ILLC Project Course in Information Theory

Crash course

13 January – 17 January 2014
12:00 to 14:00

Student presentations

27 January – 31 January 2014
12:00 to 14:00

Location

ILLC, room F1.15,
Science Park 107, Amsterdam

Materials

informationtheory.weebly.com

Contact

Mathias Winther Madsen
mathias.winther@gmail.com

Monday

Probability theory
Uncertainty and coding

Tuesday

The weak law of large numbers
The source coding theorem

Wednesday

Random processes
Arithmetic coding

Thursday

Divergence
Kelly Gambling

Friday

Kolmogorov Complexity
The limits of statistics

PLAN

- Random processes
- Markov processes
- Stationary distributions
- Entropy rates
- Ergodicity
- The source coding theorem for random processes

A **random process** is a series of random variables $X_1, X_2, X_3, \ldots$.
In each possible word, every variable in the sequence is assigned a specific value.

w	X_1	X_2	X_3	X_4	X_5	X_6	…
w_1	0	1	2	1	2	3	…
w_2	0	–1	0	–1	–2	–3	…
w_3	0	1	2	3	2	1	…

Choose *n* according to a geometric distribution; randomly pick one of the letters a, b, and c and print the letter *n* times; repeat.

babbbbbccccabaaaaaaababbbbaac ...

Let *s* be either a or b. Print *s* and replace it by the the letter that you didn't choose; repeat.

bababababababababababababababa ...

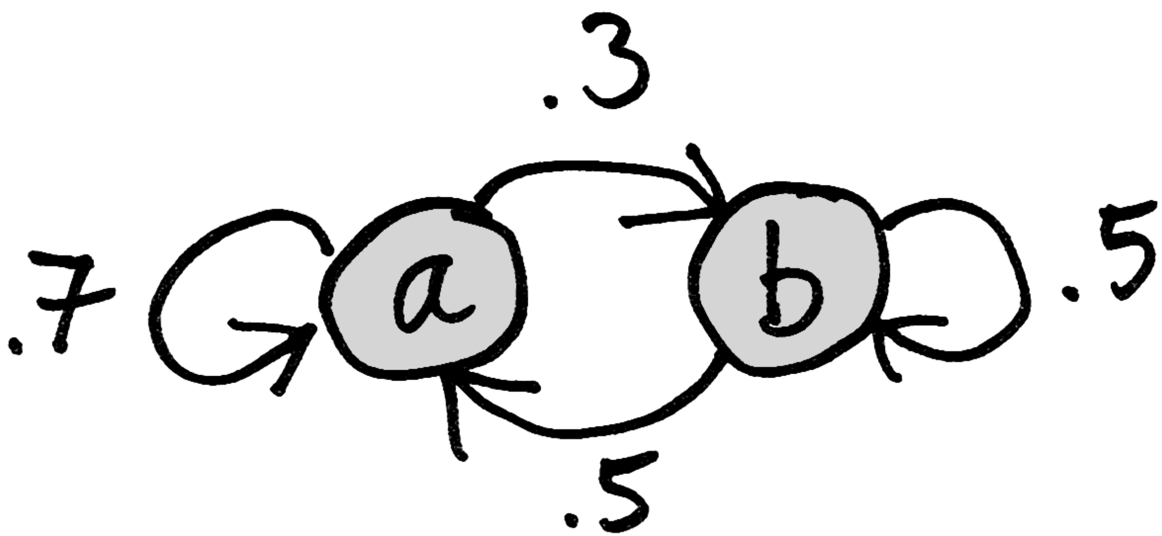
Choose *n* according to a geometric distribution; choose a random word of length *n* and print it twice; repeat.

ccccaaaaccbbababababacbacccc ...

