Stochastic processes and Markov chains (part II)

Wessel van Wieringen
w.n.van.wieringen@vu.nl

Department of Epidemiology and Biostatistics, VUmc
& Department of Mathematics, VU University
Amsterdam, The Netherlands

vrije Universiteit

VU medisch centrum

Example

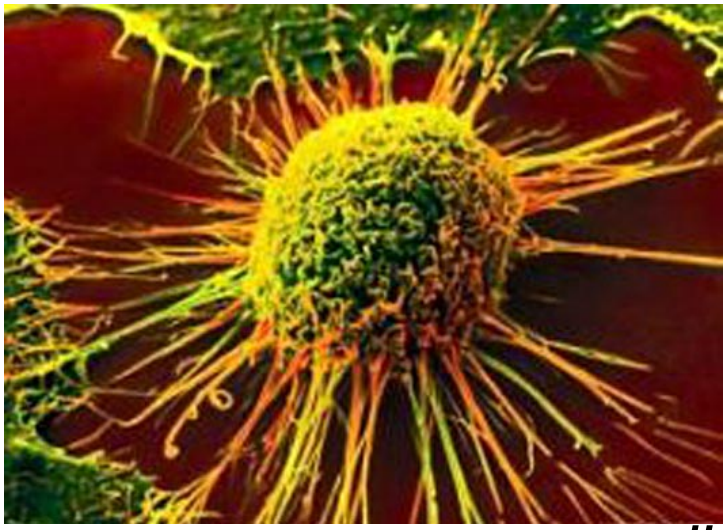
---

evolution of cancer

Example

Cancer is a genetic disease caused by abnormalities in the DNA of the cell.

A cancer cell exhibits unproliferated growth.



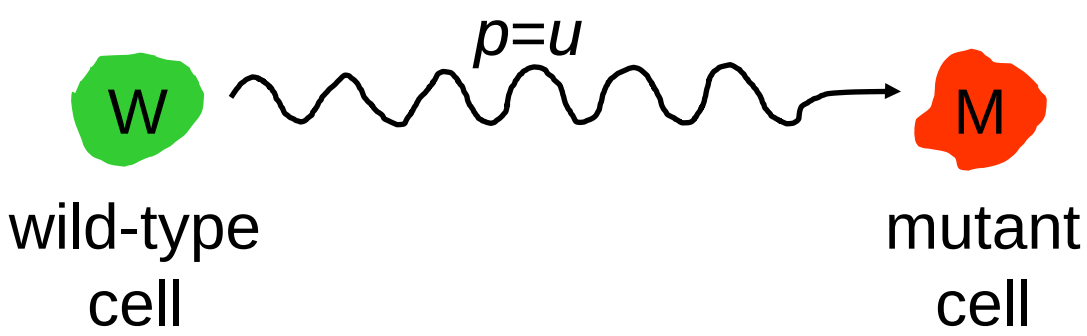
cancer cell

Knudson's two-hit hypothesis

The cell contains two copies of each gene. A single mutation (hit) in (one of the copies) of a gene is not sufficient for cancer to develop. A second hit to the non-mutated copy in the gene pair may produce cancer.

Example

One-hit model



- Wild-type cells have a probability of 1-*u* reproduce faithfully.
- With probability *u* the reproduction is not faithfully, and a mutation in the genetic material of the cell is introduced.
- Cells with a mutation are called *mutant cells*.
- Mutant cells always reproduce faithfully.

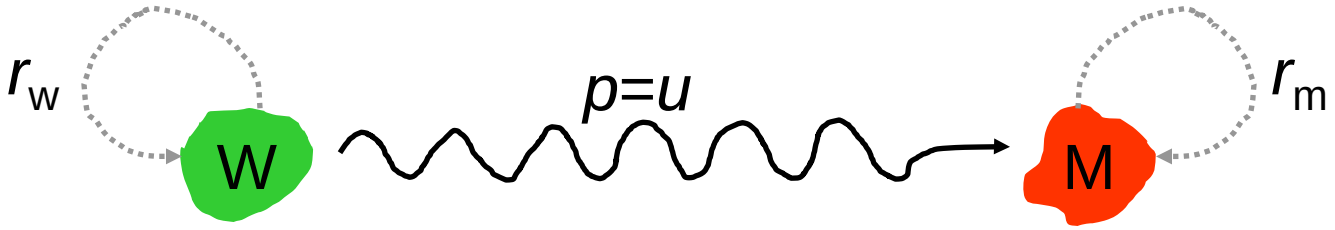
Example

Let w = #W be the number of wild-type cells.

Let m = #M be the number of mutant cells.

Let $N = w + m$ be the population size, considered fixed.

The mutation may affect the cell's reproductive rate. To accommodate this, denote the reproductive rates of the wild-type and mutant cell by r_w and r_m, respectively.



Example

At each time step, one cell reproduces and one cell dies.

The probability that a cell reproduces is proportional to its frequency and the reproductive rate, and is given by:

$$p_{+W} \propto r_w\, w \qquad \text{and} \qquad p_{+M} \propto r_m\, m$$

Rescale to turn these into probabilities:

$$p_{+W} = \frac{r_w\, w}{r_w\, w + r_m\, m}$$

and

$$p_{+M} = \frac{r_m\, m}{r_w\, w + r_m\, m}$$

Example

The probabilities on the previous slide do not take into account that a wild-type cell may not reproduce faithfully. When doing so, we end up with:

$$p_{+W} = (1-u)\,\frac{r_w\, w}{r_w\, w + r_m\, m}$$

and

$$p_{+M} = u\,\frac{r_w\, w}{r_w\, w + r_m\, m} + \frac{r_m\, m}{r_w\, w + r_m\, m}$$

Both cell types have probability to die proportional to their abundance. This leads to:

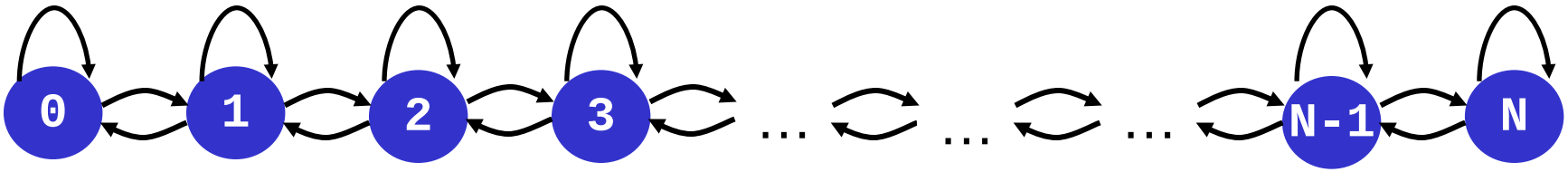
$$p_{-W} = w/N \qquad \text{and} \qquad p_{-M} = m/N$$

Example

Consider the random variable m = #M.
Its dynamical behavior is described by a Markov process.

The state space for this process S = {**0**, **1**, **2**, **3**, …, **N-1**, **N**}.

State diagram for m = #M



Due to $N = w + m$, it is sufficient to study only one.

Example

Probability of a decrease of *m* at any time step:

$$P(M_{t+1} = i - 1 \,|\, M_t = i) \;=\; p_{-M}\, p_{+W}$$

Probability of an increase of *m* at any time step:

$$P(M_{t+1} = i + 1 \,|\, M_t = i) \;=\; p_{+M}\, p_{-W}$$

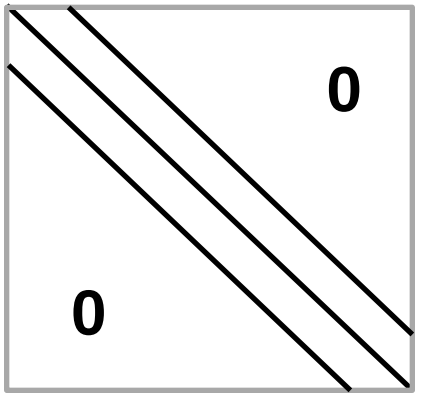
Probability of no change in *m* at any time step:

$$P(M_{t+1} = i \,|\, M_t = i) \;=\; 1 - p_{-M}\, p_{+W} - p_{+M}\, p_{-W}$$

Right-hand side depends on i.

All other transition probabilities are zero.

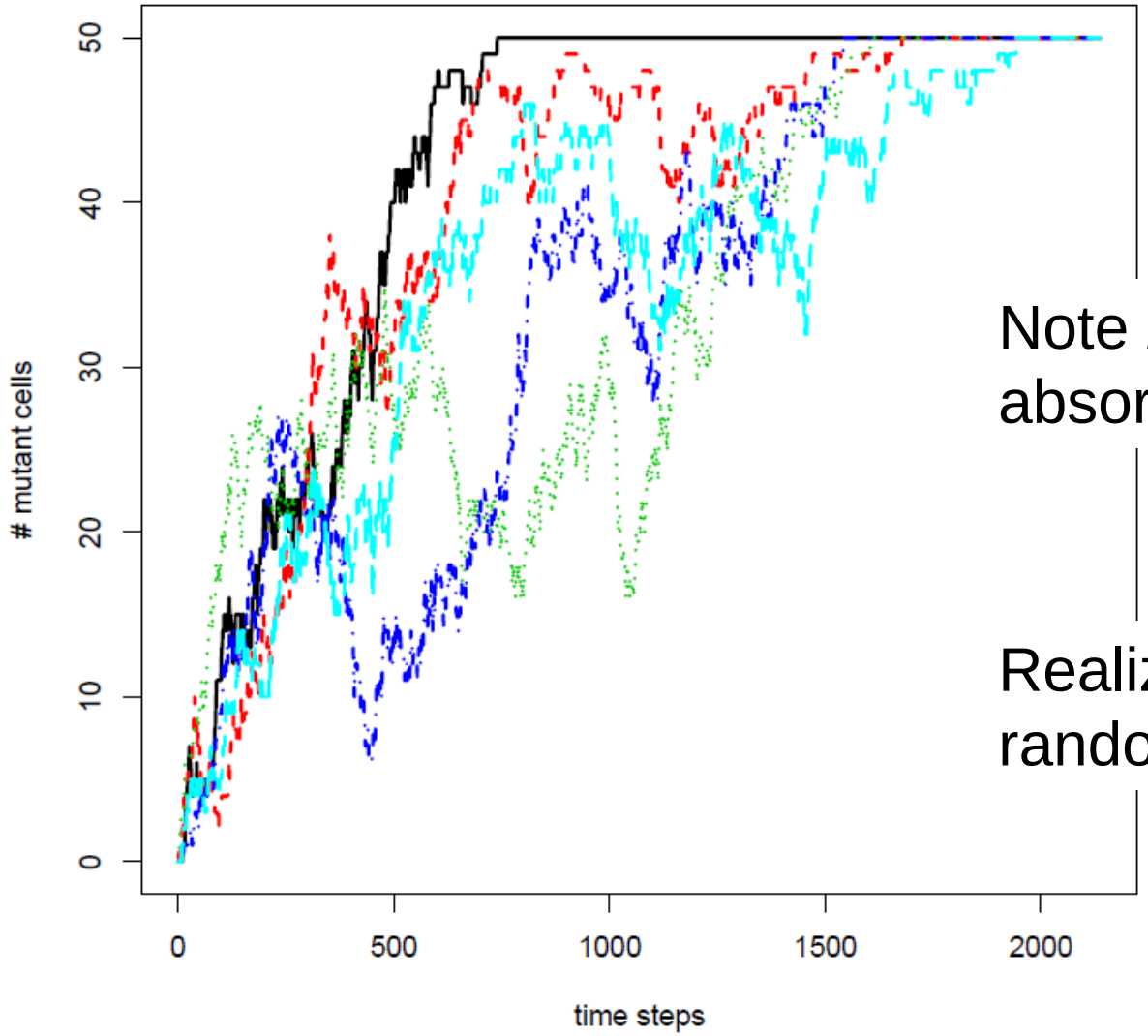
transition matrix structure



a banded matrix

Example

Five realizations of this Markov process (N=50)

