Perfect simulation and coupling from the past

Kjartan Kloster Osmundsen

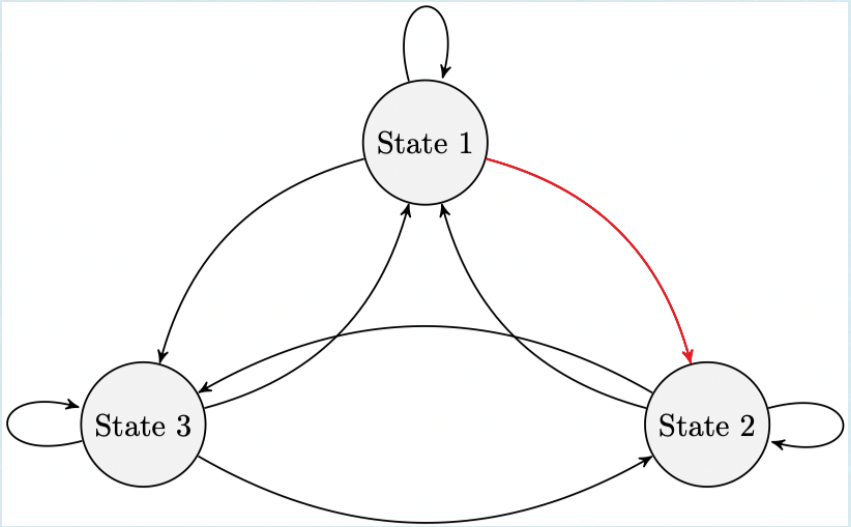
University of Stavanger

Trial lecture

Stavanger, June 10th, 2020

Markov chain

- Consider a finite number of *states*: $S = \{s_1, s_2, s_3\}$.
- Next, let a *transition matrix* $\boldsymbol{P}$ describe how likely it is to move between the different states:



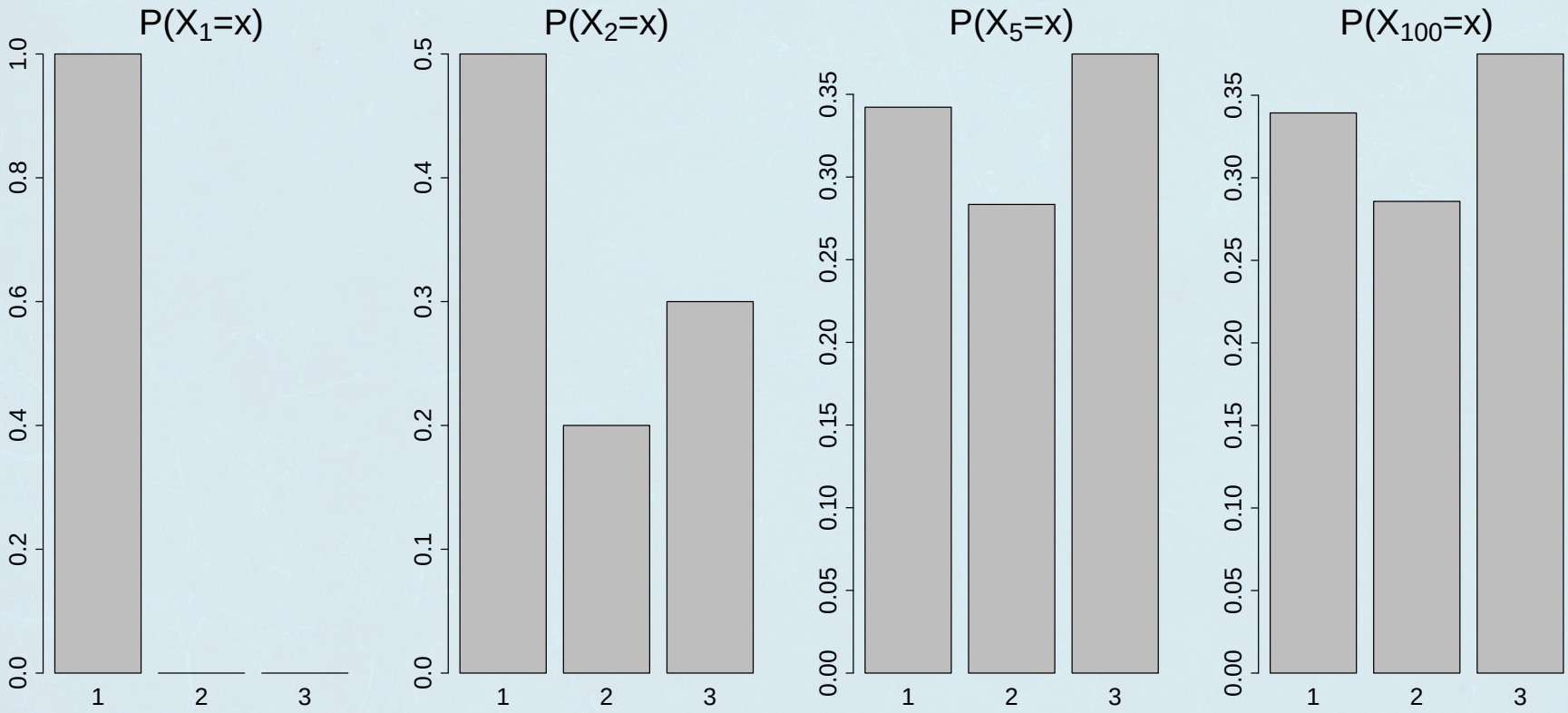
$$\boldsymbol{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}$$