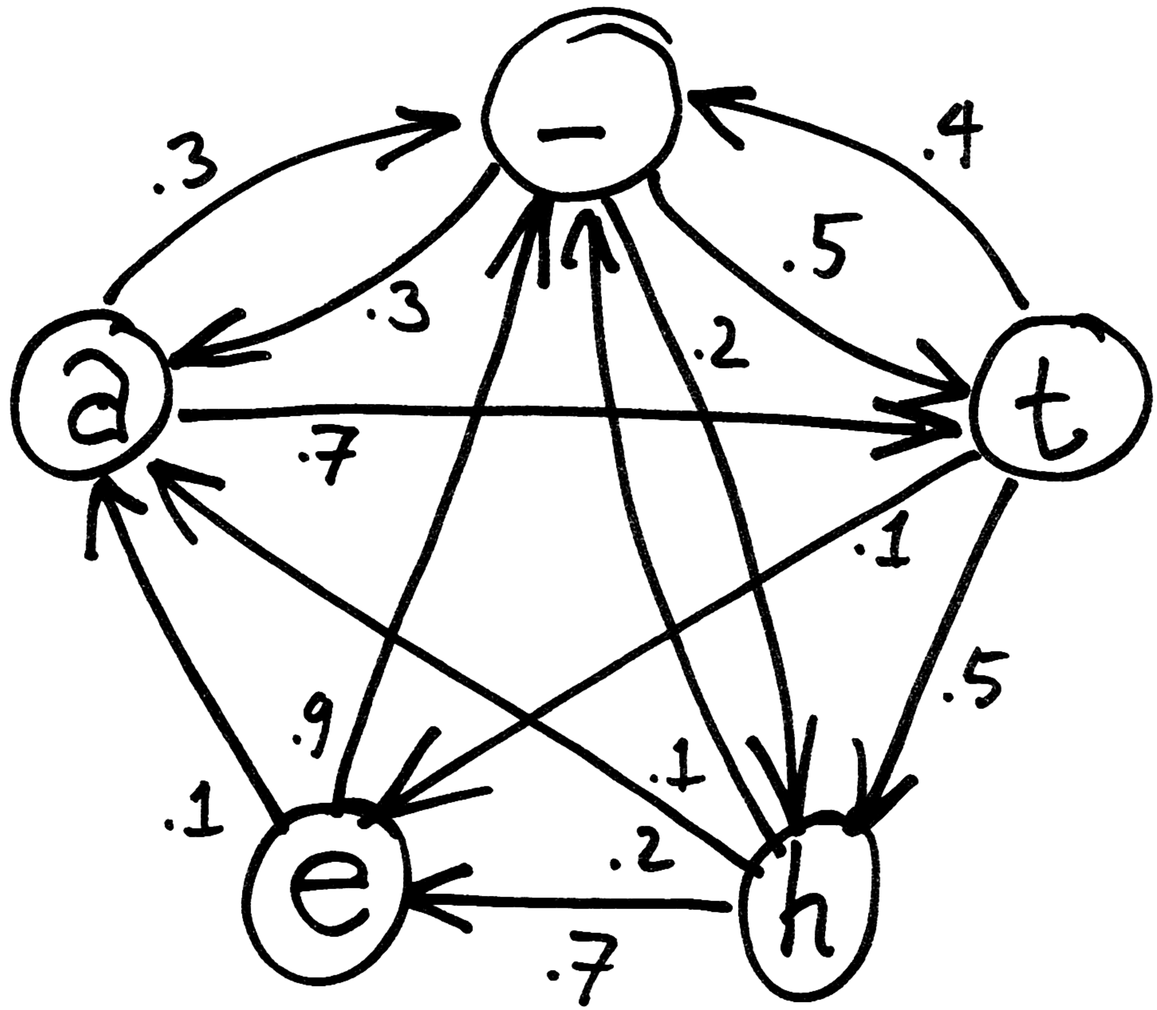
Markov chains

Choose a letter according to some initial distribution and remember your choice. Then choose the next latter based on a conditional distribution given your last choice; repeat.

