{P}^k,$$

and the (i, j)-th entry in $\boldsymbol{P}^k$ is the probability of the transition:

$$\mathrm{P}(X_{t+k} = s_j | X_t = s_i), \quad \text{for any } t \geq 1.$$

Simple random walk with a reflecting barrier at the origin

A particle moves on the set of non-negative integers: $\mathbb{Z}_+ = 0, 1, 2, \ldots$.

When the state is $j \in \mathbb{Z}_+$, for p, q such that $p + q \leq 1$,

$$\mathrm{P}(\{j+1\}|\{j\}) = p, \quad \mathrm{P}(\{j-1\}|\{j\}) = q,$$

$$\mathrm{P}(\{j\}|\{j\}) = 1 - p - q,$$

when $j \geq 1$, and

$$\mathrm{P}(\{1\}|\{0\}) = p, \quad \mathrm{P}(\{0\}|\{0\}) = 1 - p.$$

Transition matrix

$$\boldsymbol{P} = \begin{pmatrix} 1-p & p & 0 & 0 & \cdots & & \cdots \\ q & 1-p-q & p & 0 & & & \\ 0 & q & 1-p-q & p & & & \cdots \\ \vdots & \ddots & \ddots & \ddots & & & \\ & & & q & 1-p-q & p & \cdots \\ 0 & & & 0 & q & 1-p-q & \cdots \\ \vdots & \vdots & & \vdots & 0 & q & \cdots \\ \vdots & \vdots & & \vdots & \vdots & & \cdots \end{pmatrix}.$$

Simple random walk with two absorbing barriers

A particle moves on the set of $n+1$ integers: $\{0, 1, 2, \ldots, n$. When the state is $j \in \mathbb{Z}_+$, for p, q such that $p + q \leq 1$,

$$\mathrm{P}(\{j+1\}|\{j\}) = p, \quad \mathrm{P}(\{j-1\}|\{j\}) = q,$$

$$\mathrm{P}(\{j\}|\{j\}) = 1 - p - q,$$

when $1 \leq j \leq n-1$, and

$$\mathrm{P}(\{0\}|\{0\}) = \mathrm{P}(\{n\}|\{n\}) = 1.$$

Transition matrix

$$\boldsymbol{P}_n = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ q & 1-p-q & p & 0 & & 0 & 0 \\ 0 & q & 1-p-q & p & & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & & & \\ & & & q & p & 0 & 0 \\ 0 & \vdots & & 0 & 1-p-q & p & 0 \\ 0 & 0 & & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Covariance and correlation

For two sequences of numbers:

$x_1,$	$x_2,$	$x_3,$	$\ldots,$	$x_{n-2},$	$x_{n-1},$	$x_n,$
$y_1,$	$y_2,$	$y_3,$	$\ldots,$	$y_{n-2},$	$y_{n-1},$	y_n

$$s_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y}),$$

$$r_{xy} = \frac{s_{xy}}{s_x \cdot s_y}, \quad s_x = \sqrt{s_{xx}}, \quad s_y = \sqrt{s_{yy}}.$$

Lag 1 autocorrelation

Sequence y is the same x, moved one unit to the left:

$x_1,$	$x_2,$	$x_3,$	$\ldots,$	$x_{n-2},$	$x_{n-1},$	$x_n,$
$x_2,$	$x_3,$	$x_4,$	$\ldots,$	$x_{n-1},$	$x_n.$	

The *lag 1 autocorrelation*, r_1, measures the correlation between each number and the next one in the sequence.

The autocorrelation function (ACF)