- A Markov chain is *memoryless*
  - Given the present, the future does not depend on the past.

Markov chain, equilibrium distribution

For a Markov chain $\{X_1, X_2, \ldots X_n\}$ with transition matrix $\boldsymbol{P}$:

- $X_t \sim \boldsymbol{\pi}^T \Rightarrow X_{t+1} \sim \boldsymbol{\pi}^T \boldsymbol{P}$.
- $\lim_{n\to\infty} \boldsymbol{\pi_{(1)}}^T \boldsymbol{P}^n \to \boldsymbol{\pi}^*$
  - $\boldsymbol{\pi}^*$ is the **equilibrium distribution** of the Markov chain.



Monte Carlo

- Suppose you are able to simulate $X_1, X_2, \ldots, X_n$, which are all independent and identically distributed (i.i.d.).
- You wish to calculate the expectation $\mu = \mathrm{E}[X]$
- If exact calculation is infeasible, we can use a *Monte Carlo estimate*:

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- The expectation is approximated by the empirical mean of independent samples (law of large numbers)
- The statistical accuracy is inversely proportional to the square root of the sample size.
  - Increasing the sample size by a factor of 100 reduces the estimation error by a factor of 10.

Markov chain Monte Carlo (MCMC)

- Not able to generate i.i.d. samples from the desired distribution.
  - Construct a Markov chain that has the target distribution as its equilibrium distribution.

