Markov Chains and Computer Science

A not so Short Introduction

Jean-Marc Vincent

Laboratoire LIG, projet Inria-Mescal
UniversitéJoseph Fourier
Jean-Marc.Vincent@imag.fr

Winter 2013

Outline

1. **Markov Chain**
   - History
   - Approaches
2. Formalisation
3. Long run behavior
4. Cache modeling
5. Synthesis

History (Andreï Markov)

This study investigates a text excerpt containing 20,000 Russian letters of the alphabet, excluding Ь and Ъ,[2] from Pushkin's novel *Eugene Onegin* – the entire first chapter and sixteen stanzas of the second.

This sequence provides us with 20,000 connected trials, which are either a vowel or a consonant.

Accordingly, we assume the existence of an unknown constant probability p that the observed letter is a vowel. We determine the approximate value of p by observation, by counting all the vowels and consonants. Apart from p, we shall find – also through observation – the approximate values of two numbers p_1 and p_0, and four numbers $p_{1,1}$, $p_{1,0}$, $p_{0,1}$, and $p_{0,0}$. They represent the following probabilities: p_1 – a vowel follows another vowel; p_0 – a vowel follows a consonant; $p_{1,1}$ – a vowel follows two vowels; $p_{1,0}$ – a vowel follows a consonant that is preceded by a vowel; $p_{0,1}$ – a vowel follows a vowel that is preceded by a consonant; and, finally, $p_{0,0}$ – a vowel follows two consonants.

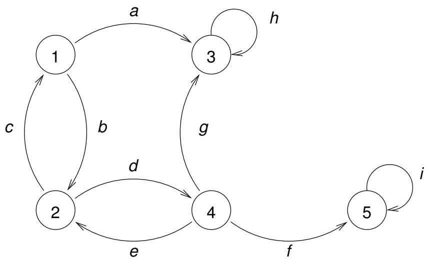
The indices follow the same system that I introduced in my paper "On a Case of Samples Connected in Complex Chain" [Markov 1911b]; with reference to my other paper, "Investigation of a Remarkable Case of Dependent Samples" [Markov 1907a], however, $p_0 = p_2$. We denote the opposite probabilities for consonants with q and indices that follow the same pattern.

If we seek the value of p, we first find 200 approximate values from which we can determine the arithmetic mean. To be precise, we divide the entire sequence of 20,000 letters into 200 separate sequences of 100 letters, and count how many vowels there are in each 100: we obtain 200 numbers, which, when divided by 100, yield 200 approximate values of p.

An example of statistical investigation in the text of "Eugene Onegin" illustrating coupling of "tests" in chains.
(1913) In Proceedings of Academic Scientific St. Petersburg, VI, pages 153-162.

1856-1922

Graphs and Paths



Random Walks

Path in a graph:
X_n n-th visited node
path : $i_0, i_1, \cdots, i_n$
normalized weight : arc $(i,j) \longrightarrow p_{i,j}$

concatenation : $. \longrightarrow \times$
$\mathcal{P}(i_0, i_1, \cdots, i_n) = p_{i_0,i_1} p_{i_1,i_2} \cdots p_{i_{n-1},i_n}$

disjoint union : $\cup \longrightarrow +$
$\mathcal{P}(i_0 \rightsquigarrow i_n) = \sum_{i_1,\cdots,i_{n-1}} p_{i_0,i_1} p_{i_1,i_2} \cdots p_{i_{n-1},i_n}$

automaton : state/transitions randomized (language)

Dynamical Systems

Figure 3. A fern drawn by a Markov chain

Diaconis-Freedman 99

Evolution Operator

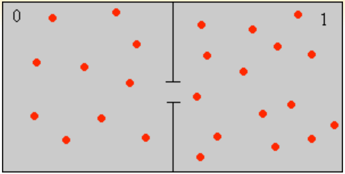
Initial value : X_0
Recurrence equation : $X_{n+1} = \Phi(X_n, \xi_{n+1})$

Innovation at step $n+1$: ξ_{n+1}
Finite set of innovations : $\{\phi_1, \phi_2, \cdots, \phi_K\}$

Random function (chosen with a given probability)

Randomized Iterated Systems

Measure Approach



Ehrenfest's Urn (1907)

Paul Ehrenfest (1880-1933)

Distribution of K particles

Initial State $X_0 = 0$
State = nb of particles in 0
Dynamic : uniform choice of a particle and jump to the other side

$$\begin{aligned}\pi_n(i) &= \mathbb{P}(X_n = i | X_0 = 0) \\ &= \pi_{n-1}(i-1).\frac{K-i+1}{K} \\ &\quad +\pi_{n-1}(i+1).\frac{i+1}{K}\end{aligned}$$

$$\pi_n = \pi_{n-1}.P$$

Iterated product of matrices

Algorithmic Interpretation

```
int minimum (T,K)
min= +∞
cpt=0;
for (k=0; k < K; k++) do
  if (T[i]< min) then
    min = T[k];
    process(min);
    cpt++;
  end if
end for
return(cpt)
```

Worst case K;
Best case 1;
on average ?