Similarly one defines r_2, r_3, r_3, ...,

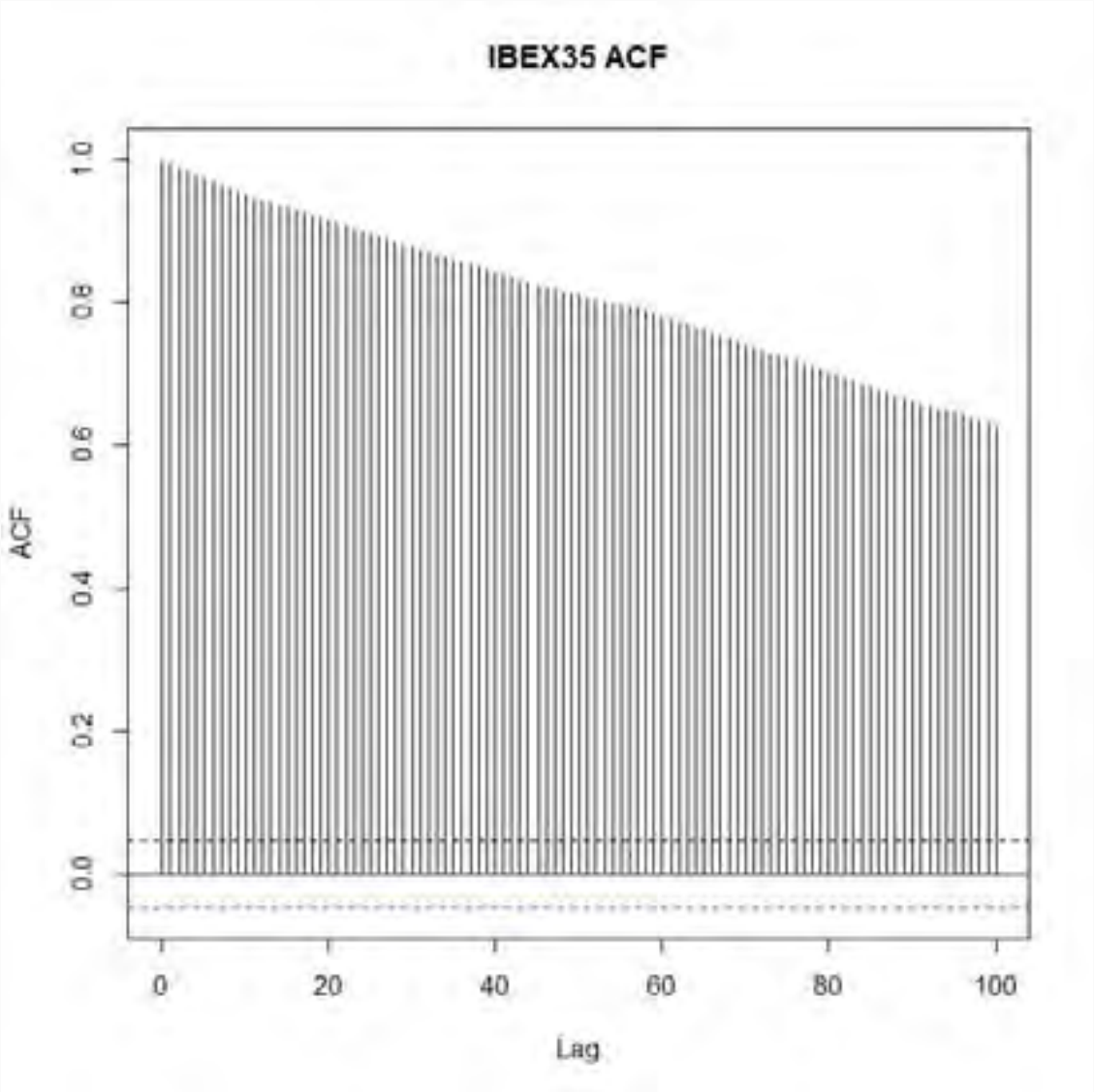
The resulting ACF,

$$r(k) = r_k, \quad k \geq 1,$$