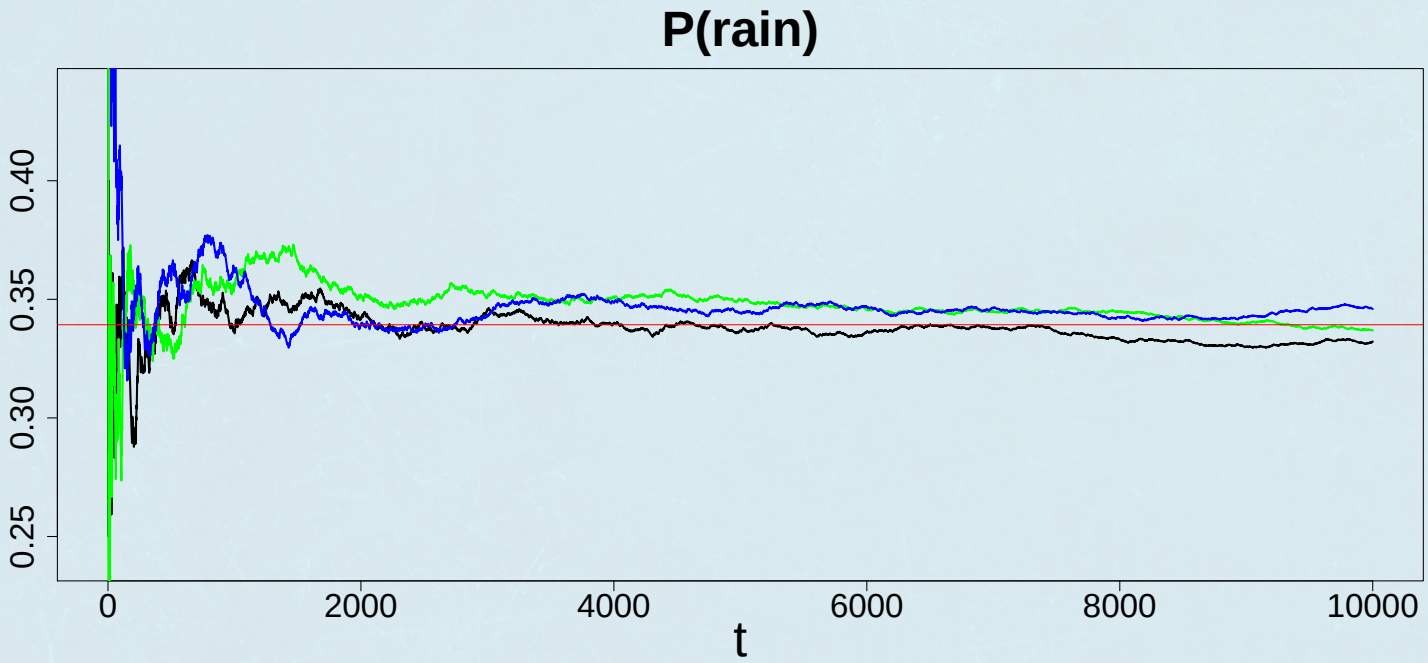
- Simulate (correlated) state values $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n$ from the Markov chain.

- Apply standard Monte Carlo to the Markov chain output:

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} \tilde{X}_i$$

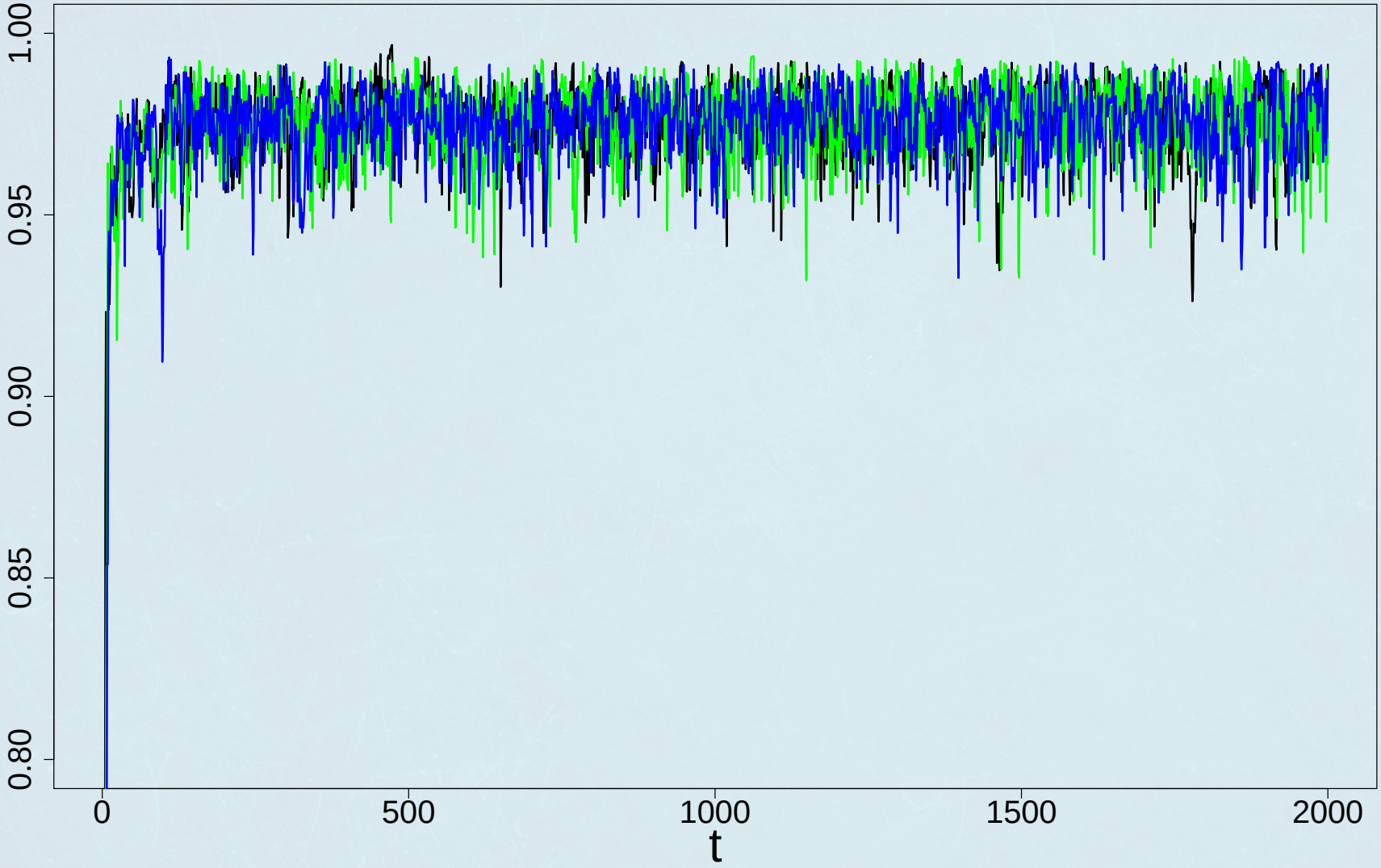
MCMC example, discrete distribution

- We use the Markov chain to model the weather:
  $S = \{s_1 = \text{rainy}, s_2 = \text{sunny}, s_3 = \text{cloudy}\}$.
- $P(\text{rain}) \approx$ The proportion of simulated time periods spent in s_1.
- As we have seen, this simple Markov chain quickly reaches its equilibrium distribution. However, Monte Carlo error is still present.



Markov chain, continuous distribution

- Three different runs of the same Markov chain, started with different initial values.



When has the Markov chain reached equilibrium?

- The chain must be run long enough so that the result is independent of the starting state.
  - May be possible to derive an upper bound of iterations needed.
  - Employ various convergence diagnostics on the simulation output.
- A popular diagnostic is to compare the within-chain variances to the variance of all the chains mixed together.
  - At convergence, these variances should be (close to) identical.
- Diagnostics can identify lack of convergence, but they do not prove convergence.

Perfect simulation

- In 1996, Propp and Wilson showed how MCMC algorithms can be modified so that they deliver *exact* draws from the Markov chains' equilibrium distribution.
- This type of sampling is referred to as **perfect simulation**.
- Perfect sampling determines automatically how long the Markov chain must run to reach the **exact** equilibrium distribution.
  - Ensures unbiased estimates.
  - Must still account for Monte Carlo error.

Coupling from the past (CFTP)

- The algorithm of Propp and Wilson is called Coupling from the past.
- Their simple, yet revolutionary, idea is to run the Markov chain **from the past** instead of into the future.