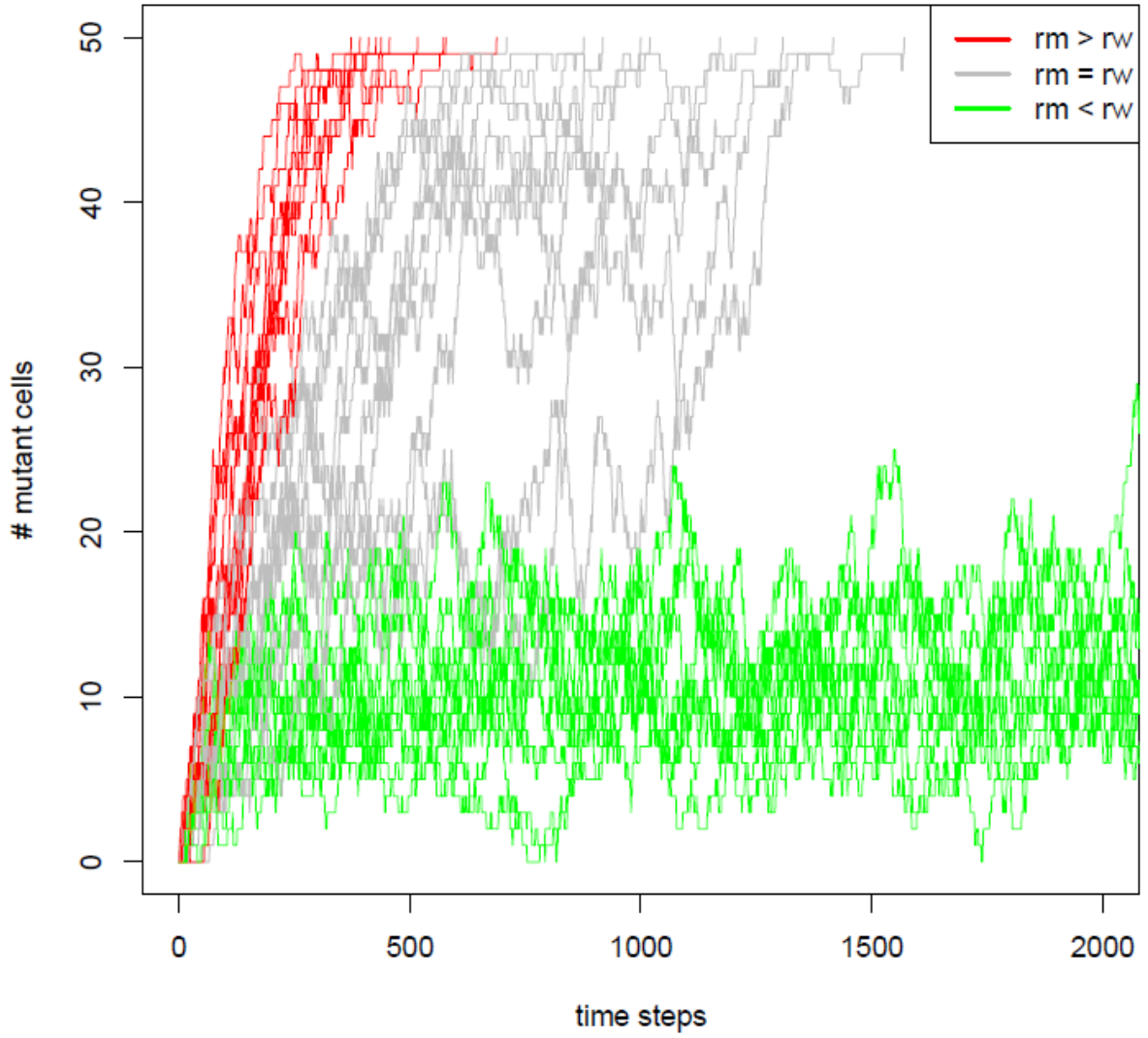


Note $m = N$ is an absorbing state.

Realization = random walk

Example

Effect of r_m / r_w, 10 realizations, N=50



Example

Conclusion

Within the one-hit model mutations that affect the reproduction rate of the cell determine whether a cancer is benign or malignant.

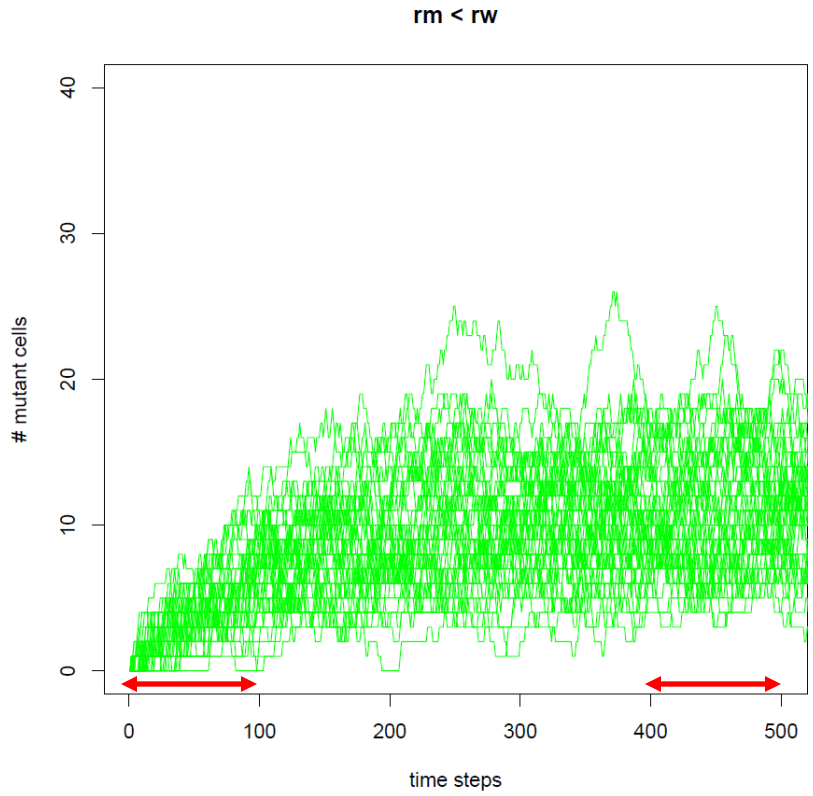
The discussed one-hit model can be extended to the two-hit model. Qualitative properties (like the expected time of absorption) of the one- and two-hit models may be derived. Comparison of these qualitative properties to empirical observations may distinguish between the two models.

Stationary distribution

Stationary distribution

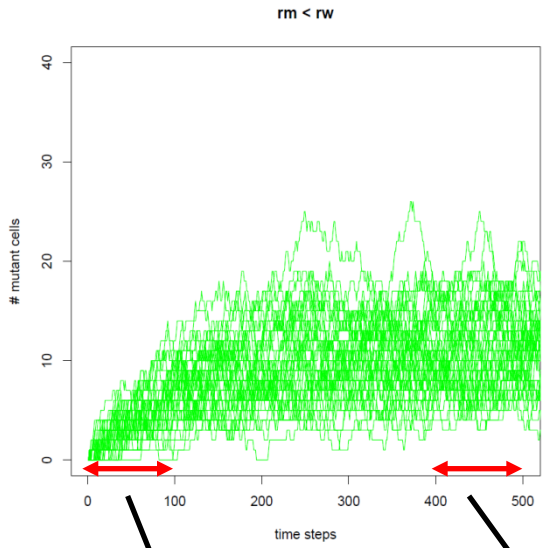
Question

Consider random walks of cancer evolution example:



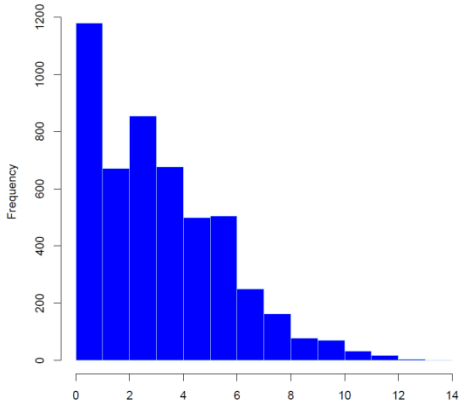
What is the difference between the random walk at generations [0,100] and those at [400, 500]?

Stationary distribution

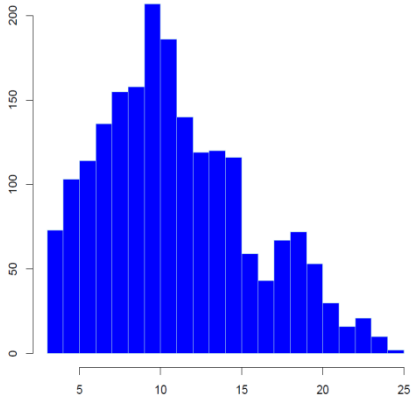


The distribution of the # mutant cells changes in the beginning, but settles after a while.

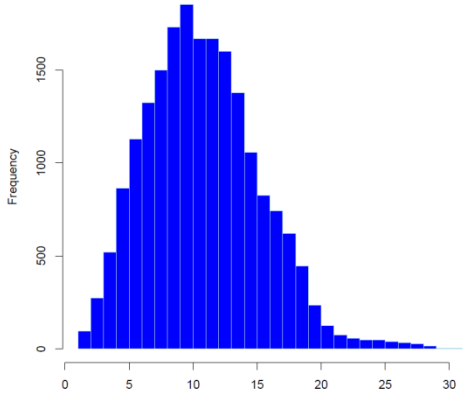
generations 1-100



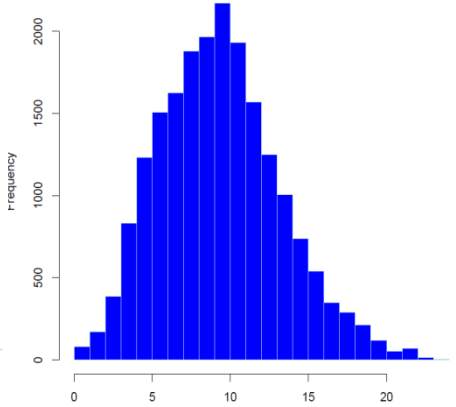
generations 400-500



generations 5000-6000



generations 9000-10000



Stationary distribution

book
Axelson-Fisk (2010): Page 34.
Ewens, Grant (2005): Section 4.8.1.

Hence, after a while an “equilibrium” sets in. Not necessarily a fixed state or pattern, but:

> *the proportion of a time period that is spent in a particular state converges to a limit value.*

The limit values of all states, e.g.:

$$\varphi_{\mathtt{A}} = \lim_{t\to\infty} P(X_t = \mathtt{A})$$

form the *stationary distribution* of the Markov process, denoted by:

$$\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \ldots, \varphi_S)^T \qquad \text{with} \sum_{k=1}^{S} \varphi_k = 1$$

Stationary distribution

For a stationary process, it holds that

$$P(X_t=E_i) = \varphi_i$$

for all *t* and *i* (follows from def. of stationary distribution.)

In particular:

$$P(X_t=E_i) = \varphi_i = P(X_{t+1}=E_i).$$

This does *not* imply:

$$P(X_t=E_i, X_{t+1}=E_i) = \varphi_i \varphi_i$$

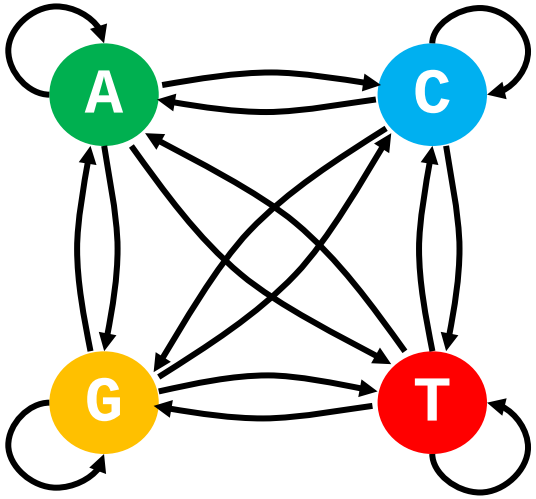
as this ignores the 1st order Markov dependency.

book Axelson-Fisk (2010): Theorem 2.1.

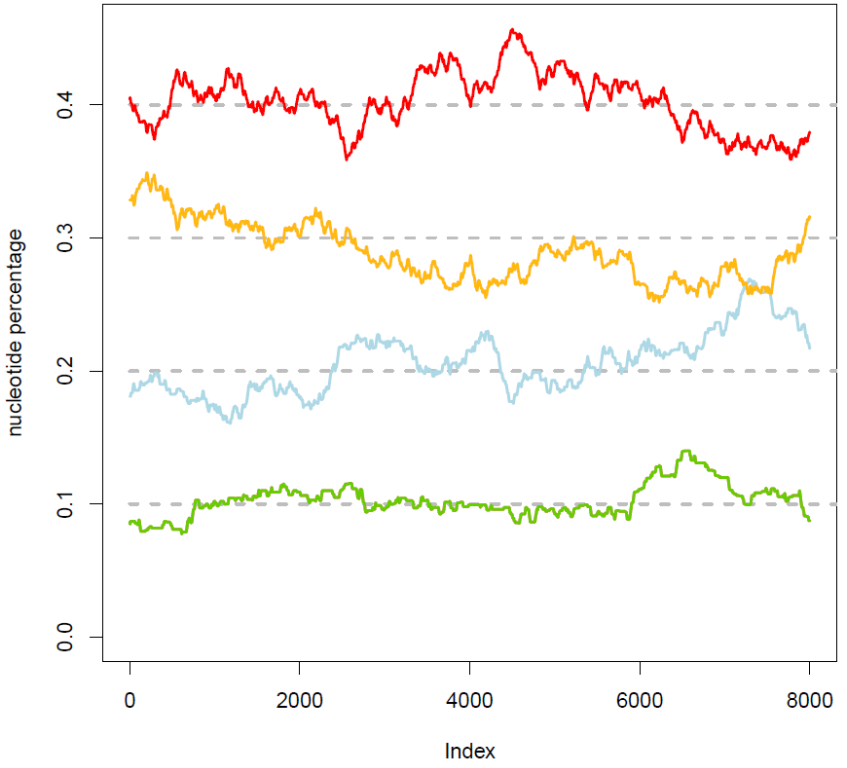
Stationary distribution

Theorem

Irreducible, aperiodic Markov chains with a finite state space S have a stationary distribution to which the chain converges as $t \to \infty$.