bbaaaaaaaaaaaaaaabbaabbbbababa …



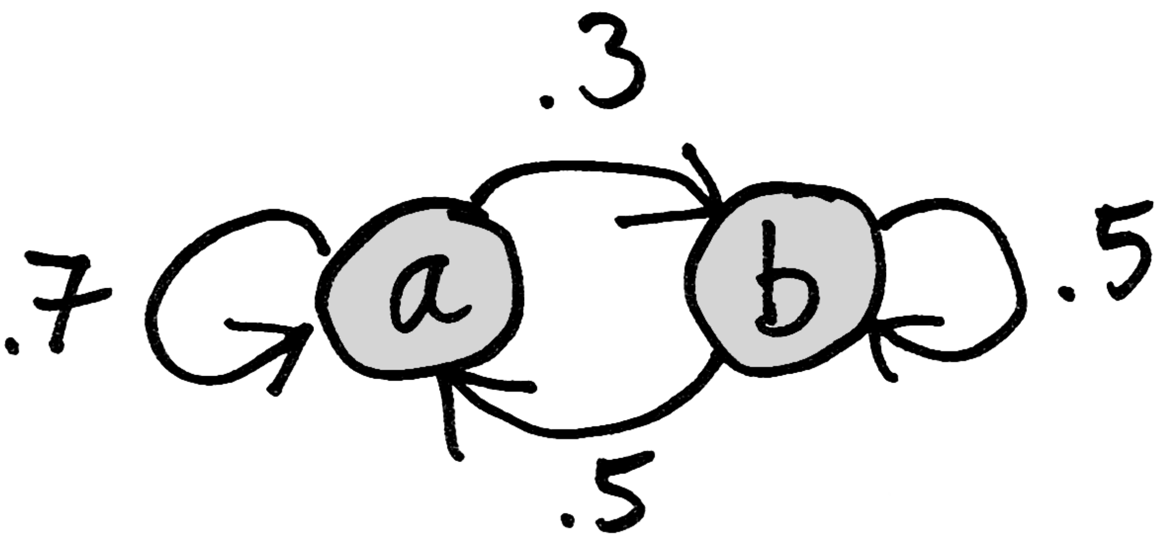
$$
\begin{array}{c} \begin{matrix} a & b \\ \downarrow & \downarrow \end{matrix} \\ \begin{pmatrix} .7 & .5 \\ .3 & .5 \end{pmatrix} \end{array}
\begin{array}{l} \\ \\ \rightarrow a \\ \rightarrow b \end{array}
$$



t_ate_t_he_te_the_the_that_t_te_
athe_at_athe_t_athe_te_ath_th_a_
a_the_the_thatea_the_he_a_t_ …

Stationary distributions

If a distribution is a f xed point for the time transition we call is a stationary distribution. These distribution thus remain unchanged if you take one step ahead in the future: $TP = P$.



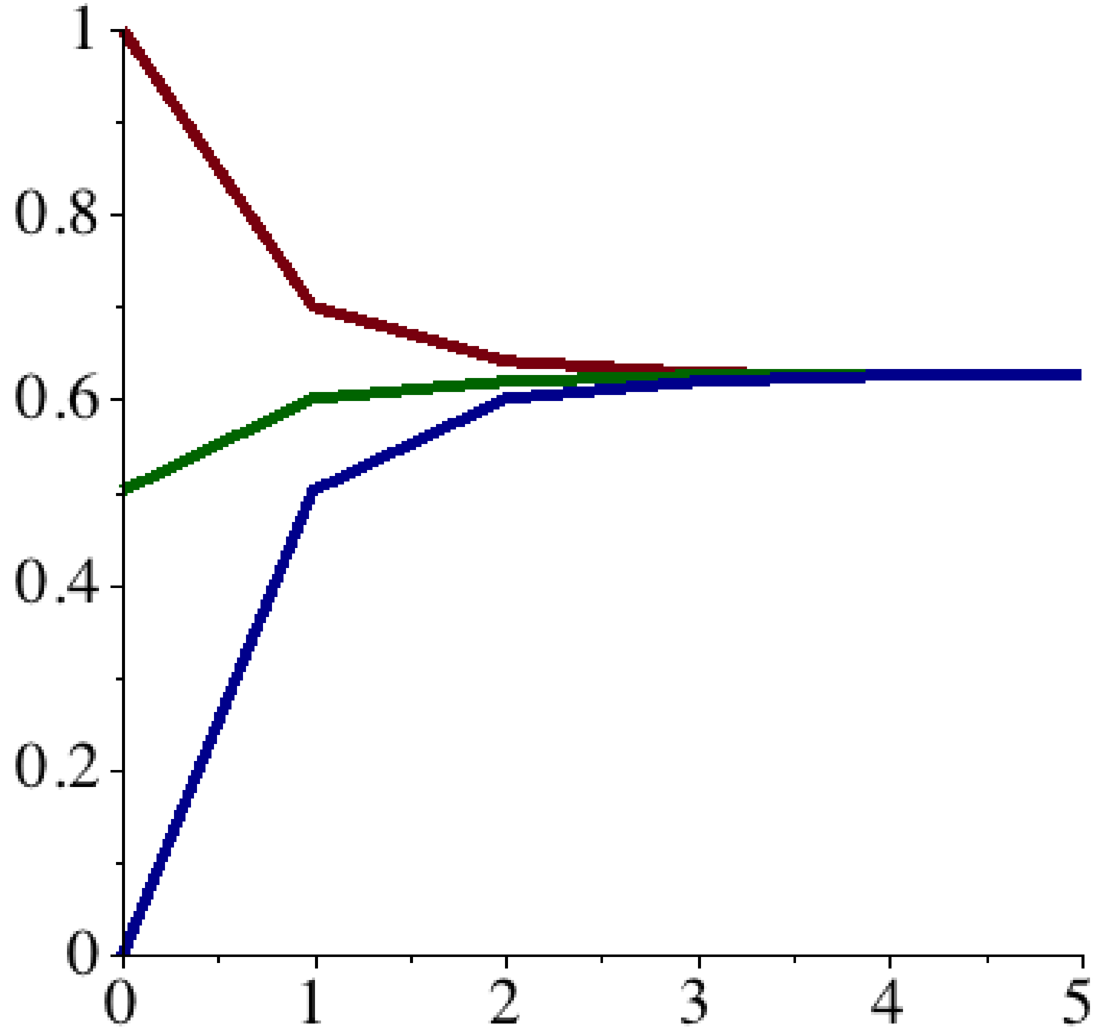
$$\begin{array}{cc} \begin{matrix} a & b \\ \downarrow & \downarrow \end{matrix} & \\ \begin{pmatrix} .7 & .5 \\ .3 & .5 \end{pmatrix} & \begin{matrix} \to a \\ \to b \end{matrix} \end{array}$$