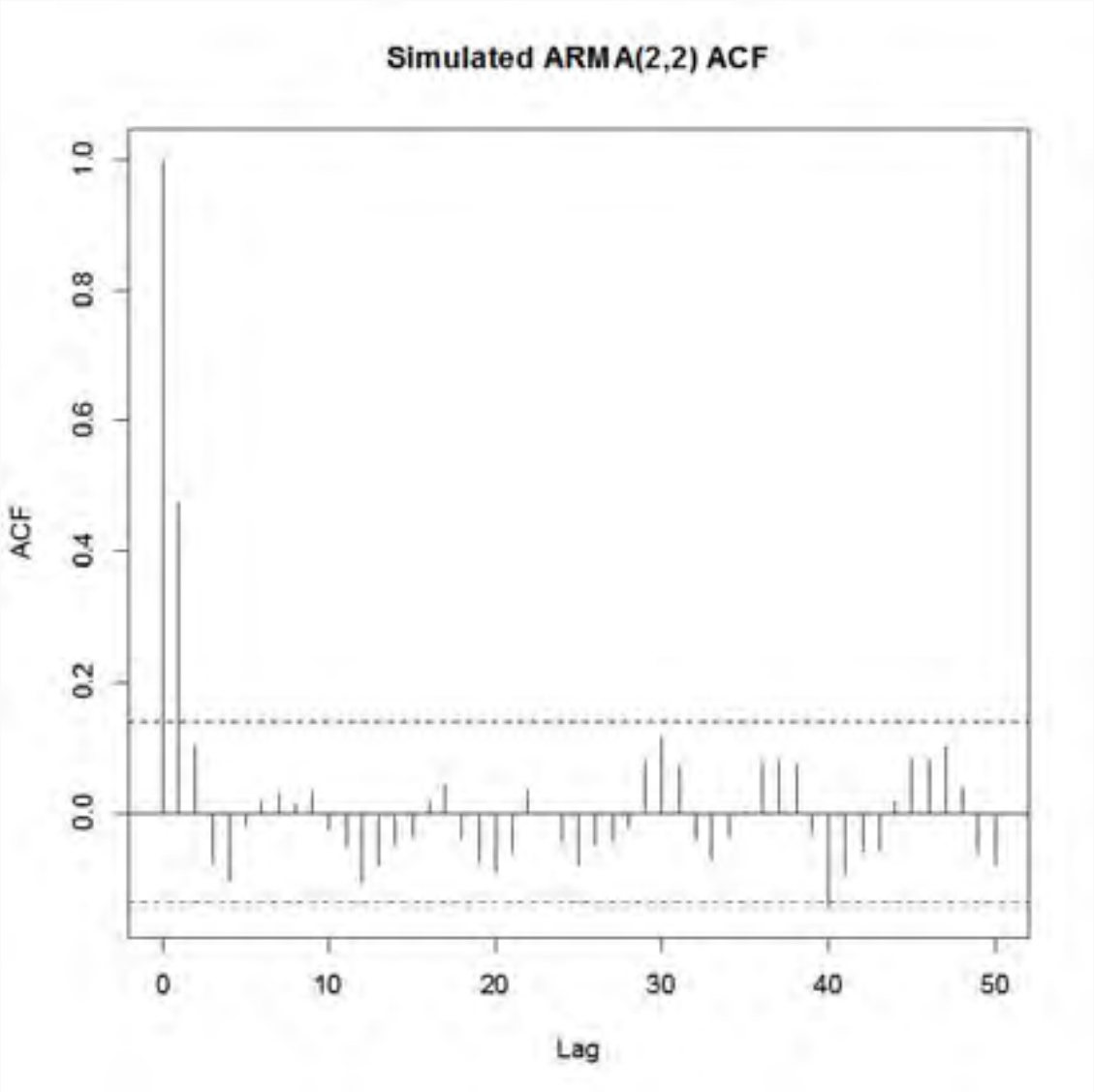
is a widely used descriptive tool for sequences, often time series data.

IBEX35 data, from Jan 02, 2012 to yesterday (Nov 20, 2018)