φ_A=0.1, φ_C=0.2, φ_G=0.3, φ_T=0.4



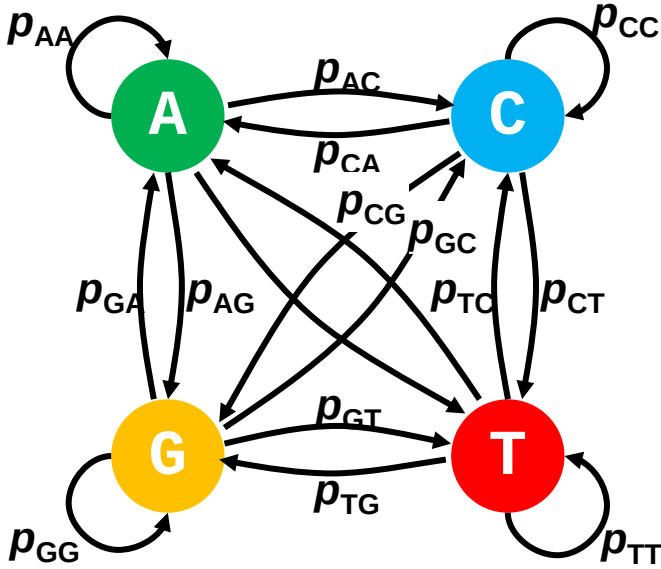
Stationary distribution

book Axelson-Fisk (2010): Definition 2.7.

A Markov chain is *aperiodic* if there is no state that can only be reached at multiples of a certain period. E.g., state E_i only at $t = 0, 3, 6, 9$, et cetera.

Example of an aperiodic Markov chain

$$\mathbf{P} = \begin{array}{c} \\ \mathrm{A} \\ \mathrm{C} \\ \mathrm{G} \\ \mathrm{T} \end{array} \begin{array}{c} \begin{array}{cccc} \mathrm{A} & \mathrm{C} & \mathrm{G} & \mathrm{T} \end{array} \\ \begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix} \end{array}$$

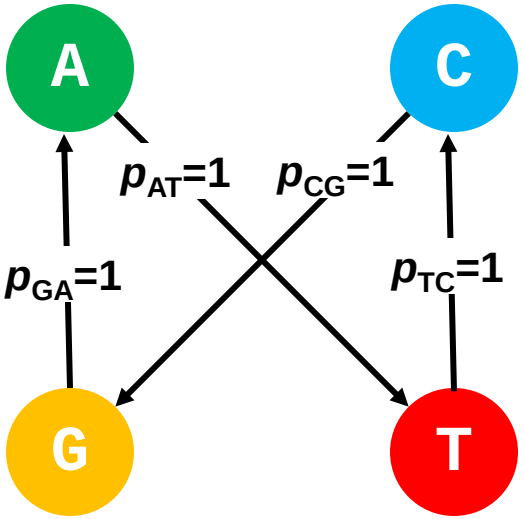


...ATGGTACGCTCCCGTA...

Stationary distribution

Example of a periodic Markov chain

$$\mathbf{P} = \begin{array}{c} \\ \mathbf{A} \\ \mathbf{C} \\ \mathbf{G} \\ \mathbf{T} \end{array} \begin{array}{c} \begin{array}{cccc} \mathbf{A} & \mathbf{C} & \mathbf{G} & \mathbf{T} \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \end{array}$$



…ATCGATCGATCGATCG…

The fraction of time spent in **A** (roughly $\varphi_{\mathbf{A}}$):

$$P(X_{t+1000} = \mathbf{A}) = ¼$$

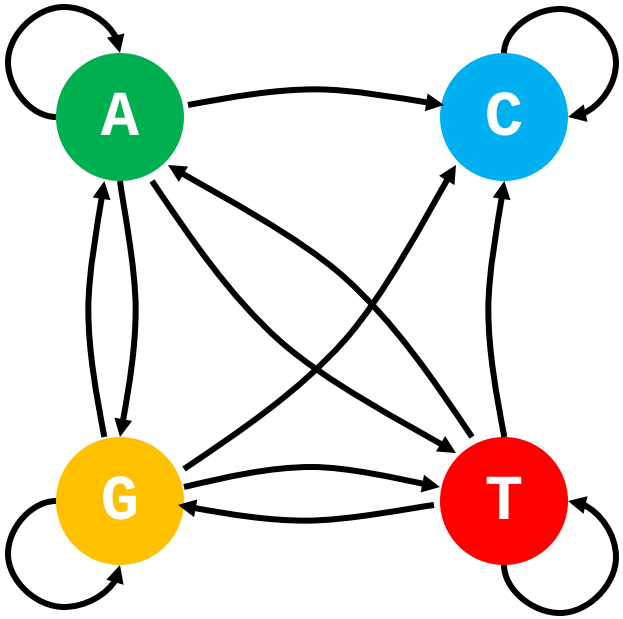
whereas $P(X_{4+1000} = \mathbf{A} \mid X_4 = \mathbf{A}) = 1$.

book Axelson-Fisk (2010): Definition 2.4.

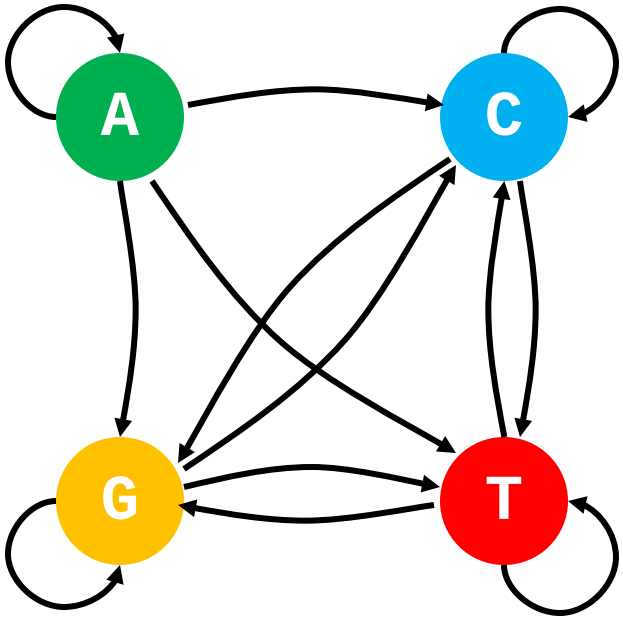
Stationary distribution

A Markov chain is *irreducible* if every state can (in principle) be reached (after enough time has passed) from every other state.

Examples of a reducible Markov chain



C is an absorbing state.



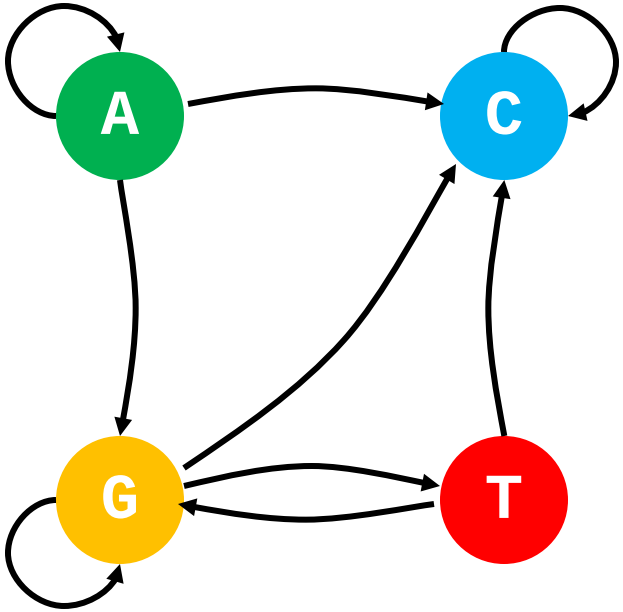
A will not be reached.

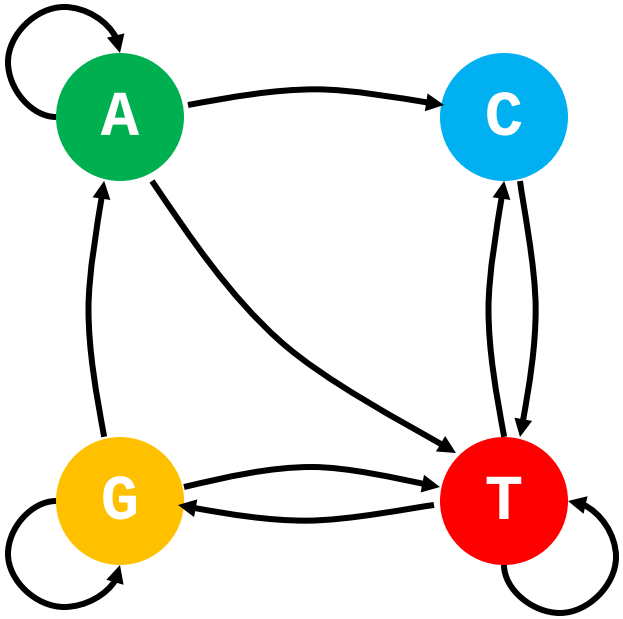
Stationary distribution

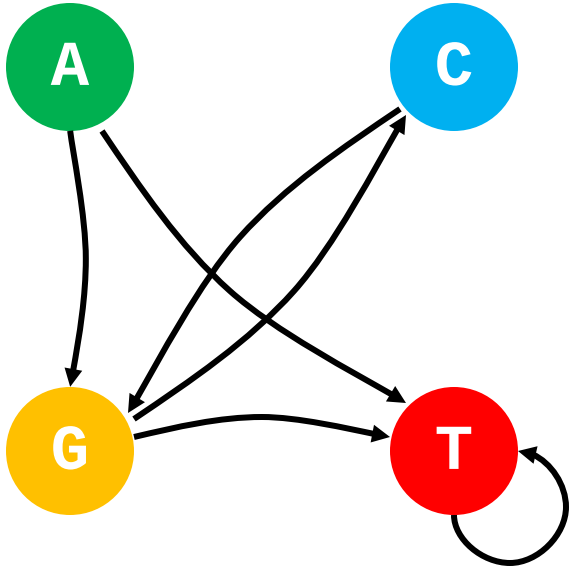
Question

→ Which is (a)periodic?

→ Which is (ir)reducible?







Stationary distribution

The stationary distribution is associated with the first-order Markov process, parameterized by ($\boldsymbol{\pi}$, $\mathbf{P}$).

Question

How do $\boldsymbol{\varphi}$ and ($\boldsymbol{\pi}$, $\mathbf{P}$) relate?

Hereto, recall:

$$\varphi_i = P(X_t = E_i)$$

(due to stationarity).

Stationary distribution

We then obtain the following relation:

$$\varphi_i \quad = \quad P(X_{t+1} = E_i)$$

definition

Stationary distribution

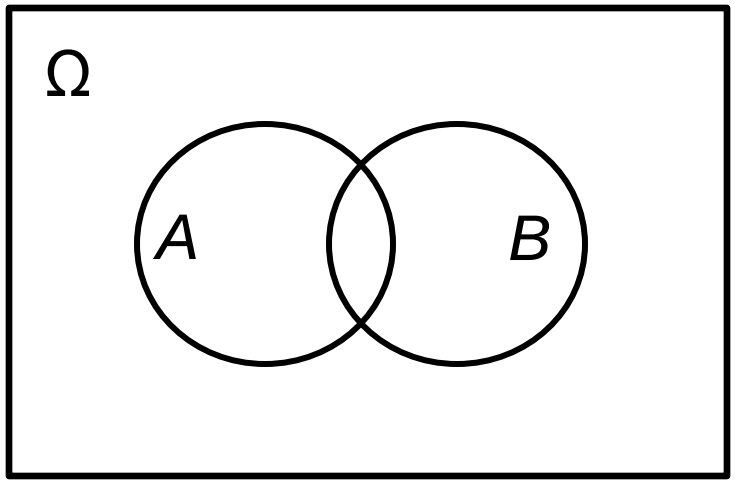
We then obtain the following relation:

$$\begin{aligned} \varphi_i &= P(X_{t+1} = E_i) \\ &= \sum_{k=1}^{S} P(X_{t+1} = E_i, X_t = E_k) \end{aligned}$$

use the fact that

P(*A*, *B*) + P(*A*, B^C) = P(*A*)

(*total probability law*)



Stationary distribution

We then obtain the following relation:

$$\begin{aligned}
\varphi_i &= P(X_{t+1} = E_i) \\
&= \sum_{k=1}^{S} P(X_{t+1} = E_i, X_t = E_k) \\
&= \sum_{k=1}^{S} P(X_{t+1} = E_i \mid X_t = E_k)\, P(X_t = E_k)
\end{aligned}$$

use the definition of conditional probability:

P(*A* | *B*) = P(*A*, *B*) / P(*B*)

Or: P(*A*, *B*) = P(*A* | *B*) P(B)

Stationary distribution

We then obtain the following relation:

$$
\begin{aligned}
\varphi_i &= P(X_{t+1} = E_i) \\
&= \sum_{k=1}^{S} P(X_{t+1} = E_i, X_t = E_k) \\
&= \sum_{k=1}^{S} P(X_{t+1} = E_i \mid X_t = E_k)\, P(X_t = E_k) \\
&= \sum_{k=1}^{S} p_{ki}\, \varphi_k
\end{aligned}
$$

Stationary distribution

We then obtain the following relation:

$$
\begin{aligned}
\varphi_i &= P(X_{t+1} = E_i) \\
&= \sum_{k=1}^{S} P(X_{t+1} = E_i, X_t = E_k) \\
&= \sum_{k=1}^{S} P(X_{t+1} = E_i \mid X_t = E_k)\, P(X_t = E_k) \\
&= \sum_{k=1}^{S} p_{ki}\, \varphi_k
\end{aligned}
$$

Thus:

$$\boldsymbol{\varphi}^T = \boldsymbol{\varphi}^T \mathbf{P} \longrightarrow \text{Eigenvectors!}$$

Stationary distribution

How to find the stationary distribution?