Stationary distributions

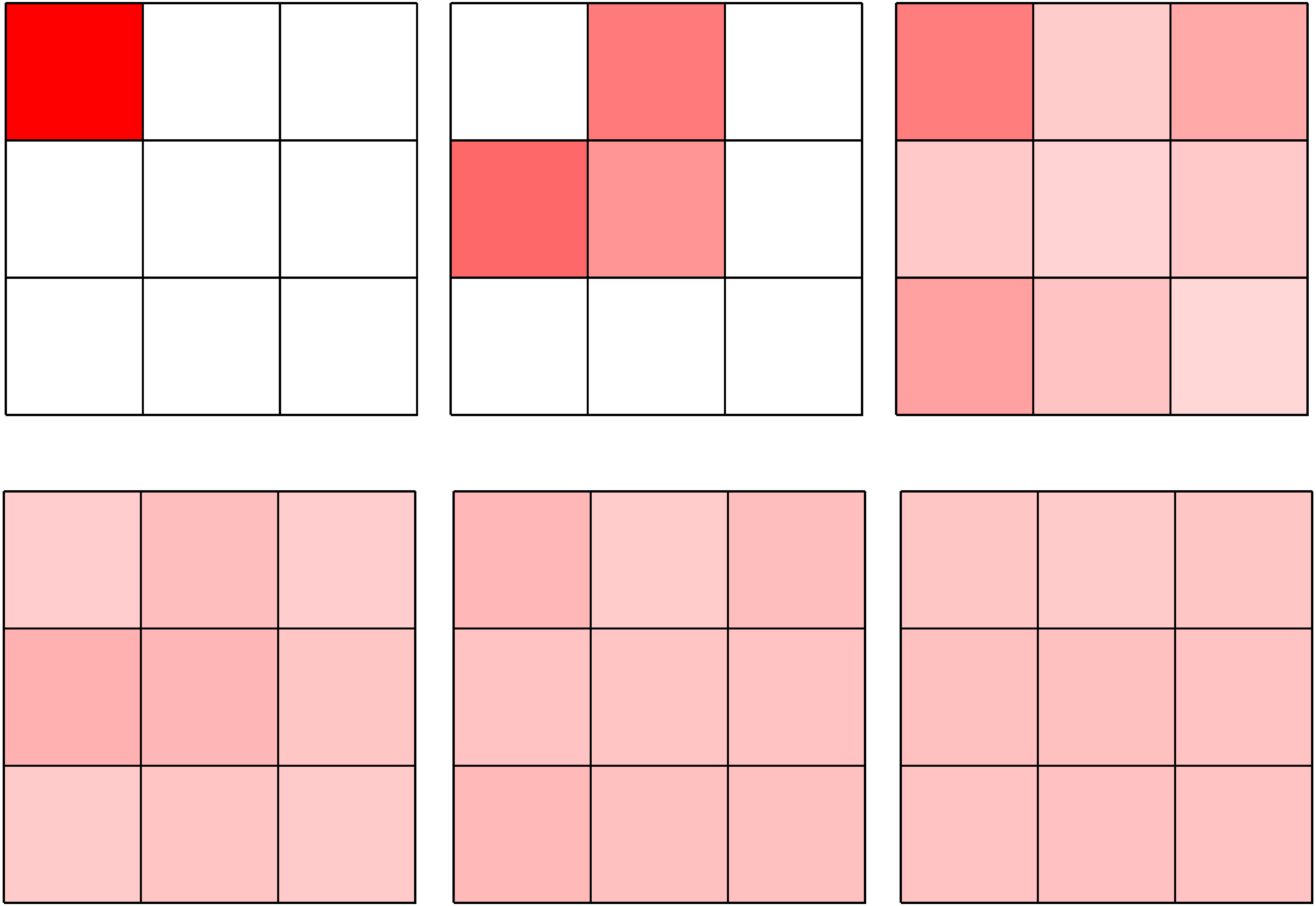
A Markov process converges to a unique stationary distribution on all sample paths if both of the following conditions are met:

- All states are connected by a path with positive probability;
- The greatest common divisor of the length of the cycles in the state graph is 1.

Pr('a')



.625



Entropy rate

A reminder:

The conditional entropy $H(X \mid Y)$ is the expected value of the posterior entropy $H(X \mid Y = y)$.

For instance,

$$H(X \mid Y) = H(X \mid Y = 0)\, p(0) + H(X \mid Y = 1)\, p(1)$$

where $p(y)$ is shorthand for $\Pr(Y = y)$, and we have assumed that Y only takes the two values $Y = 0$ and $Y = 1$.

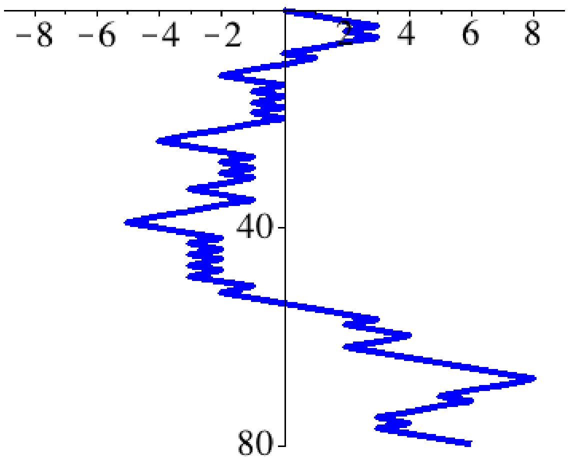
Entropy rate

For a stochastic process $X_1, X_2, X_3, \ldots,$ the entropy rate is the limit of the conditional entropies

$$H(X_n \mid X_1, X_2, X_3, \ldots, X_{n-1})$$

for $n \to \infty$ (when such a limit exists).

Entropy rate



A dust particle starts at $X = 0$ and randomly moves left or right.