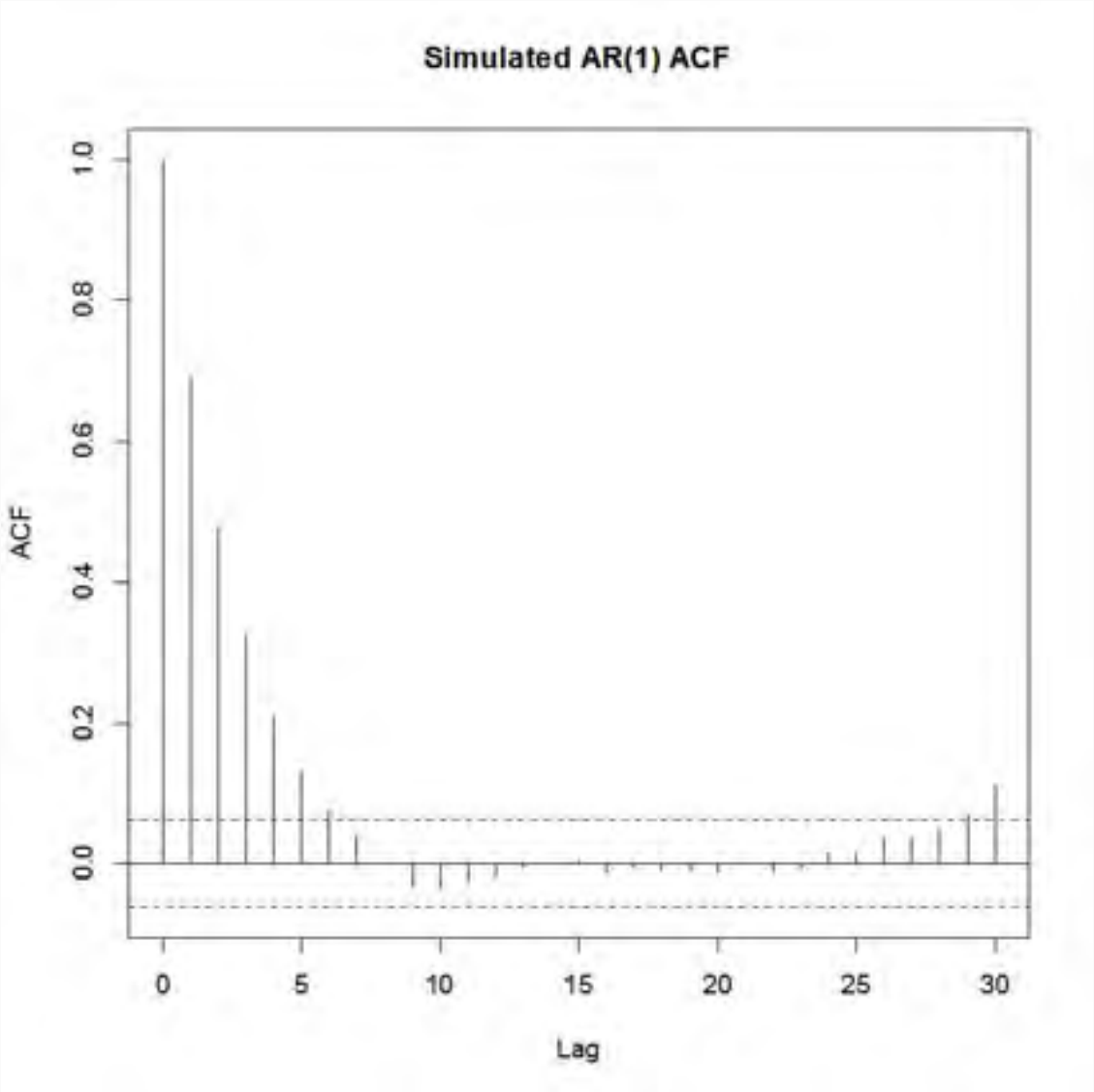
ACF of the IBEX35 data



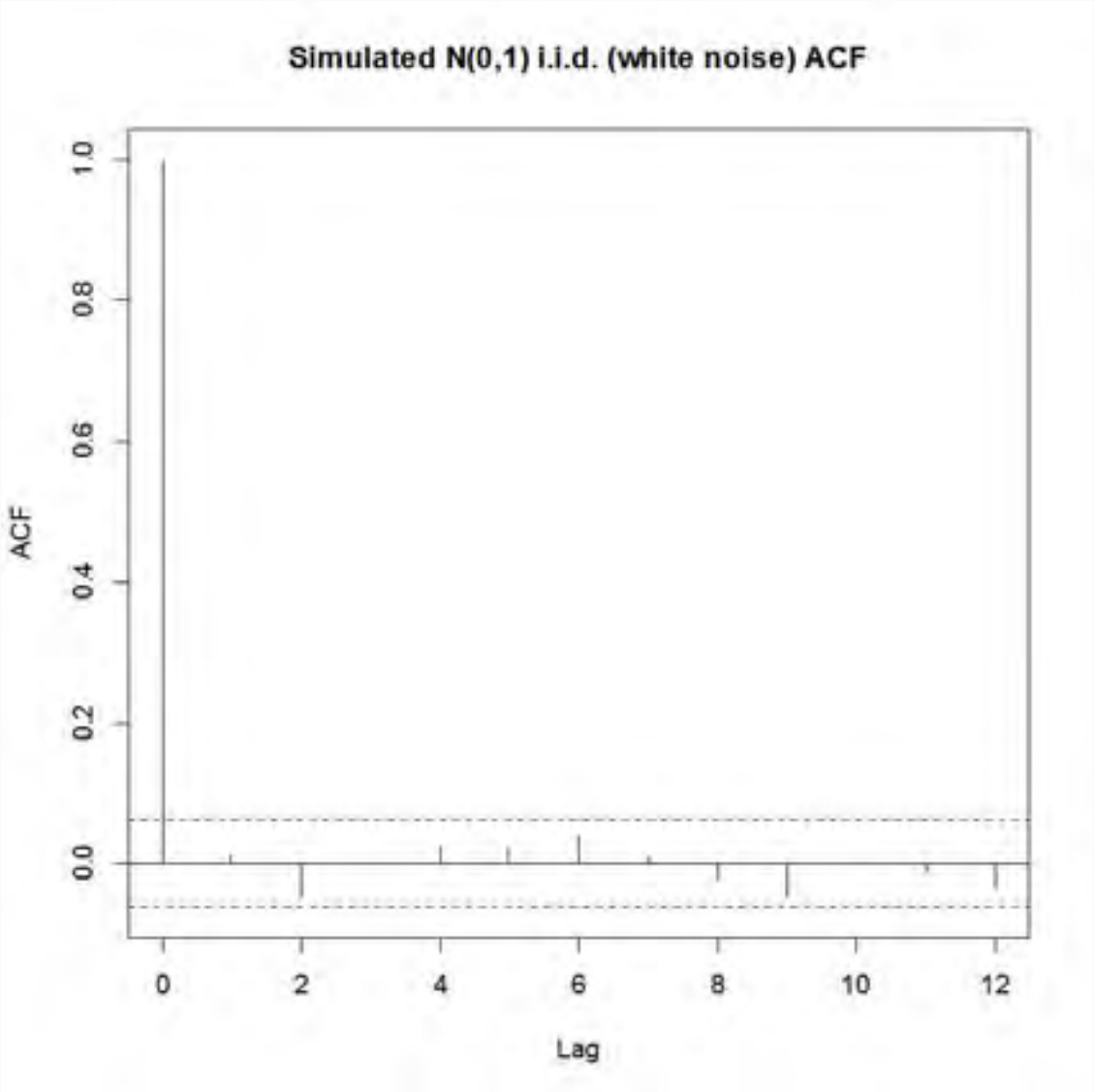
ACF of a simulated ARMA(2,2) series



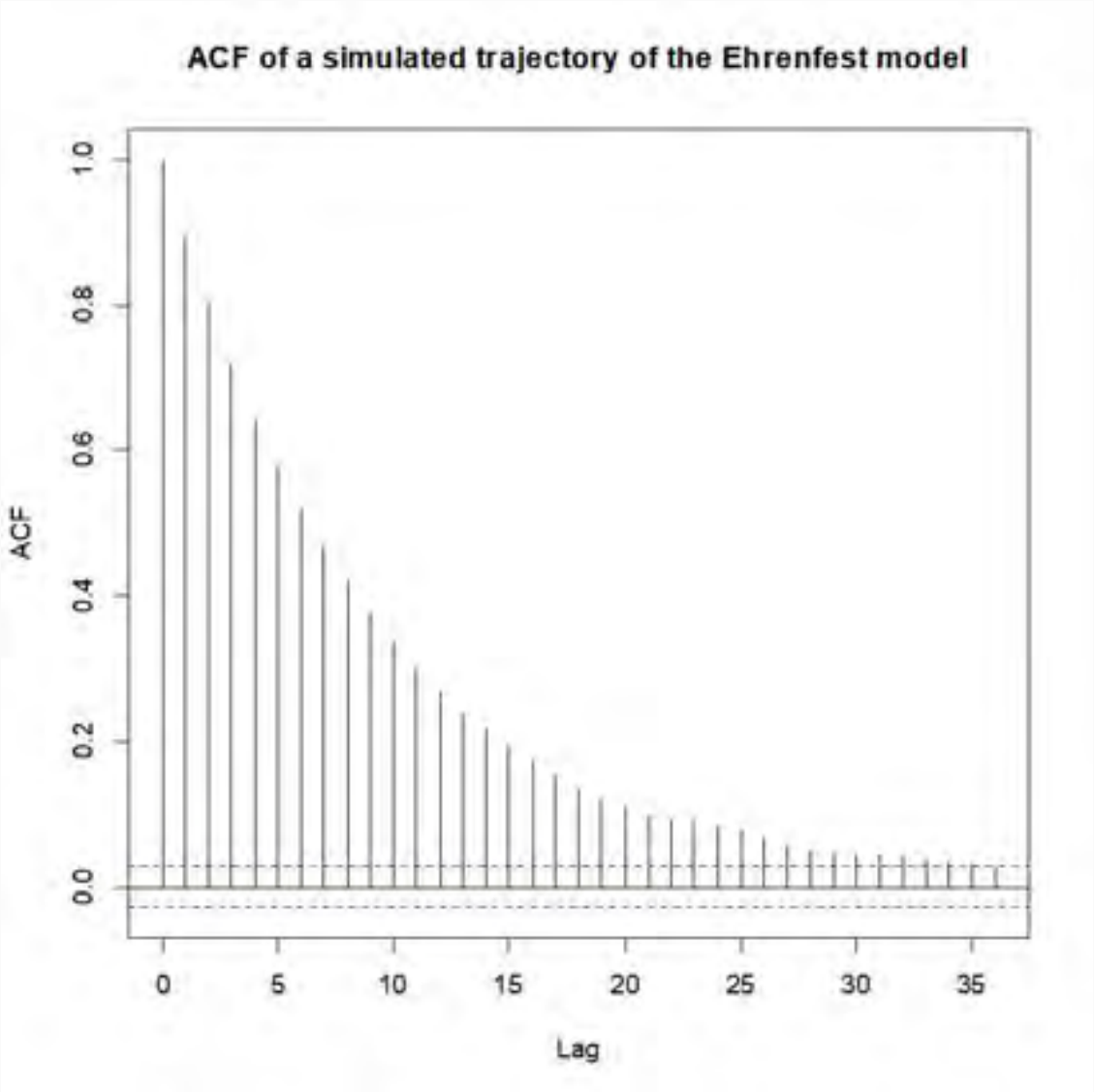
ACF of a simulated AR(1) series



ACF of `rnorm(1000)`



ACF of a simulated MC trajectory



09 - Markov chains - 01

General idea

MCMC procedures construct a *Markov chain* $X = \{X_t\}_{t\in\mathbb{N}}$ converging towards (i.e., having stationary distribution) the target pdf π.

After a sufficiently long time *(mixing time)* finite terms of the sequence will be close to the limit.

To simulate π we generate finite random trajectories in the chain X_t.

Example

The Markov chain with transition matrix:

$$\boldsymbol{P} = \begin{pmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1/3 & 2/3 \end{pmatrix}.$$