We know the stationary distribution satisfies:

$$\boldsymbol{\varphi}^T = \boldsymbol{\varphi}^T \mathbf{P} \qquad (*)$$

and

$$\varphi_1 + \varphi_2 + \ldots + \varphi_S = 1$$

We thus have *S*+1 equations for *S* unknowns: one of the equations in (*) can be dropped (which is irrelevant), and the system of *S* remaining equations needs to be solved.

Stationary distribution

Example

Consider the transition matrix:

$$\mathbf{P} = \begin{pmatrix} 0.35 & 0.65 \\ 0.81 & 0.19 \end{pmatrix}$$

In order to find the stationary distribution we need to solve the following system of equations:

$$\begin{aligned} (\varphi_1, \varphi_2) &= (\varphi_1, \varphi_2) \begin{pmatrix} 0.35 & 0.65 \\ 0.81 & 0.19 \end{pmatrix} \\ 1 &= \varphi_1 + \varphi_2 \end{aligned}$$

Stationary distribution

Example (continued)

Rewritten:

$$\begin{aligned} 0.35\varphi_1 + 0.81\varphi_2 &= \varphi_1 \\ 0.65\varphi_1 + 0.19\varphi_2 &= \varphi_2 \\ \varphi_1 + \varphi_2 &= 1 \end{aligned}$$

This system of equations is overdetermined (3 equations, only 2 variables). Solution: skip one of the first two equations. Now solve:

$$\begin{aligned} 0.35\varphi_1 + 0.81\varphi_2 &= \varphi_1 \\ \varphi_1 + \varphi_2 &= 1 \end{aligned}$$

This yields: $(\varphi_1, \varphi_2)^T = (0.5547945, 0.4452055)^T$

Stationary distribution

How do we find the stationary distribution? (II)

For very large transition steps (in an irreducible and aperiodic 1st order Markov chain) the effect of the previous state vanishes:

$$(\mathbf{P}^{(n)})_{ij} = P(X_{t+n}=E_j \mid X_t=E_i) \approx P(X_{t+n}=E_j)$$

for large enough *n*.

The $(\mathbf{P}^{(n)})_{ij}$ are thus independent of *i* (for large *n*)

On the other hand, stationarity: $\varphi_j = P(X_{t+n}=E_j)$.

Together this yields:

$$(\mathbf{P}^{(n)})_{ij} \approx P(X_{t+n}=E_j) = \varphi_j,$$

for large *n*.

Stationary distribution

How to find the stationary distribution? (II)

Hence, the *n*-step transition matrix $\mathbf{P}^{(n)}$ has identical rows:

$$\lim_{n\to\infty} \mathbf{P}^{(n)} = \lim_{n\to\infty} \mathbf{P}^{n} = \begin{pmatrix} \varphi_1 & \varphi_2 & \cdots & \varphi_S \\ \varphi_1 & \varphi_2 & \cdots & \varphi_S \\ \vdots & \vdots & & \vdots \\ \varphi_1 & \varphi_2 & \cdots & \varphi_S \end{pmatrix}$$

This motivates a numerical way to find the stationary distribution:

→ Study powers of **P**: **P** x **P** x **P** x **P** x **P** x **P** x ….

book Ewens, Grant (2005): Section 4.8.2.

Stationary distribution

Example (continued)

The transition matrix:

$$\mathbf{P} = \begin{pmatrix} 0.35 & 0.65 \\ 0.81 & 0.19 \end{pmatrix}$$

with stationary distribution: $(\varphi_1, \varphi_2)^T = (0.5547945, 0.4452055)^T$

Then:

$$\mathbf{P}^{(2)} = \begin{pmatrix} 0.35 & 0.65 \\ 0.81 & 0.19 \end{pmatrix} \begin{pmatrix} 0.35 & 0.65 \\ 0.81 & 0.19 \end{pmatrix}$$

$$= \begin{pmatrix} 0.35 \times 0.35 + 0.65 \times 0.81 & 0.35 \times 0.65 + 0.65 \times 0.19 \\ 0.81 \times 0.35 + 0.19 \times 0.81 & 0.81 \times 0.65 + 0.19 \times 0.19 \end{pmatrix}$$

matrix multiplication (“rows times columns”)

Stationary distribution

Example (continued)

Thus:

$$\mathbf{P}^{(2)} = \begin{pmatrix} 0.6490 & 0.3510 \\ 0.4374 & 0.5626 \end{pmatrix}$$

In similar fashion we obtain:

$$\mathbf{P}^{(5)} = \begin{pmatrix} 0.5456249 & 0.4543751 \\ 0.5662212 & 0.4337788 \end{pmatrix}$$

$$\mathbf{P}^{(20)} = \begin{pmatrix} 0.5547946 & 0.4452054 \\ 0.5547944 & 0.4452056 \end{pmatrix}$$

Stationary distribution

book Axelson-Fisk (2010): Theorem 2.1.

The stationary distribution, the “proportion of time spent in each state”, of a Markov chain is related to the “waiting time” to observe another instance of each state.

Consider an irreducible, aperiodic Markov chain.
Define the random variable T_i by:

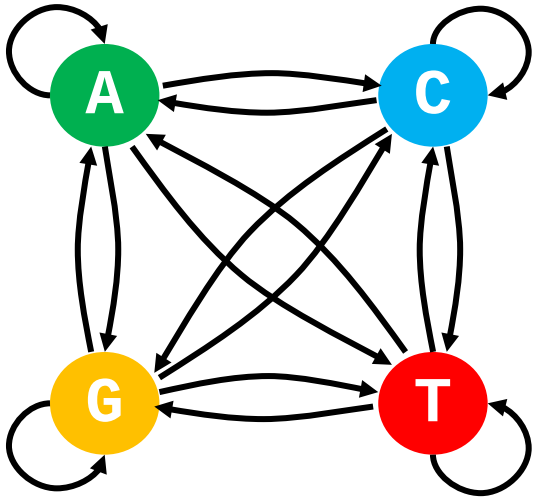
$$T_i = \min_n \{n \geq 1 : X_{t+n} = E_i \mid X_t = E_i\}$$

The *recurrence time* of state E_i after visiting it at time t.

The *mean recurrence time* of state E_i , the expectation of T_i, is then (proof omitted):

$$E(T_i) = 1/\varphi_i$$

Stationary distribution



Question

Recall the DNA example with stationary distribution:

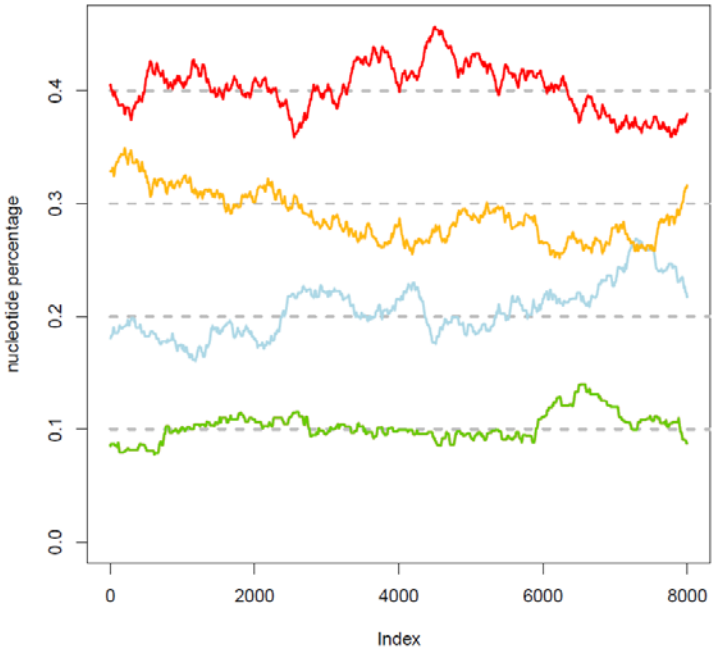
$\varphi_A=0.1$, $\varphi_C=0.2$,

$\varphi_G=0.3$, $\varphi_T=0.4$.

The mean recurrence time for the nucleotides then is:

$E(T_A)$ = …

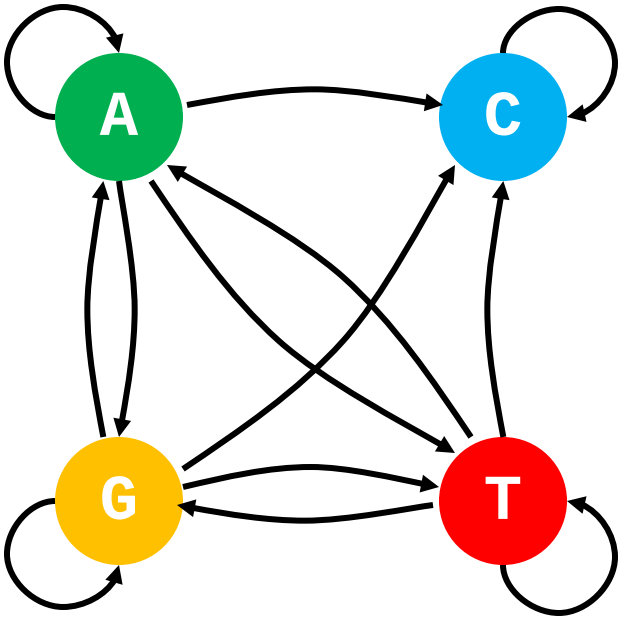
$E(T_C)$ = …



Stationary distribution

Question

What is the mean recurrence time of each state of the 1st order Markov process (describing the DNA sequence) corresponding to the state diagram below?