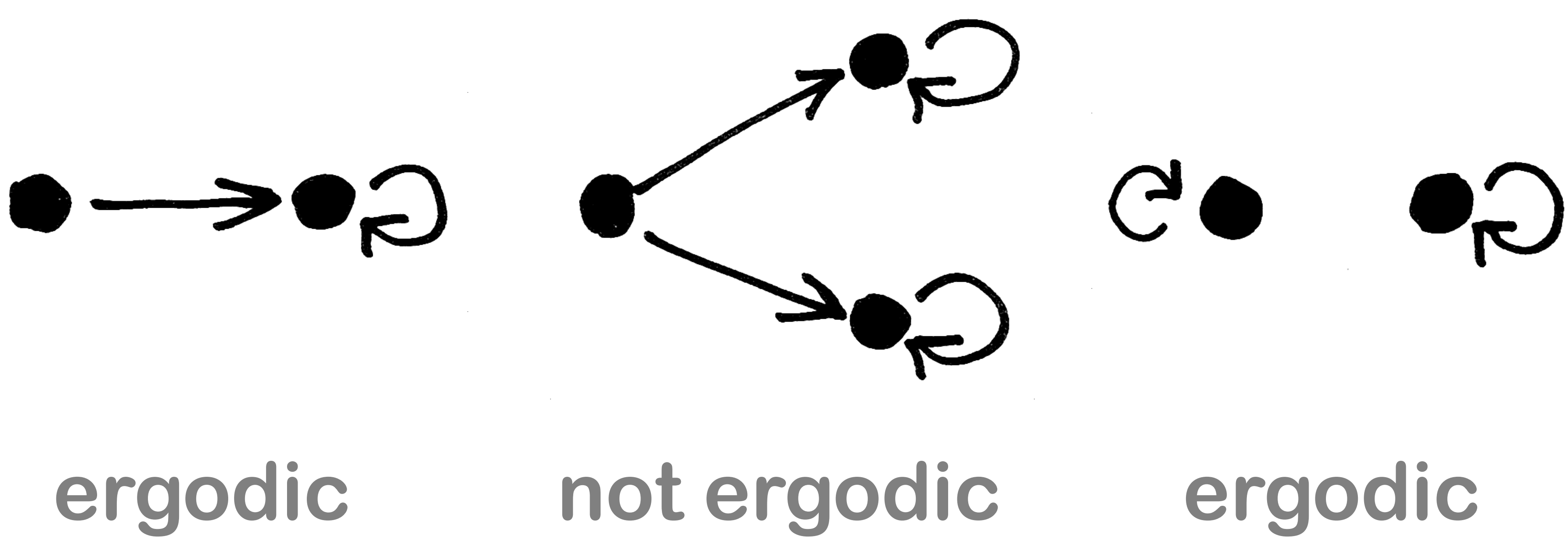
0, –1, –2, –3, –2, –1, 0, 1, 0, –1, –2, –1, 0, –1, 0, –1, 0, –1, …

Choose an n according to a geometric distribution; then a letter from {a, b, c, d}; print it n times.

aadbdccccdccccccdaabbbbbaaacccccccccdddbbbbddabdbbbccccbaaaddbb …

Let's call a set of states is a **trapping set** if being inside that set at time t implies that you are also inside that set at time $t + 1$.

A random process is then called **ergodic** if all its trapping sets have probability 0 or 1.



Birkoff's Ergodicity Theorem

Suppose each state in the state space is associated with a certain reward, and that we are interested in its long-term average (reward per unit of time).

If the process is ergodic, then the expected reward is the same on all sample paths.

George David Birkhoff: "Proof of the ergodic theorem," Proceedings of the National Academy of Sciences of the USA, 1931.

Birkoff's Ergodicity Theorem

If the process is ergodic **and** stationary, then an entropy rate exists.

If the process is ergodic **and** converges to a stationary distribution for any sample path, then an entropy rate also exists.

Random walk with gravity A molecule moves around in a glass of water which we consider as divided up into three compartments. Whenever possible, the molecule moves one compartment down with probability $1/5$, and one compartment up with probability $1/20$.

1. Write down the transition probabilities associated with this system in an exhaustive and explicit fashion.
2. Find the associated equilibrium distribution.
3. What would you guess the equilibrium distribution would look if we had started with k compartments instead of three?

Tiny chess What's the entropy rate of a knight walking on a 3×3 chess board?

What about a bishop?

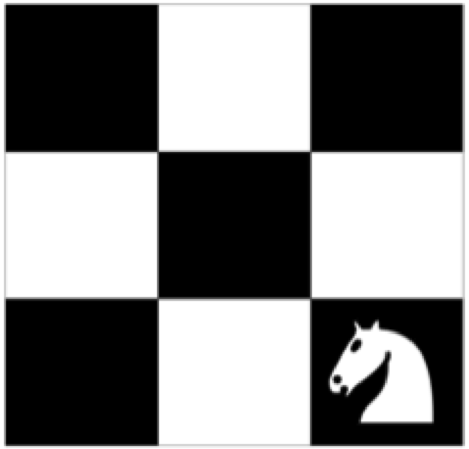


Figure 1: A knight on a 3×3 chess board.

Morse code (Cover and Thomas, Ex. 4.8) An alphabet contains a dot which takes one unit of time to transmit, and a dash which takes two.

1. When the two symbols have probability p and $q = 1 - p$, what's the entropy rate of this process?
2. For which choice of p and q is this entropy rate the largest?