has limiting distribution:

$$\boldsymbol{p} = (1/6, 1/3, 1/2).$$

What do we want to do?

We are interested in the *inverse problem:*

> *Given a probability* $\boldsymbol{p}$ *on* S*, construct a Markov chain having* S *as its state space and* $\boldsymbol{p}$ *as its limit probability.*

What do we want to do?

We are interested in the *inverse problem:*

> *Given a probability* $\boldsymbol{p}$ *on* S*, construct a Markov chain having* S *as its state space and* $\boldsymbol{p}$ *as its limit probability.*

Is this feasible?

What do we want to do?

We are interested in the *inverse problem:*

> *Given a probability* $\boldsymbol{p}$ *on* S*, construct a Markov chain having* S *as its state space and* $\boldsymbol{p}$ *as its limit probability.*

Is this feasible?

Yes, for *reversible Markov chains* (with respect to $\boldsymbol{p}$.)

Time reversal in a Markov chain

Consider a 2×2 transition matrix:

$$\boldsymbol{P}_{1\to 2} = \begin{pmatrix} \mathrm{P}(A_1\ [t=2] | A_1\ [t=1]), \mathrm{P}(A_2\ [t=2] | A_1\ [t=1]) \\ \\ \mathrm{P}(A_1\ [t=2] | A_2\ [t=1]), \mathrm{P}(A_2\ [t=2] | A_2\ [t=1]) \end{pmatrix}$$

Rows in $\boldsymbol{P}_{1\to 2}$ are conditional probabilities

Time reversal in a Markov chain

Assume we know:

$$\boldsymbol{p}_1 = (\mathrm{P}(A_1\ [t=1]), \mathrm{P}(A_2\ [t=1]))\,.$$

Multiply:

- the first row of $\boldsymbol{P}_{1\to 2}$ by $\mathrm{P}(A_1\ [t=1])$,
- and the second row by $\mathrm{P}(A_2\ [t=1])$.