- Into the future: $X_0\ ,\ X_1\quad ,X_2,\ \ldots\quad ,X_{n-2},X_{n-1},X_n$
- From the past: $X_{-n},X_{-n+1},X_{-n+2},\ldots,X_2\quad ,X_1\quad ,X_0$

- At first glance, this is simply a re-labelling of the Markov chain.
  - How is it helpful?

Update function

- Instead of describing a Markov chain with a transition matrix, we can use an *update function*.
  - $X_{t+1} = \phi(X_t, \varepsilon_{t+1})$
- Letting the distribution of all ε_t be independent standard uniform, our three-state example get the following update function:.

$$\boldsymbol{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \quad \phi(s_i, \varepsilon) = \begin{cases} 1, \text{ for } \varepsilon \in [0, p_{i1}) \\ 2, \text{ for } \varepsilon \in [p_{i1}, p_{i1} + p_{i2}) \\ 3, \text{ for } \varepsilon \in [p_{i1} + p_{i2}, 1]. \end{cases}$$

- Unlike the transition matrix, the update function explicitly describes how the next state depends on the simulated random variable.

Into the future vs. from the past

- Into the future: $X_0\ ,\ X_1\quad ,X_2,\ \ \ldots\quad ,X_{n-2},X_{n-1},X_n$
- From the past: $X_{-n},X_{-n+1},X_{-n+2},\ldots,X_{-2}\quad ,X_{-1}\quad ,X_0$

- Into the future:
  $(X_n|X_0 = x) = \phi(\phi(\ldots\phi(\phi(x,\varepsilon_1),\varepsilon_2),\ldots\varepsilon_{n-1}),\varepsilon_n)$
- From the past:
  $(X_0|X_{-n} = x) = \phi(\phi(\ldots\phi(\phi(x,\varepsilon_{-n+1}),\varepsilon_{-n+2}),\ldots\varepsilon_{-1}),\varepsilon_0)$

- The sequences have identical distribution, given that ε_t is i.i.d., which typically is the case.
- This means that the limits as $n \to \infty$ also have the same distribution:
  - $\lim_{n\to\infty}(X_n|X_0 = x) \sim \lim_{n\to\infty}(X_0|X_{-n} = x)$

Why CFTP works

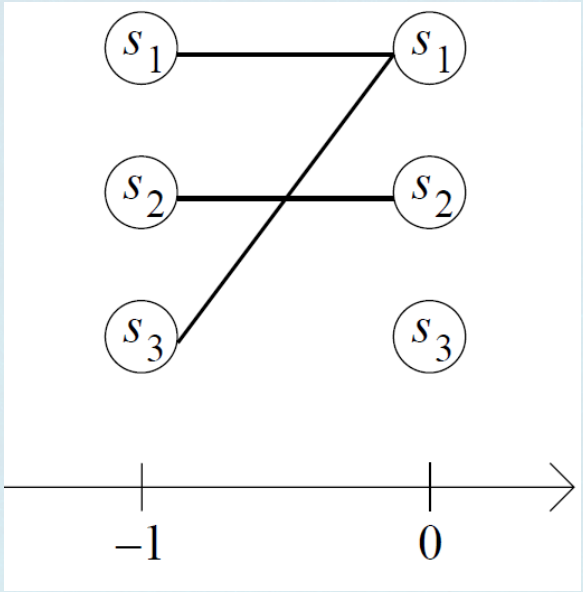
- The limit is impossible to compute when simulating into the future. Why should it be easier to calculate when simulating from the past?
- A simple example shows why. Assume $\phi(X_t, \varepsilon_{t+1}) = \varepsilon_{t+1}$.
  - With the standard i.i.d assumption for $\varepsilon_t,\ t = 1, 2, \ldots$, it is clear this Markov chain forms an i.i.d. sequence.
- Into the future:
  - $X_0, \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$
  - $\lim_{n\to\infty}(X_n | X_0 = x) \sim \epsilon_0$
- From the past:
  - $X_{-n}, \varepsilon_{-n+1}, \varepsilon_{-n+2}, \ldots, \varepsilon_0$
  - $\lim_{n\to\infty}(X_0 | X_{-n} = x) = \epsilon_0$