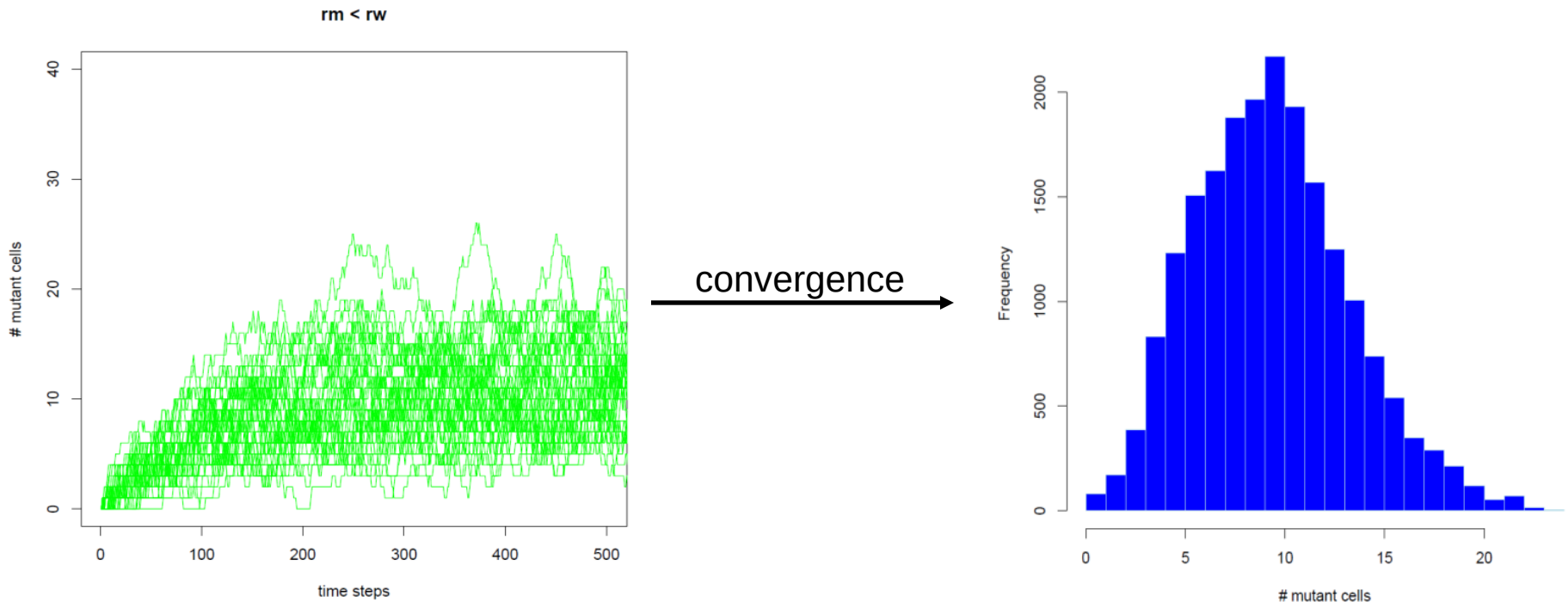


Convergence to the stationary distribution

Convergence

Motivation

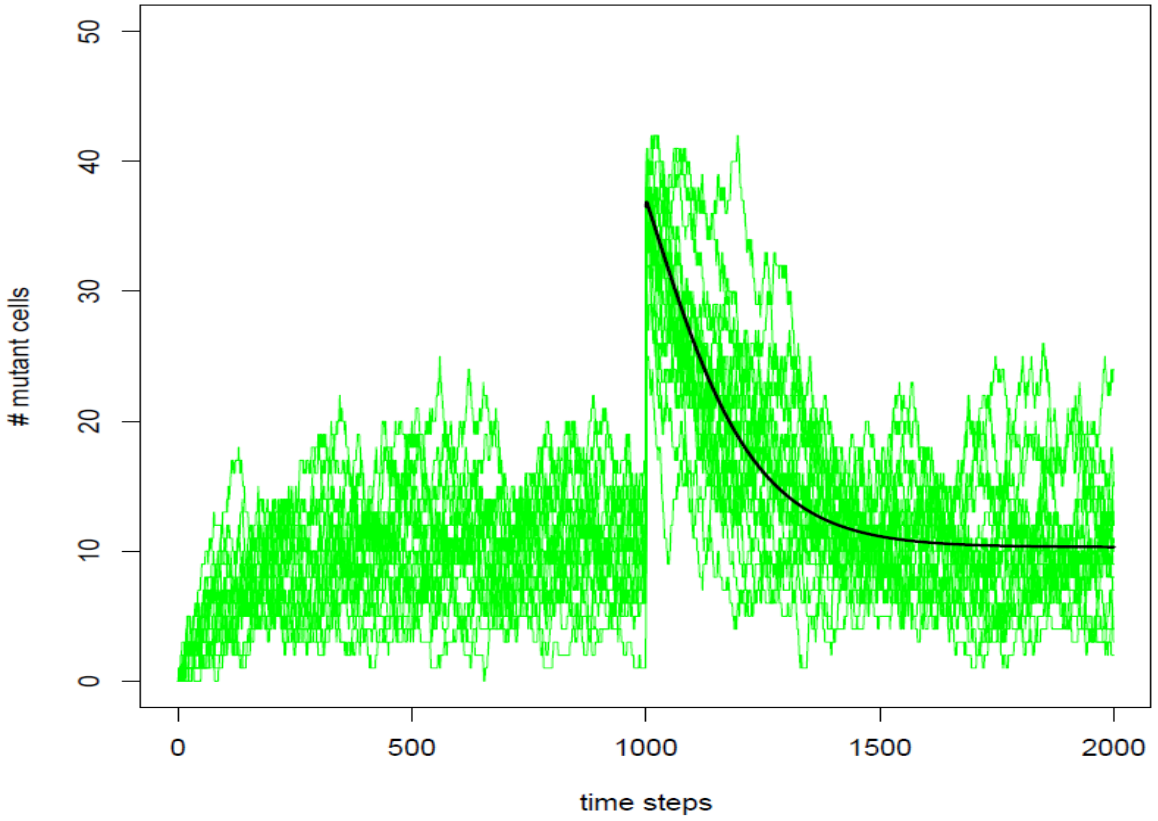
Recall the one-hit model for cancer evolution. After how many generations are the cell types in “equilibrium”?



Convergence

Motivation

Due to (say) radiation, not part of the 1st order Markov process, the number of mutant cells suddenly increases. How fast does its effect in the cell population vanish?



Convergence

Define the vector $\mathbf{1} = (1,\ldots,1)^T$.

We have already seen:

$$\mathbf{P}^n = \mathbf{1}\,\boldsymbol{\varphi}^T \qquad \text{for large } n$$

Question

How fast does $\mathbf{P}^n$ go to $\mathbf{1}\,\boldsymbol{\varphi}^T$ as $n \to \infty$?

Answer

1) Use linear algebra
2) Calculate numerically

Convergence

Fact

The transition matrix **P** of a finite, aperiodic, irreducible Markov chain has an eigenvalue equal to 1 ($\lambda_1=1$), while all other eigenvalues are (in the absolute sense) smaller than one: $|\lambda_k| < 1$, k=2,…,3.

Focus on $\lambda_1=1$

We know $\boldsymbol{\varphi}^T \mathbf{P} = \boldsymbol{\varphi}^T$ for the stationary distribution. Hence, $\boldsymbol{\varphi}$ is the left eigenvector of eigenvalue $\lambda=1$.

Also, row sums of **P** equal **1**: $\mathbf{P}\,\mathbf{1} = \mathbf{1}$. Hence, **1** is a right eigenvector of eigenvalue λ=1.

Convergence

Example (continued)

The transition matrix:

$$\mathbf{P} = \begin{pmatrix} 0.35 & 0.65 \\ 0.81 & 0.19 \end{pmatrix}$$

has eigenvalues $\lambda_1 = 1$ and $\lambda_2 = -0.46$.

```
> # Verify in R
> P <- matrix(c(0.35, 0.81, 0.65, 0.19), 2, 2)
> D <- diag(eigen(P)$value)
> Vl <- eigen(P)$vectors
> # reconstruct P
> Vl %*% D %*% solve(Vl)
```

Convergence

The *spectral decomposition* of a square matrix **P** is given by:

$$\mathbf{P} = \mathbf{V}\,\mathbf{D}\,\mathbf{V}^{-1}$$

where:

$\mathbf{D}$ diagonal matrix containing the eigenvalues,

$\mathbf{V}$ columns contain the corresponding eigenvectors.

In case of **P** is symmetric, $\mathbf{V}$ is orthogonal: $\mathbf{V}^{-1} = \mathbf{V}^{T}$

Then:

$$\mathbf{P} = \mathbf{V}\,\mathbf{D}\,\mathbf{V}^{T}$$

Convergence

The *spectral decomposition* of **P** can be reformulated as:

$$\mathbf{P} = \sum_{k=1}^{S} \lambda_k \, \mathbf{r}_k \, \boldsymbol{\ell}_k^T$$

eigenvalues — λ_k

right eigenvectors — $\mathbf{r}_k$

left eigenvectors — $\boldsymbol{\ell}_k^T$

The eigenvectors are normalized:

$$\begin{aligned} \mathbf{r}_k^T \, \boldsymbol{\ell}_k &= 1, && \text{and} \\ \mathbf{r}_{k_1}^T \, \boldsymbol{\ell}_{k_2} &= 0 && \text{if } k_1 \neq k_2 \end{aligned}$$

Convergence

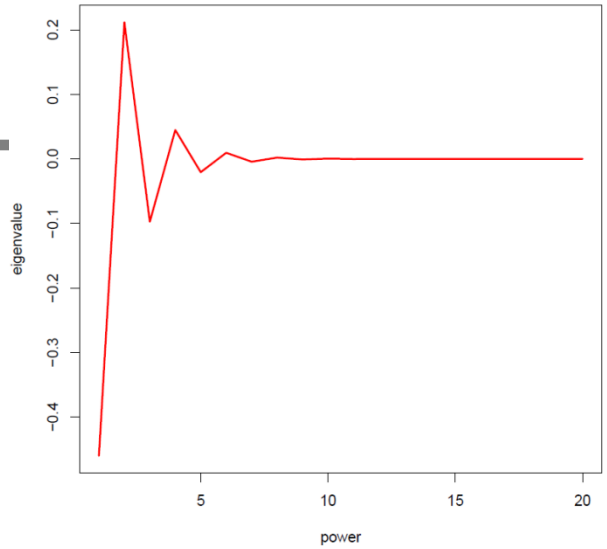
Suppose spectral decomposition of one-step transition matrix known, i.e.:

$$\begin{aligned} \mathbf{P}\mathbf{r}_k &= \lambda_k \mathbf{r}_k \\ \boldsymbol{\ell}_k^\top \mathbf{P} &= \lambda_k \boldsymbol{\ell}_k^\top \end{aligned}$$

Then (see SM) , $\boldsymbol{\ell}_k$ and $\mathbf{r}_k$ are left and right eigenvector with eigenvalue λ_k^n of $\mathbf{P}^n$. Thus:

$$\mathbf{P}^n = \sum_{k=1}^{S} \lambda_k^n \, \mathbf{r}_k \, \boldsymbol{\ell}_k^T$$

Convergence



Example (continued)

The transition matrix:

$$\mathbf{P}^{(2)} = \begin{pmatrix} 0.6490 & 0.3510 \\ 0.4374 & 0.5626 \end{pmatrix}$$

has eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 0.2116$;

$$\mathbf{P}^{(5)} = \begin{pmatrix} 0.5456249 & 0.4543751 \\ 0.5662212 & 0.4337788 \end{pmatrix}$$

has eigenvalues $\lambda_1 = 1$ and $\lambda_2 = -0.0205963$;

$$\mathbf{P}^{(20)} = \begin{pmatrix} 0.5547946 & 0.4452054 \\ 0.5547944 & 0.4452056 \end{pmatrix}$$

has eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 1.79952e\text{-}07$.

Convergence

Use the spectral decomposition of $\mathbf{P}^n$ to show how fast $\mathbf{P}^n$ converges to $\mathbf{1}\,\boldsymbol{\varphi}^T$ as $n \to \infty$.

We know:

$\lambda_1=1$, $|\lambda_k| < 1$ for k=2, …,S, $\boldsymbol{\ell}_1 = \boldsymbol{\varphi}$ and $\mathbf{r}_1 = \mathbf{1}$

Then:

$$\mathbf{P}^n = 1^n\,\mathbf{1}\,\boldsymbol{\varphi}^T + \sum_{k=2}^{S} \lambda_k^n\,\mathbf{r}_k\,\boldsymbol{\ell}_k^T$$

Convergence

Expanding this:

$$\mathbf{P}^n = 1^n \mathbf{1}\boldsymbol{\varphi}^T + \sum_{k=2}^{S} \lambda_k^n \mathbf{r}_k \boldsymbol{\ell}_k^T$$

$$= \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} (\varphi_1, \varphi_2, \ldots, \varphi_S) + \lambda_2^n \mathbf{r}_2 \boldsymbol{\ell}_2^T + \ldots + \lambda_S^n \mathbf{r}_S \boldsymbol{\ell}_S^T$$

$$= \begin{pmatrix} \varphi_1 & \varphi_2 & \cdots & \varphi_S \\ \varphi_1 & \varphi_2 & \cdots & \varphi_S \\ \vdots & \vdots & & \vdots \\ \varphi_1 & \varphi_2 & \cdots & \varphi_S \end{pmatrix} + \lambda_2^n \mathbf{r}_2 \boldsymbol{\ell}_2^T + \ldots + \lambda_S^n \mathbf{r}_S \boldsymbol{\ell}_S^T$$

as n → ∞

$\lambda_2^n \to 0$ $\qquad$ $\lambda_S^n \to 0$

Convergence

Clearly:

$$\lim_{n\to\infty} \mathbf{P}^n = \begin{pmatrix} \varphi_1 & \varphi_2 & \cdots & \varphi_S \\ \varphi_1 & \varphi_2 & \cdots & \varphi_S \\ \vdots & \vdots & & \vdots \\ \varphi_1 & \varphi_2 & \cdots & \varphi_S \end{pmatrix}$$

Furthermore, as:

$$0 \leq |\lambda_{k_1}| < |\lambda_{k_2}| < 1 \qquad \Rightarrow \qquad 0 \leq |\lambda_{k_1}^n| < |\lambda_{k_2}^n| < 1$$

It is the second largest (in absolute sense) eigenvalue that dominates, and thus determines the convergence speed to the stationary distribution.

Convergence

Fact (used in exercises)

A Markov chain with a symmetric **P** has a uniform stationary distribution.

Proof

- Symmetry of **P** implies that left- and right eigenvectors are identical (up to a constant).
- First right eigenvector corresponds to vector of ones, **1**.
- Hence, the left eigenvector equals c**1**.
- The left eigenvector is the stationary distribution and should sum to one: c = 1 / (number of states).

Convergence

Question

Suppose the DNA may reasonably be described by a first order Markov model with transition matrix **P**:

$$\mathbf{P} = \begin{pmatrix} 0.2 & 0.3 & 0.3 & 0.2 \\ 0.1 & 0.4 & 0.4 & 0.1 \\ 0.3 & 0.2 & 0.2 & 0.3 \\ 0.4 & 0.1 & 0.1 & 0.4 \end{pmatrix}$$

and stationary distribution:

$$\boldsymbol{\varphi}^T = (\varphi_A, \varphi_C, \varphi_G, \varphi_T) = (1/4, 1/4, 1/4, 1/4)$$

and eigenvalues:

$$\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)^T = (1, 0.2, 0, 0)^T$$

Convergence

Question

What is the probability of a **G** at position 2, 3, 4, 5, 10?
And how does this depend on the starting nucleotide?

In other words, give:

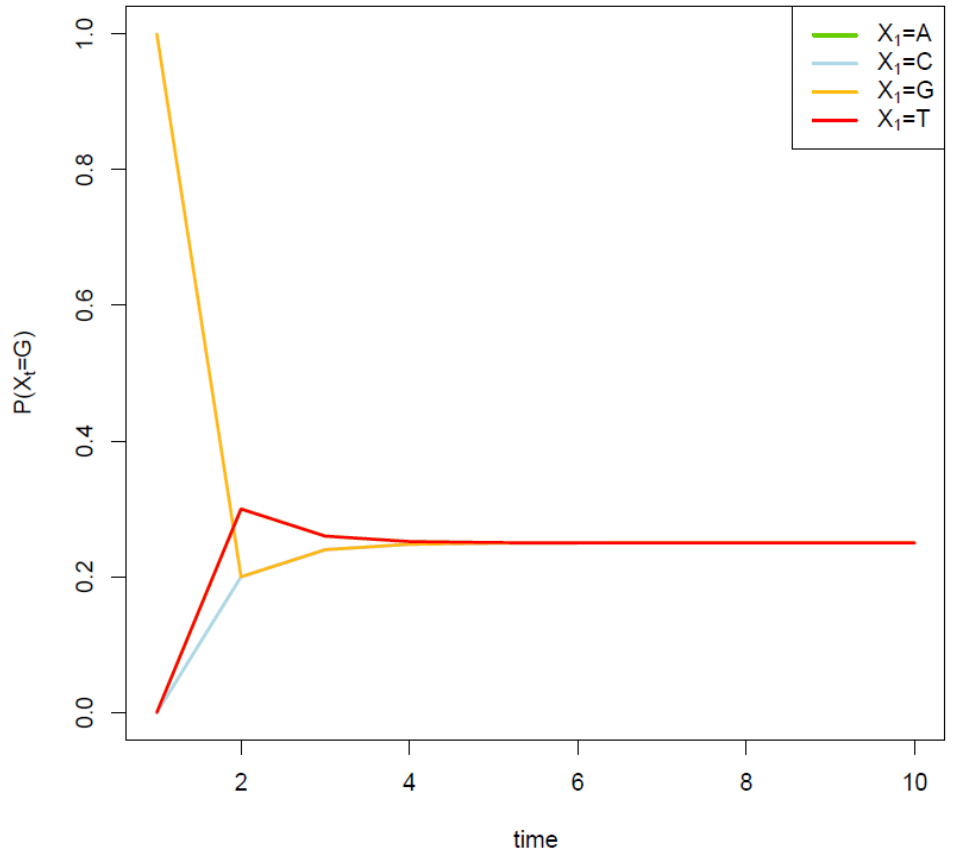
$P(X_2 = \mathbf{G} \mid X_1 = \mathbf{A}) = \ldots$
$P(X_2 = \mathbf{G} \mid X_1 = \mathbf{C}) = \ldots$
$P(X_2 = \mathbf{G} \mid X_1 = \mathbf{G}) = \ldots$
$P(X_2 = \mathbf{G} \mid X_1 = \mathbf{T}) = \ldots$

But also:

$P(X_3 = \mathbf{G} \mid X_1 = \mathbf{A}) = \ldots$
et cetera.