In each case the probability of the conditioning event

Time reversal in a Markov chain

We obtain a matrix with probabilities of intersections:

$$\begin{pmatrix} \mathrm{P}(A_1\ [t=2]\,, A_1\ [t=1]) & \mathrm{P}(A_2\ [t=2]\,, A_1\ [t=1]) \\ \mathrm{P}(A_1\ [t=2]\,, A_2\ [t=1]) & \mathrm{P}(A_2\ [t=2]\,, A_2\ [t=1]) \end{pmatrix}$$

Time reversal in a Markov chain

Transpose it:

$$\begin{pmatrix} \mathrm{P}(A_1\ [t=2], A_1\ [t=1]) & \mathrm{P}(A_1\ [t=2], A_2\ [t=1]) \\ \mathrm{P}(A_2\ [t=2], A_1\ [t=1]) & \mathrm{P}(A_2\ [t=2], A_2\ [t=1]) \end{pmatrix}$$

Time reversal in a Markov chain

Intersection is symmetrical: Switch the $[t = 1]$, $[t = 2]$ halves.

$$\begin{pmatrix} \mathrm{P}(A_1\ [t=1], A_1\ [t=2]) & \mathrm{P}(A_2\ [t=1], A_1\ [t=2]) \\ \mathrm{P}(A_1\ [t=1], A_2\ [t=2]) & \mathrm{P}(A_2\ [t=1], A_2\ [t=2]) \end{pmatrix}$$

Sum of the first row is: $\mathrm{P}(A_1\ [t=2])$, and of the second row: $\mathrm{P}(A_2\ [t=2])$.

Time reversal in a Markov chain

Finally, divide each row by its total: first row by $\mathrm{P}(A_1\ [t=2])$, and second row by $\mathrm{P}(A_2\ [t=2])$.
The result is:

$$\boldsymbol{P}_{2\to 1} = \begin{pmatrix} \mathrm{P}(A_1\ [t=1]|A_1\ [t=2]) & \mathrm{P}(A_2\ [t=1]|A_1\ [t=2]) \\ \mathrm{P}(A_1\ [t=1]|A_2\ [t=2]) & \mathrm{P}(A_2\ [t=1]|A_2\ [t=2]) \end{pmatrix}$$

The matrix of transition probabilities $2 \to 1$.

Time reversal in a general homogeneous Markov chain

The *time reversal operation* at time t is:

$$\widetilde{P}_{t+1\to t} = \boldsymbol{D}_{\boldsymbol{p}_{t+1}}{}^{-1} \cdot \boldsymbol{P}' \cdot \boldsymbol{D}_{\boldsymbol{p}_t}$$

To reverse time along the whole span of t and still get a homogeneous Markov chain we need the existence of a limit probability (and to know it).

Time reversal of a Markov chain with limit probability

If $\boldsymbol{p}$ is the limit probability, then:

$$\widetilde{P} = {D_p}^{-1} \cdot P' \cdot D_p$$

is the transition matrix of the new chain.

Reversible chains

A homogeneous Markov chain such that:

$$\widetilde{\boldsymbol{P}} = \boldsymbol{D_p}^{-1} \cdot \boldsymbol{P}' \cdot \boldsymbol{D_p} = \boldsymbol{P}. \qquad (\star)$$

is called *time-reversible.*

Equation $(\star)$ is called the *detailed balance* condition.
In terms of the individual entries:

$$p_i \cdot P_{ij} = p_j \cdot P_{ji}, \qquad 1 \leq i, j \leq n.$$

Reversible chains

Writing the time-reversibility condition requires to know the stationary pdf $\boldsymbol{p}$ in advance.

But this is precisely what happens when solving the *inverse problem:* given a pdf $\boldsymbol{p}$ derive a transition matrix $\boldsymbol{P}$ such that $\boldsymbol{p}$ is the stationary limit pdf (of the associated Markov chain).

Reversible chains

Theorem:

If, for a Markov chain with transition matrix $\boldsymbol{P}$,
the *detailed balance* equation $(\star)$ is satisfied
for a probability $\boldsymbol{p}$, then the chain is time-reversible
and $\boldsymbol{p}$ is the unique stationary distribution.

09 - Markov chains - 01

Finite Metropolis algorithm

$\theta \in S$, a state space with n elements, $S = \{1, \ldots, n\}$.

A target probability:

$$\pi = (\pi_1, \ldots, \pi_n), \qquad \sum_{i=1}^{n} \pi_i = 1$$

with support on S.

Finite Metropolis algorithm: Candidate transition matrix

An $n \times n$ stochastic matrix $\boldsymbol{K} = (k(i,j) : i,j \in S)$ of *transition proposal probabilities*.

We assume here $\boldsymbol{K}$ is a symmetric matrix

(see below the general case).