CFTP algorithm

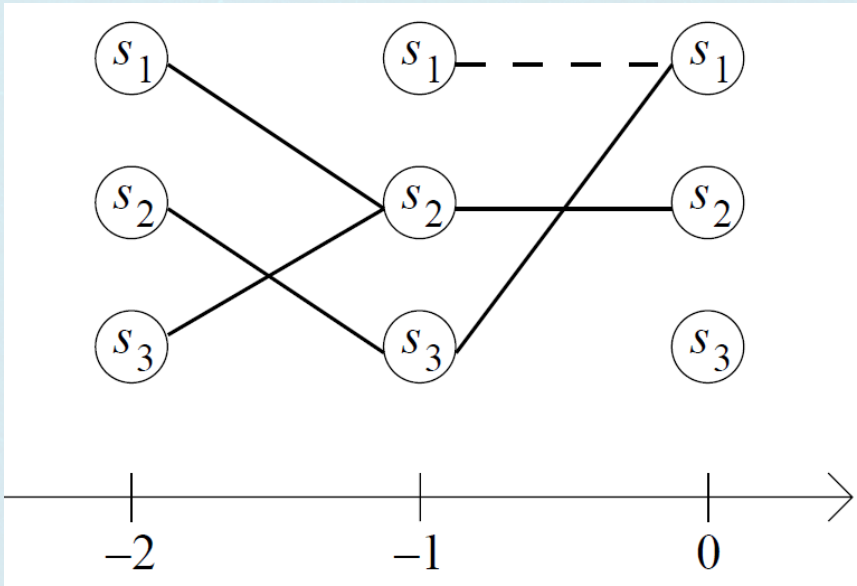
- Given certain conditions, it can be shown that there exists a finite *stopping time* T such that the distribution of $(X_0 | X_{-T} = x)$ is exactly the equilibrium distribution of the Markov chain.
- The algorithm start by defining a sequence of "starting times".
  - Typically $\{-1, -2, -4, -8, -16, \dots\}$.
- The goal of the algorithm is to find a starting time $-T$, where multiple chains with different initial states end up in the same state at $t = 0$.

Coupling from the past (CFTP)

- $\phi(s_1, \varepsilon_0) = s_1$
- $\phi(s_2, \varepsilon_0) = s_2$
- $\phi(s_3, \varepsilon_0) = s_1$