Convergence

Thus, calculate $P(X_t = \mathbf{G} \mid X_1 = x_1)$ with t=2, 3, 4, et cetera for x_1 = **A**, **C**, **G**, **T**.



Study the influence of the first nucleotide on the calculated probability for increasing t.

Convergence

```
> # define π and P
> pi <- matrix(c(1, 0, 0, 0), ncol=1)
> P <- matrix(c(2, 3, 3, 2, 1, 4, 4, 1, 3, 2, 2, 3,
      4, 1, 1, 4), ncol=4, byrow=TRUE)/10

> # define function that calculates the powers of a
> # matrix (inefficiently though)
> matrixPower <- function(X, power){
      Xpower <- X
      for (i in 2:power){
            Xpower <- Xpower %*% X
      }
      return(Xpower)
  }

> # calculate P to the power 100
> matrixPower(P, 100)
```

Convergence

Question

Suppose the DNA may reasonably be described by a first order Markov model with transition matrix **P**:

$$\mathbf{P} = \begin{pmatrix} 0.77450 & 0.22500 & 0.00025 & 0.00025 \\ 0.22500 & 0.77450 & 0.00025 & 0.00025 \\ 0.00025 & 0.00025 & 0.77450 & 0.22500 \\ 0.00025 & 0.00025 & 0.22500 & 0.77450 \end{pmatrix}$$

and stationary distribution:

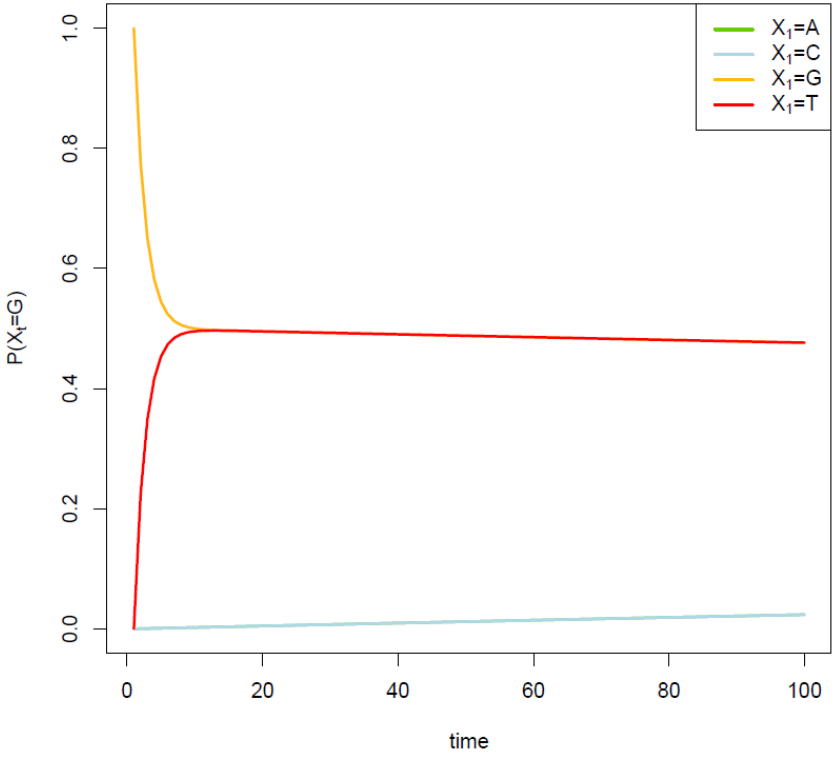
$$\boldsymbol{\varphi}^T = (\varphi_A, \varphi_C, \varphi_G, \varphi_T) = (1/4, 1/4, 1/4, 1/4)$$

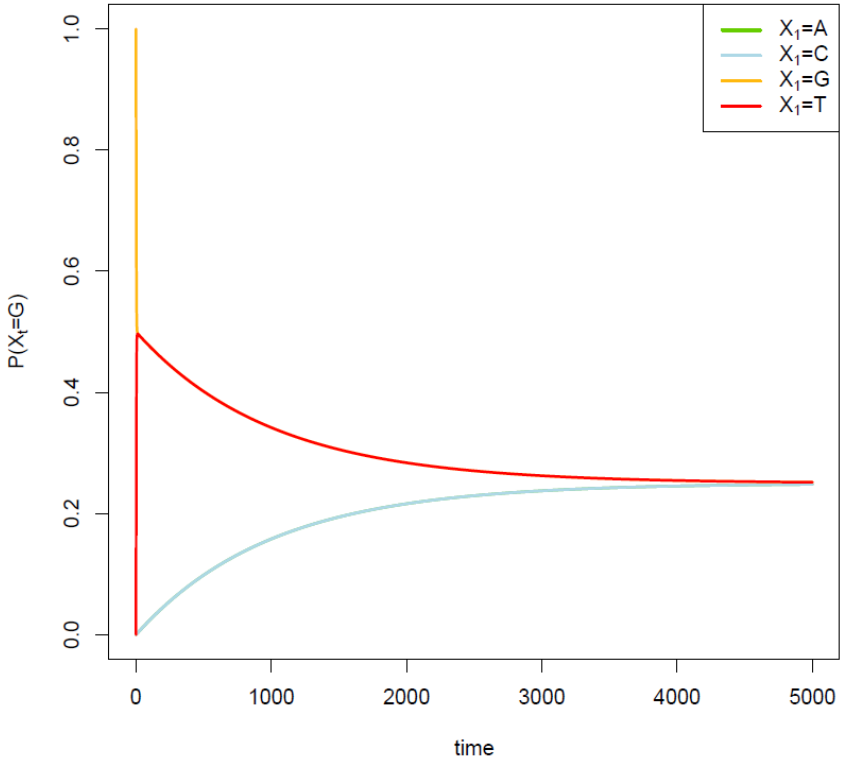
and eigenvalues:

$$\boldsymbol{\lambda} = (1, 0.9990, 0.5495, 0.5495)^T$$

Convergence

Again calculate $P(X_t = \mathbf{G} \mid X_1 = x_1)$ with t=2, 3, 4, 5, et cetera for x_1 = **A**, **C**, **G**, **T**.





Now the influence of the first nucleotide fades slowly. This can be explained by the large 2nd eigenvalue.

Convergence

Question

Is the difference in convergence speed surprising? Think of them generating DNA sequences, while a perturbation replaces an **A** by **T**.

Recall:

$$\mathbf{P} = \begin{pmatrix} 0.2 & 0.3 & 0.3 & 0.2 \\ 0.1 & 0.4 & 0.4 & 0.1 \\ 0.3 & 0.2 & 0.2 & 0.3 \\ 0.4 & 0.1 & 0.1 & 0.4 \end{pmatrix}$$

and

$$\mathbf{P} = \begin{pmatrix} 0.77450 & 0.22500 & 0.00025 & 0.00025 \\ 0.22500 & 0.77450 & 0.00025 & 0.00025 \\ 0.00025 & 0.00025 & 0.77450 & 0.22500 \\ 0.00025 & 0.00025 & 0.22500 & 0.77450 \end{pmatrix}$$

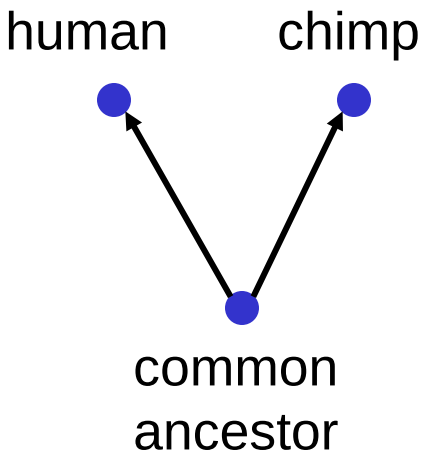
Processes
back in time

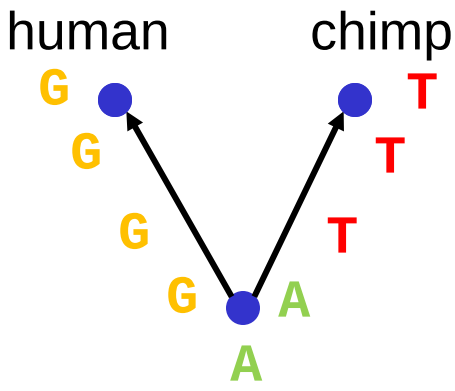
Processes back in time

So far, we have studied Markov chains forward in time. In practice, we may wish to study processes back in time.

Example

Evolutionary models that describe occurrence of SNPs in DNA sequences. We aim to attribute two DNA sequences to a common ancestor.





Processes back in time

Example

In the original time direction the sequence is:

… **CGATACGATGC** … 

In the reverse direction

 … **CGATACGATGC** …

Or,

… **CGTAGCATAGC** … 

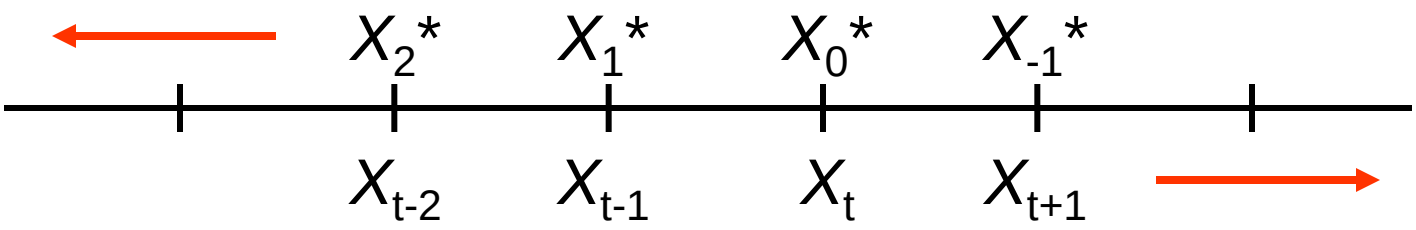
Estimated transition probability from both sequences:

$\widehat{P(\mathbf{A} \mid \mathbf{G})} = 0$ and $\widehat{P(\mathbf{A} \mid \mathbf{G})} = ¼$.

Processes back in time

Consider a Markov chain $\{X_t\}_{t=1,2,\ldots}$. The reverse Markov chain $\{X_r^*\}_{r=1,2,\ldots}$ is then defined by:

$$X_r^* = X_{t-r}$$



With transition probabilities:

$$p_{ij} = P(X_t = E_j \mid X_{t-1} = E_i)$$

$$p_{ij}^* = P(X_r^* = E_j \mid X_{r-1}^* = E_i)$$

Processes back in time

Hence, we now have two Markov processes defined on the same random variables.

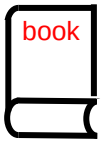
These two Markov processes:
- model the dependency between the variables in different directions, and
- have different parameters.

Example

Evolution occurs as time marches on. Phylogenetics looks in the opposite direction.

Question

How are the two related? Do they give the same “results”?



Axelson-Fisk (2010): Formula 2.24.

Processes back in time

Question

How to relate the transition probabilities in the forward and backward direction?

Show how the transition probabilities p_{ji}^* relate to p_{ij}:

$$p_{ij}^* = P(X_r^* = E_j \mid X_{r-1}^* = E_i)$$

just the definition

book Axelson-Fisk (2010): Formula 2.24.

Processes back in time

Show how the transition probabilities p_{ji}* relate to p_{ij}:

$$\begin{aligned} p^*_{ij} &= P(X^*_r = E_j \mid X^*_{r-1} = E_i) \\ &= P(X_{t-r} = E_j \mid X_{t-r+1} = E_i) \end{aligned}$$

express this in terms of the original Markov chain using that $X_r^* = X_{t-r}$

book Axelson-Fisk (2010): Formula 2.24.

Processes back in time

Show how the transition probabilities p_{ji}^* relate to p_{ij}:

$$
\begin{aligned}
p_{ij}^* &= P(X_r^* = E_j \mid X_{r-1}^* = E_i) \\
&= P(X_{t-r} = E_j \mid X_{t-r+1} = E_i) \\
&= P(X_{t-r+1} = E_i \mid X_{t-r} = E_j) \, \frac{P(X_{t-r} = E_j)}{P(X_{t-r+1} = E_i)}
\end{aligned}
$$

apply definition of conditional probability:
P(A | B) = P(B | A) P(A) / P(B)

book Axelson-Fisk (2010): Formula 2.24.