Finite Metropolis algorithm: Candidate transition matrix

A possible $\boldsymbol{K}$, the transition matrix of a *symmetric cyclic random walk* on S.

At a given time, either move to the right or to the left with equal probability, except if the current position is an extreme, where there is 50% probability to go to the opposite one.

Finite Metropolis algorithm: Candidate transition matrix

$$K = \begin{pmatrix} 0.0 & 0.5 & 0.0 & 0.0 & 0.0 & 0.5 \\ 0.5 & 0.0 & 0.5 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.5 & 0.0 & 0.5 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.5 & 0.0 & 0.5 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.5 & 0.0 & 0.5 \\ 0.5 & 0.0 & 0.0 & 0.0 & 0.5 & 0.0 \end{pmatrix}$$

Finite Metropolis algorithm: candidate generation

The step $\theta^{(m)} \longrightarrow \theta^{(m+1)}$ is:

(1) Generate the candidate $\theta' \in S$ from the pmf:

$$k(\theta^{(m)}, \cdot),$$

(the $\theta^{(m)}$-th row in $\boldsymbol{K}$).

Finite Metropolis algorithm: Accept-reject step

(2) Generate a random $I \sim \text{Ber}(p)$, with probability:

$$p = \min\left\{1, \frac{\pi_{\theta'}}{\pi_{\theta^{(m)}}}\right\}.$$

If $I = 1$, accept the displacement, that is, set

$$\theta^{(m+1)} = \theta',$$

Otherwise keep $\theta^{(m+1)} = \theta^{(m)}$.

Transition matrix of the resulting chain

This defines a Markov chain with transition matrix:

$$\boldsymbol{P} = (P_{ij} : \ i, j \in S),$$

where:

$$P_{ij} = k(i,j) \cdot \min\left\{1, \ \frac{\pi_j}{\pi_i}\right\}, \qquad i \neq j,$$

and

$$P_{ii} = 1 - \sum_{j \neq i} P_{ij}.$$

The resulting chain satisfies the detailed balance

Since:

$$\pi_i P_{ij} = k(i,j) \cdot \min\{\pi_i, \pi_j\},$$

The symmetry of $\boldsymbol{K}$ entails this is equal to:

$$\pi_j P_{ji} = k(j,i) \cdot \min\{\pi_j, \pi_i\}.$$

Hence it is a *time-reversible* Markov chain.

The target pdf is the limit probability

Such a Markov chain has π as its limit distribution.

Indeed:

$$\begin{aligned}(\pi \cdot \boldsymbol{P})_j &= \sum_{i \in S} \pi_i \, P_{ij} = \sum_{i \in S} \pi_j \, P_{ji} \\ &= \pi_j \sum_{i \in S} P_{ji} = \pi_j,\end{aligned}$$

since rows in $\boldsymbol{P}$ have unit sum.

09 - Markov chains - 01

References

Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller (1953), *Equation of State Calculations by Fast Computing Machines*, J. Chemical Physics, Vol. 21, pp. 1087–1092.

Wilfred Keith Hastings (1970), *Monte Carlo sampling methods using Markov chains and their applications*, Biometrika 57, 97-109.

Description

A generalization of the above basic algorithm, with a non-symmetric matrix $\boldsymbol{K}$.

The difference is in the acceptance rule.

Now we accept the displacement, setting $\theta^{(m+1)} = \theta'$, with probability:

$$\min\left\{1, \frac{\pi_{\theta'} \cdot k\left(\theta', \theta^{(m)}\right)}{\pi_{\theta^{(m)}} \cdot k\left(\theta^{(m)}, \theta'\right)}\right\},$$

Description

The transition matrix is:

$$P_{ij} = k(i,j) \cdot \min\left\{1, \frac{\pi_j \cdot k(j,i)}{\pi_i \cdot k(i,j)}\right\}, \qquad i \neq j,$$

and

$$P_{ii} = 1 - \sum_{j \neq i} P_{ij}.$$

Conditions of detailed balance

They are also satisfied, since:

$$\begin{aligned} \pi_i \cdot P_{ij} &= k(i,j) \cdot \min\left\{\pi_i, \frac{\pi_j\, k(j,i)}{k(i,j)}\right\} \\ &= \min\left\{\pi_i\, k(i,j), \pi_j\, k(j,i)\right\} \\ &= \pi_j \cdot P_{ji}, \end{aligned}$$

hence the chain is reversible with respect to π, the stationary distribution.