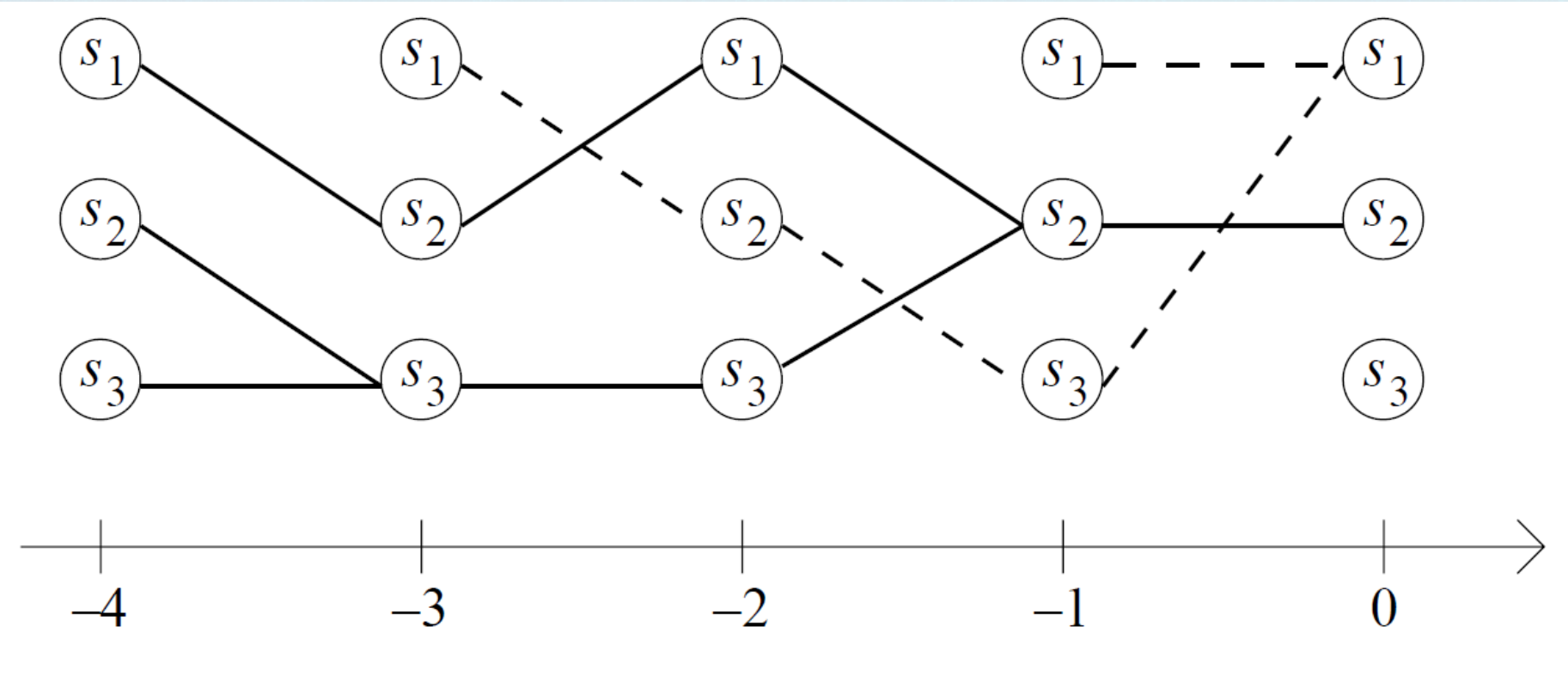
- $\phi(s_1, \varepsilon_{-1}) = s_2$
- $\phi(s_2, \varepsilon_{-1}) = s_3$
- $\phi(s_3, \varepsilon_{-1}) = s_2$



Coupling from the past (CFTP)