Processes back in time

Show how the transition probabilities p_{ji}^* relate to p_{ij}:

$$
\begin{aligned}
p_{ij}^* &= P(X_r^* = E_j \mid X_{r-1}^* = E_i) \\
&= P(X_{t-r} = E_j \mid X_{t-r+1} = E_i) \\
&= P(X_{t-r+1} = E_i \mid X_{t-r} = E_j) \, \frac{P(X_{t-r} = E_j)}{P(X_{t-r+1} = E_i)} \\
&= p_{ji} \, \varphi_j \, / \, \varphi_i
\end{aligned}
$$

Hence:

$$
p_{ij}^* = p_{ji} \, \frac{\varphi_j}{\varphi_i}
$$

Processes back in time

Check that rows of the transition matrix **P*** sum to one, i.e.:

$$p_{i1}^* + p_{i2}^* + \ldots + p_{jS}^* = 1$$

Hereto:

$$\sum_{j=1}^{S} p_{ij}^* = \sum_{j=1}^{S} p_{ji} \frac{\varphi_j}{\varphi_i}$$

$$= \frac{1}{\varphi_i} \sum_{j=1}^{S} p_{ji}\, \varphi_j = \frac{1}{\varphi_i} \varphi_i = 1$$

use the fact that $\boldsymbol{\varphi}^T = \boldsymbol{\varphi}^T \mathbf{P}$

Processes back in time

The two Markov chains defined by **P** and **P*** have the same stationary distribution. Indeed, as:

$$\sum_{k=1}^{S} \varphi_k \, p^*_{kj} \quad = \quad \sum_{k=1}^{S} p_{jk} \, \varphi_j$$

$$= \quad \varphi_j \sum_{k=1}^{S} p_{jk} \quad = \quad \varphi_j$$

we have:

$$\boldsymbol{\varphi}^T \quad = \quad \boldsymbol{\varphi}^T \mathbf{P}^*$$

Question

Why is this no surprise? Consider a “stationary DNA sequence”.

Processes back in time

Definition

A Markov chain is called *reversible* if the forward and backward process are 1-1 related.

More precisely, if $p_{ij}^* = p_{ij}$.

In that case:

$$p_{ij}^* = p_{ij} = p_{ji}\, \varphi_j / \varphi_i$$

Or,

$$\varphi_i\, p_{ij} = \varphi_j\, p_{ji} \qquad \text{for all i and j.}$$

These are the so-called *detailed balance equations*.

Theorem

A Markov chain is reversible if and only if the detailed balance equations hold.

Processes back in time

A closer look at the *detailed balance equations*:

$$(\mathbf{P})_{ij}\,\varphi_i \quad = \quad (\mathbf{P})_{ji}\,\varphi_j$$

Rewritten:

$$P(X_{t+1} = E_j \,|\, X_t = E_i)\, P(X_t = E_i)$$
$$= \quad P(X_{t+1} = E_i \,|\, X_t = E_j) P(X_t = E_j)$$

Or:

$$P(X_{t+1} = E_j, X_t = E_i) \quad = \quad P(X_{t+1} = E_i, X_t = E_j)$$

Interpretation

It is irrelevant whether one goes from state E_j to state E_i or vice versa.

Processes back in time

Example 1

The 1st order Markov chain with transition matrix:

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

is irreversible. Check that this (deterministic) Markov chain does not satisfy the detailed balance equations.

Irreversibility can be seen from a sample of this chain:

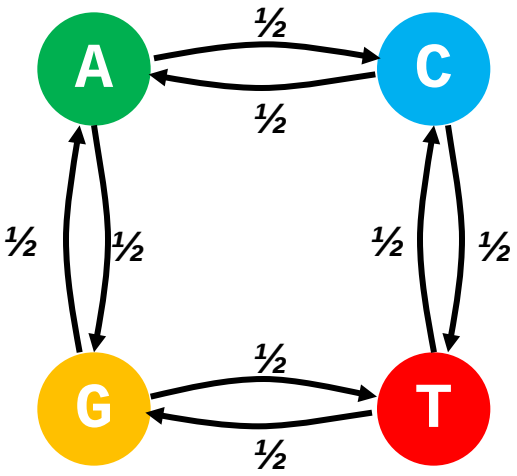
… ABCABCABCABCABCABC…

In the reverse direction transitions from **B** to **C** do not occur!

Processes back in time

Question

Consider a “stationary DNA sequence”, which may be modelled by a 1st order Markov process with transition matrix:



The process has a uniform stationary distribution.

Is this Markov process reversible?

Processes back in time

Example 2

The 1st order Markov chain with transition matrix:

$$\mathbf{P} = \begin{pmatrix} 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \\ 0.8 & 0.1 & 0.1 \end{pmatrix}$$

is irreversible. Again, a uniform stationary distribution:

$$\boldsymbol{\varphi}^T = (\varphi_A, \varphi_B, \varphi_C) = (1/3, 1/3, 1/3)$$

As **P** is not symmetric, the detailed balance equations are not satisfied:

$p_{ij} / 3 \neq p_{ji} / 3$ for all i and j.

Processes back in time

Example 2 (continued)

The irreversibility of this chain implies:

$P(X_t=\mathbf{A}, X_{t+1}=\mathbf{B}, X_{t+2}=\mathbf{C}, X_{t+3}=\mathbf{A})$

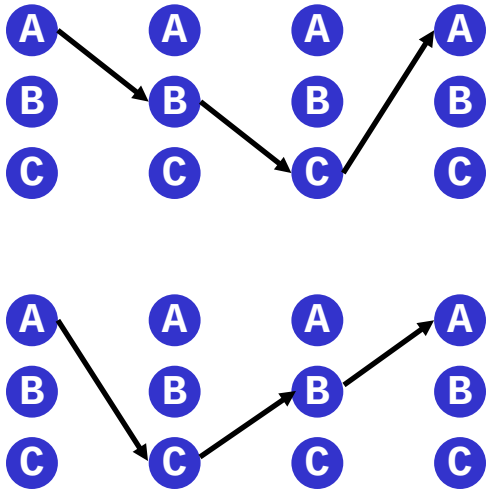
$= \mathbf{P(A)\ P(B|A)\ P(C|B)\ P(A|C)}$

$= 1/3 * 0.8 * 0.8 * 0.8$

$\neq 1/3 * 0.1 * 0.1 * 0.1$

$= \mathbf{P(A)\ P(C|A)\ P(B|C)\ P(A|B)}$

$= P(X_t=\mathbf{A}, X_{t+1}=\mathbf{C}, X_{t+2}=\mathbf{B}, X_{t+3}=\mathbf{A})$.



It matters how one walks from **A** to **A**.

Or, it matters whether one walks forward or backward.

Processes back in time

Kolmogorov condition for reversibility

A stationary Markov chain is reversible if and only if any path from state E_i to state E_i has the same probability as the path in the opposite direction.

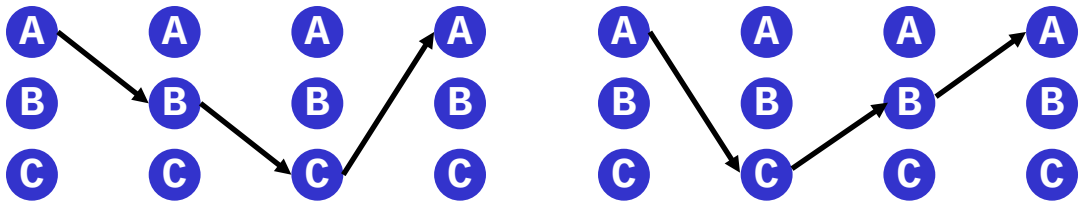
Or, the Markov chain is reversible if and only if:

$$\varphi_i \, p_{i,i_1} \cdot p_{i_1,i_2} \cdot \ldots \cdot p_{i_k,i} \quad = \quad \varphi_i \, p_{i,i_k} \cdot p_{i_2,i_1} \cdot \ldots \cdot p_{i_1,i}$$

for all i, i_1, i_2, …, i_k.

E.g.: P(X_t=**A**,X_{t+1}=**B**,X_{t+2}=**C**, X_{t+3}=**A**)

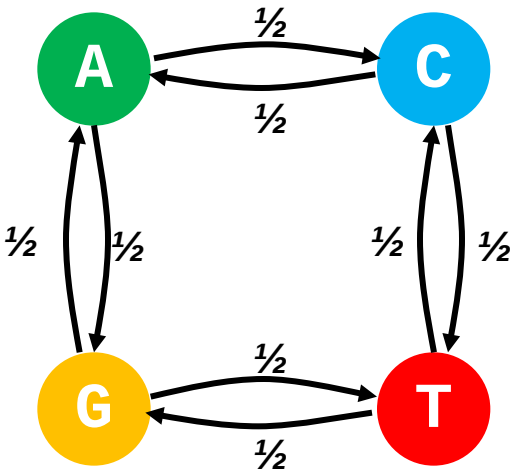
= P(X_t=**A**,X_{t+1}=**C**,X_{t+2}=**B**, X_{t+3}=**A**).



Processes back in time

Question (revisited)

Consider a “stationary DNA sequence”, which may be modelled by a 1st order Markov process with transition matrix:



The process has a uniform stationary distribution.

Is this Markov process reversible? Use Kolmogorov condition.

Processes back in time

Interpretation of Kolmogorov condition

For a reversible Markov chain it is impossible to determine the direction of the process from the observed state sequence alone.

- Molecular phylogenetics reconstructs evolutionary relationships among present day species from DNA sequences. Reversibility is an essential assumption.
- Genes are transcribed in one direction only (from the 3' end to the 5' end). The promoter is on the 3' end, which suggests irreversibility.

Codons **CAG** and **GAC** encode for an Glutamine and Aspartic amino acid, respectively.

Processes back in time

E. Coli

For a gene in the E. Coli genome, we estimate:

```
Transition matrix
      [,1]      [,2]      [,3]      [,4]
[1,] 0.2296984 0.3155452 0.2273782 0.2273782
[2,] 0.1929134 0.2421260 0.2933071 0.2716535
[3,] 0.1979167 0.2854167 0.2166667 0.3000000
[4,] 0.2522686 0.2032668 0.2341198 0.3103448

Stationary distribution
[1] 0.2187817 0.2578680 0.2436548 0.2796954
```

Then, the detailed balance equation do not hold, e.g.:

$\pi_1 p_{12} \neq \pi_2 p_{21}$.

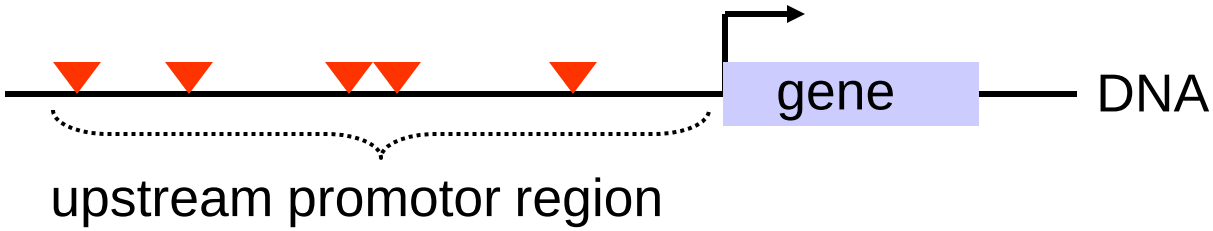
Example

---

motif discovery

Example: motif discovery

A gene's promotor region contains binding sites for the transcription factors that regulate its transcription.



These binding sites (that may regulate multiple genes) share certain sequence patterns, *motifs*.

Not all transcription factors and motifs are known. A high occurrence of a particular sequence pattern in a gene's upstream regions may indicate that it has a regulatory function.

Question

When does a motif occurs exceptionally frequent?

Example: motif discovery

An *h*-letter *word* $W = w_1\ w_2\ \ldots\ w_h$ is a map from {1, …, *h*} to $\mathcal{A}^h$, where $\mathcal{A}$ some non-empty set, called the *alphabet*.

In the DNA example:

$$\mathcal{A} = \{\mathtt{A},\ \mathtt{C},\ \mathtt{G},\ \mathtt{T}\}$$

and, e.g.:

$$W = \mathtt{CAGTACGACT}$$

$$W = \mathtt{TACGACTGCATATGCGTA}$$

are, respectively, 10- and 18-letter words.

Example: motif discovery

Let $N(W)$ be the number of occurrences of an h-letter word W in a random sequence of length n.

If $Y_t(W)$ is the random variable defined by:

$$Y_t = I_{\{\text{an occurence of } W \text{ starts at position } t\}}$$

then

$$N(W) = \sum_{t=1}^{n-h+1} Y_t$$

For W = **GTA** and sequence **GCTCGTAAACTGCT**,
$Y_t(W)$=0 for all t, except for t = 5: $Y_5(W)$=1.

Example: motif discovery

Problem

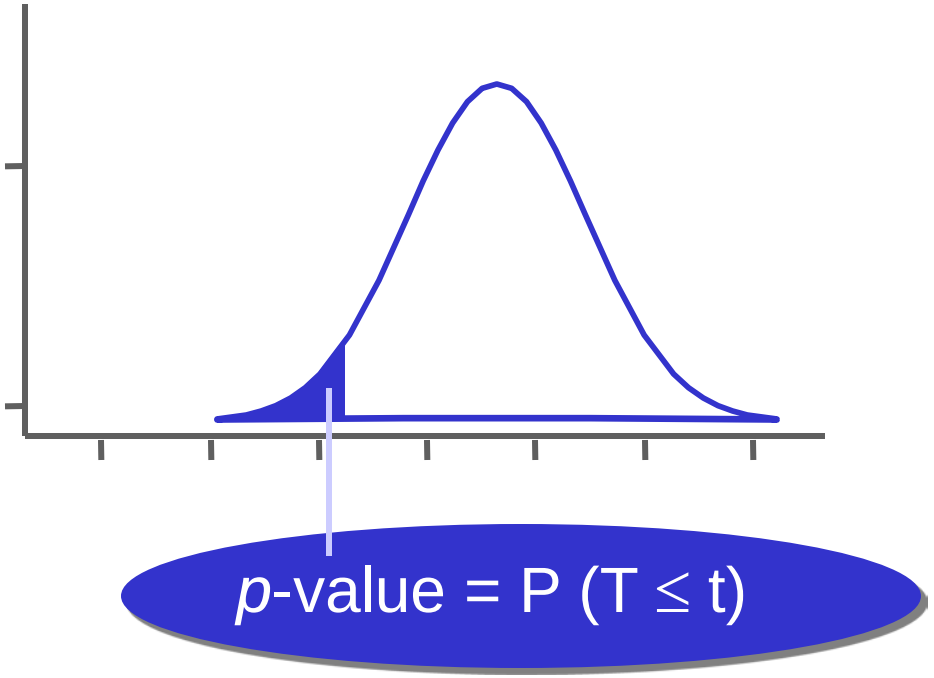
Determine the probability of observing *m* motifs in a background generated by a 1st order stationary Markov chain.

Compare the observed frequency to the expected one through hypothesis testing:

→ Summarize evidence in test statistics

→ Compare test statistics to reference distribution.

→ Calculate p-value.



Example: motif discovery

The random variable $Y_t(W)$ is Bernouilli distributed with parameter $\mu(W) = P(Y_t(W) = 1)$.

Assume a stationary 1st order Markov model for the random sequence of length n. The probability of an occurrence of W in the random sequence is given by:

$$\mu(W) = \mu(w_1) \prod_{t=1}^{h-1} (\mathbf{P})_{w_t, w_{t+1}}$$

Furthermore, the variance of $Y_t(W)$ is:

$$\mu(W)[1 - \mu(W)]$$

Example: motif discovery

The expectation and variance of the word count *N*(*W*) are then:

$$\begin{aligned} E[N(W)] &= \sum_{t=1}^{n-h+1} E[Y_t(W)] \\ &= (n-h+1)\,\mu(W) \end{aligned}$$

and:

$$\mathrm{Var}[N(W)] = (n-h+1)\,\mu(W)[1-\mu(W)]$$

Example: motif discovery

To find words with exceptionally frequency in the DNA, the following (asymptotically) standard normal statistic is used:

$$Z(W) = \frac{N(W) - E[N(W)]}{\sqrt{\mathrm{Var}[N(W)]}} \sim N(0,1)$$

The *p*-value of word *W* is then given by:

$$\begin{aligned} p(W) &= P(Z \geq Z(W)) \\ &= 1 - \Phi_{0,1}(Z(W)) \end{aligned}$$

Example: motif discovery

E. Coli sequence

In a non-coding region, length 557 bases, of the E. Coli genome the motif **TTAA** has been found 12 times.

For this region, the nucleotide percentage is estimated as

A	C	G	T
0.3196	0.1364	0.1993	0.3447

and the transition matrix as:

	A	C	G	T
A	0.3427	0.0899	0.1742	0.3933
C	0.3816	0.1316	0.1711	0.3158
G	0.3333	0.2072	0.2162	0.2432
T	0.2670	0.1414	0.2199	0.3717

Example: motif discovery

E. Coli sequence

Then:

P(TTAA) = 0.3447 * 0.3717 * 0.2670 * 0.3427
= 0.0117

Consequently,

E[N(**TTAA**)] = 6.496, Var[N(**TTAA**)] = 6.420,

and the test statistic equals Z(**TTAA**) = 2.172.

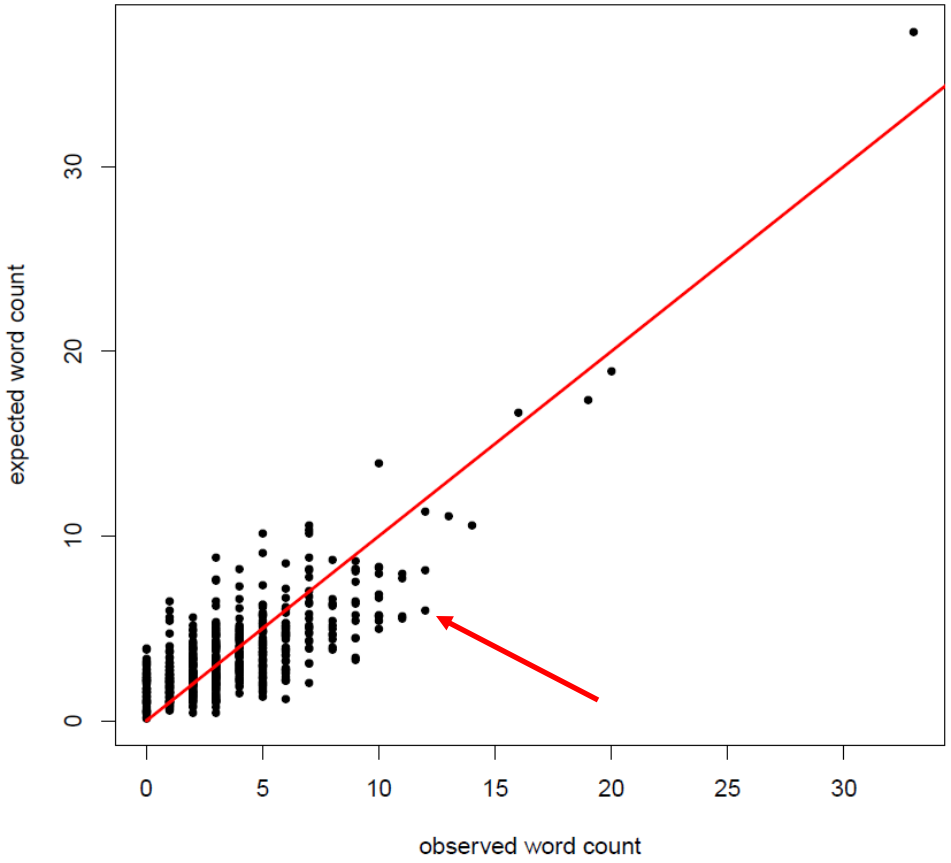
Using the standard normal approximation we find:

p-value = 0.0149

Example: motif discovery

E. Coli sequence

Compare the observed and expected word count for non-coding regions larger than 250 bases.



Example: motif discovery

Notes

Words are assumed non-periodic, i.e. “they do not repeat themselves”. For instance, W_1 = **CGATCGATCG** is periodic:

```
123456789
CGATCGATC
    CGATCGATC
        CGATCGATC
```

Periodic motif: variance of its test statistic needs modification.

Robin, Daudon (1999) provide exact probabilities of word occurrences in random sequences. However, calculation of the exact probabilities is computationally intensive (Robin, Schbath, 2001). Hence, the use of an approximation here.

Supplementary material:
Spectral decomposition
of the *n*-step
transition matrix

Convergence to the stationary distribution

We now obtain the spectral decomposition of the *n*-step transition matrix $\mathbf{P}^n$. Hereto observe that:

$$\mathbf{P}^n \mathbf{r}_k = \mathbf{P}^{n-1} \left(\sum_{k_0=1}^{S} \lambda_{k_0} \mathbf{r}_{k_0} \boldsymbol{\ell}_{k_0}^T \right) \mathbf{r}_k$$

plug in the spectral decomposition of **P**

Convergence to the stationary distribution

We now obtain the spectral decomposition of the *n*-step transition matrix $\mathbf{P}^n$. Hereto observe that:

$$\mathbf{P}^n \mathbf{r}_k = \mathbf{P}^{n-1} \left(\sum_{k_0=1}^{S} \lambda_{k_0} \mathbf{r}_{k_0} \boldsymbol{\ell}_{k_0}^T \right) \mathbf{r}_k$$

$$= \mathbf{P}^{n-1} \left(\sum_{k_0=1}^{S} \lambda_{k_0} \mathbf{r}_{k_0} [\mathbf{r}_k^T \boldsymbol{\ell}_{k_0}]^T \right)$$

bring the right eigenvector in the sum,
and use the properties of the transpose operator

Convergence to the stationary distribution

We now obtain the spectral decomposition of the *n*-step transition matrix $\mathbf{P}^n$. Hereto observe that:

$$\begin{aligned}
\mathbf{P}^n \mathbf{r}_k &= \mathbf{P}^{n-1} \Big(\sum_{k_0=1}^{S} \lambda_{k_0} \mathbf{r}_{k_0} \boldsymbol{\ell}_{k_0}^T \Big) \mathbf{r}_k \\
&= \mathbf{P}^{n-1} \Big(\sum_{k_0=1}^{S} \lambda_{k_0} \mathbf{r}_{k_0} [\mathbf{r}_k^T \boldsymbol{\ell}_{k_0}]^T \Big) \\
&= \mathbf{P}^{n-1} \lambda_k \mathbf{r}_k
\end{aligned}$$

the eigenvectors are normalized

Convergence to the stationary distribution

We now obtain the spectral decomposition of the *n*-step transition matrix $\mathbf{P}^n$. Hereto observe that:

$$
\begin{aligned}
\mathbf{P}^n \, \mathbf{r}_k &= \mathbf{P}^{n-1} \Big(\sum_{k_0=1}^{S} \lambda_{k_0} \, \mathbf{r}_{k_0} \, \boldsymbol{\ell}_{k_0}^T \Big) \mathbf{r}_k \\
&= \mathbf{P}^{n-1} \Big(\sum_{k_0=1}^{S} \lambda_{k_0} \, \mathbf{r}_{k_0} \, [\mathbf{r}_k^T \, \boldsymbol{\ell}_{k_0}]^T \Big) \\
&= \mathbf{P}^{n-1} \, \lambda_k \, \mathbf{r}_k \\
&= \lambda_k \, \mathbf{P}^{n-1} \, \mathbf{r}_k
\end{aligned}
$$

Convergence to the stationary distribution

Verify the spectral decomposition for $\mathbf{P}^2$:

$$
\begin{aligned}
\mathbf{P}^2 &= \left(\sum_{k=1}^{S} \lambda_k \, \mathbf{r}_k \, \boldsymbol{\ell}_k^T\right) \left(\sum_{k=1}^{S} \lambda_k \, \mathbf{r}_k \, \boldsymbol{\ell}_k^T\right) \\
&= \sum_{k_1=1}^{S} \sum_{k_2=1}^{S} \lambda_{k_1} \, \mathbf{r}_{k_1} \, \boldsymbol{\ell}_{k_1}^T \, \lambda_{k_2} \, \mathbf{r}_{k_2} \, \boldsymbol{\ell}_{k_2}^T \\
&= \sum_{k_1=1}^{S} \sum_{k_2=1}^{S} \lambda_{k_1} \, \lambda_{k_2} \, \mathbf{r}_{k_1} \, \boldsymbol{\ell}_{k_1}^T \, \mathbf{r}_{k_2} \, \boldsymbol{\ell}_{k_2}^T \\
&= \sum_{k_1=1}^{S} \sum_{k_2=1}^{S} \lambda_{k_1} \, \lambda_{k_2} \, \mathbf{r}_{k_1} \, (\mathbf{r}_{k_2}^T \, \boldsymbol{\ell}_{k_1})^T \, \boldsymbol{\ell}_{k_2}^T \\
&= \sum_{k=1}^{S} \lambda_k^2 \, \mathbf{r}_k \, \boldsymbol{\ell}_k^T
\end{aligned}
$$

Supplementary material:
Irreversibility in
evolutionary theory

Processes back in time

Note

Within evolution theory the notion of irreversibility refers to the presumption that complex organisms once lost evolution will not appear in the same form.

Indeed, the likelihood of reconstructing a particular phylogenic system is infinitesimal small.

References & further reading

References and further reading

Ewens, W.J, Grant, G (2006), *Statistical Methods for Bioinformatics*, Springer, New York.

Reinert, G., Schbath, S., Waterman, M.S. (2000), “Probabilistic and statistical properties of words: an overview”, *Journal of Computational Biology*, **7**, 1-46.

Robin, S., Daudin, J.-J. (1999), “Exact distribution of word occurrences in a random sequence of letters”, *Journal of Applied Probability,* **36**, 179-193.

Robin, S., Schbath, S. (2001), “Numerical comparison of several approximations of the word count distribution in random sequences”, *Journal of Computational Biology*, **8**(4), 349-359.

Schbath, S. (2000), “An overview on the distribution of word counts in Markov chains”, *Journal of Computational. Biology,* **7**, 193-202.

Schbath, S., Robin, R. (2009), “How can pattern statistics be useful for DNA motif discovery?”. In *Scan Statistics: Methods and Applications* by Glaz, J. *et al*. (eds.).

Wodarz, D., Komarova, N.L. (2008), *Computational biology of cancer: lecture notes and mathematical modeling*, Singapore: World Scientific.