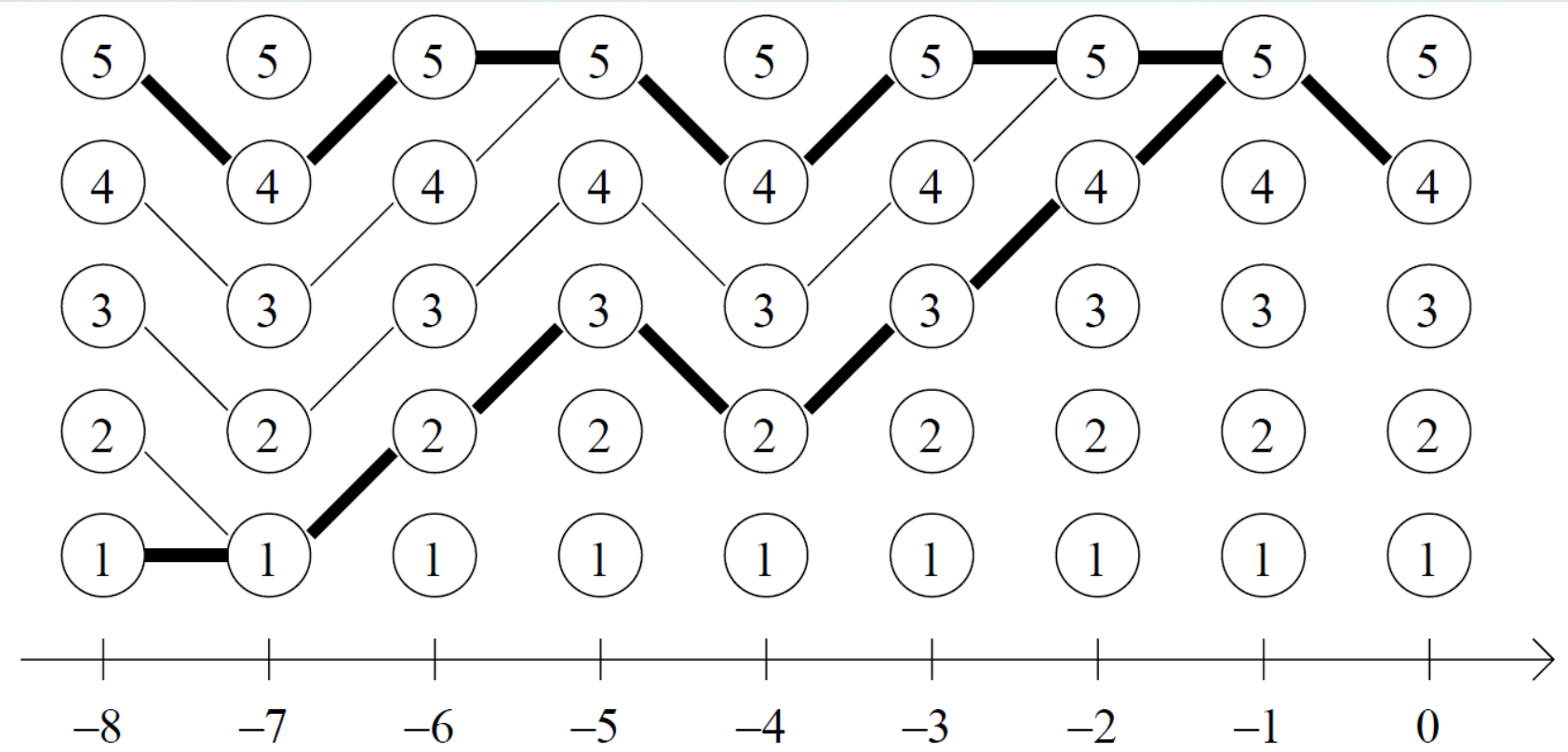
- $\phi(s_1, \varepsilon_{-2}) = s_2$
- $\phi(s_2, \varepsilon_{-2}) = s_1$
- $\phi(s_3, \varepsilon_{-2}) = s_3$

- $\phi(s_1, \varepsilon_{-3}) = s_2$
- $\phi(s_2, \varepsilon_{-3}) = s_3$
- $\phi(s_3, \varepsilon_{-3}) = s_3$



CFTP extension - Sandwiching

- For the CFTP algorithm to be of practical use, it needs to work in cases where the state-space S of the Markov chain is very large.



CFTP extensions

- Read-once CFTP (Wilson)
  - Requires less computer memory, as it does not need to store all the random variables ε_t.
- Fill's algorithm
  - Makes it possible to abandon chains that runs too long, without introducing sampling bias.
- Small-set CFTP (Green & Murdoch)
  - Can be applied to continuous state-spaces.

Final remarks

- Perfect sampling determines automatically how long the Markov chain must run to reach the **exact** equilibrium distribution.
- The most successful applications of perfect sampling involves large, discrete state-spaces.
- Unlike standard MCMC, there is no universal method for perfect sampling that is suitable for all Markov chains.

Bibliography


Craiu, R. V. and Meng, X.-L. (2011).
Perfection within reach: exact MCMC sampling.
Handbook of Markov Chain Monte Carlo, pages 199–226.


Häggström, O. et al. (2002).
Finite Markov chains and algorithmic applications, volume 52.
Cambridge University Press.


Kendall, W. S.
Introduction to CFTP using R.


Kendall, W. S. (2005).
Notes on perfect simulation.
Markov chain Monte Carlo: innovations and applications, 7.