Number of processing min

State : $X_n =$ rank of the n^{th} processing

$$\mathbb{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{k-1}, \cdots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j | X_n = i)$$

$$\mathbb{P}(X_{n+1} = j | X_n = i) = \begin{cases} \frac{1}{K-i+1} & \text{si } j < i; \\ 0 & \text{sinon.} \end{cases}$$

All the information of for the step $n+1$ is contained in the state at step n

$$\tau = \min\{n; \quad X_n = 1\}$$

Correlation of length 1

Outline

Formal definition

Let $\{X_n\}_{n\in\mathbb{N}}$ a random sequence of variables in a discrete state-space $\mathcal{X}$

$\{X_n\}_{n\in\mathbb{N}}$ is a **Markov chain** with initial law $\pi(0)$ iff

- $X_0 \sim \pi(0)$ and
- for all $n \in \mathbb{N}$ and for all $(j, i, i_{n-1}, \cdots, i_0) \in \mathcal{X}^{n+2}$

$$\mathbb{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \cdots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j | X_n = i).$$

$\{X_n\}_{n\in\mathbb{N}}$ is a **homogeneous** Markov chain iff

- for all $n \in \mathbb{N}$ and for all $(j, i) \in \mathcal{X}^2$

$$\mathbb{P}(X_{n+1} = j | X_n = i) = \mathbb{P}(X_1 = j | X_0 = i) \stackrel{def}{=} p_{i,j}.$$

(invariance during time of probability transition)

Formal definition

Let $\{X_n\}_{n\in\mathbb{N}}$ a random sequence of variables in a discrete state-space $\mathcal{X}$

$\{X_n\}_{n\in\mathbb{N}}$ is a **Markov chain** with initial law $\pi(0)$ iff

- $X_0 \sim \pi(0)$ and
- for all $n \in \mathbb{N}$ and for all $(j, i, i_{n-1}, \cdots, i_0) \in \mathcal{X}^{n+2}$

$$\mathbb{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \cdots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j | X_n = i).$$

$\{X_n\}_{n\in\mathbb{N}}$ is a **homogeneous** Markov chain iff

- for all $n \in \mathbb{N}$ and for all $(j, i) \in \mathcal{X}^2$

$$\mathbb{P}(X_{n+1} = j | X_n = i) = \mathbb{P}(X_1 = j | X_0 = i) \stackrel{def}{=} p_{i,j}.$$

(invariance during time of probability transition)

Algebraic representation

$P = ((p_{i,j}))$ is the **transition matrix** of the chain

- P is a **stochastic matrix**

$$p_{i,j} \geqslant 0; \quad \sum_j p_{i,j} = 1.$$

Linear recurrence equation $\pi_i(n) = \mathbb{P}(X_n = i)$

$$\pi_n = \pi_{n-1} P.$$

- Equation of **Chapman-Kolmogorov** (homogeneous): $P^n = ((p_{i,j}^{(n)}))$

$$p_{i,j}^{(n)} = \mathbb{P}(X_n = j | X_0 = i); \quad P^{n+m} = P^n . P^m;$$

$$\begin{aligned} \mathbb{P}(X_{n+m} = j | X_0 = i) &= \sum_k \mathbb{P}(X_{n+m} = j | X_m = k) \mathbb{P}(X_m = k | X_0 = i); \\ &= \sum_k \mathbb{P}(X_n = j | X_0 = k) \mathbb{P}(X_m = k | X_0 = i). \end{aligned}$$

Interpretation: decomposition of the set of paths with length $n + m$ from i to j.

Problems

Finite horizon

- Estimation of $\pi(n)$
- Estimation of stopping times

$$\tau_A = \inf\{n \geqslant 0;\ X_n \in A\}$$

- $\ldots$

Infinite horizon

- Convergence properties
- Estimation of the asymptotics
- Estimation speed of convergence
- $\ldots$

Problems

Finite horizon

- Estimation of $\pi(n)$
- Estimation of stopping times

$$\tau_A = \inf\{n \geqslant 0;\ X_n \in A\}$$

- $\cdots$

Infinite horizon

- Convergence properties
- Estimation of the asymptotics
- Estimation speed of convergence
- $\cdots$

Applications in computer science

Applications in most of scientific domains ...
In computer science :

Markov chain : an algorithmic tool

- Numerical methods (Monte-Carlo methods)
- Randomized algorithms (ex: TCP, searching, pageRank...)
- Learning machines (hidden Markov chains)
- . . .

Markov chains : a modeling tool

- Performance evaluation (quantification and dimensionning)
- Stochastic control
- Program verification
- . . .

Applications in computer science

Applications in most of scientific domains ...
In computer science :

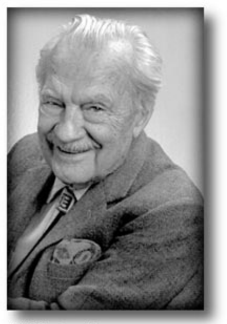
Markov chain : an algorithmic tool

- Numerical methods (Monte-Carlo methods)
- Randomized algorithms (ex: TCP, searching, pageRank...)
- Learning machines (hidden Markov chains)
-. . .

Markov chains : a modeling tool

- Performance evaluation (quantification and dimensionning)
- Stochastic control
- Program verification
-. . .

Nicholas Metropolis (1915-1999)



Nick Metropolis

Metropolis contributed several original ideas to mathematics and physics. Perhaps the most widely known is the Monte Carlo method. Also, in 1953 Metropolis co-authored the first paper on a technique that was central to the method known now as simulated annealing. He also developed an algorithm (the Metropolis algorithm or Metropolis-Hastings algorithm) for generating samples from the Boltzmann distribution, later generalized by W.K. Hastings.

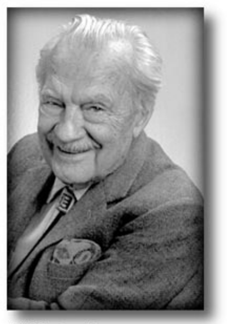
Simulated annealing

Convergence to a global minimum by a stochastic gradient scheme.

$$X_{n+1} = X_n - \vec{grad}\Phi(X_n)\Delta_n(Random).$$

$$\Delta_n(random) \xrightarrow{n\to\infty} 0.$$

Nicholas Metropolis (1915-1999)



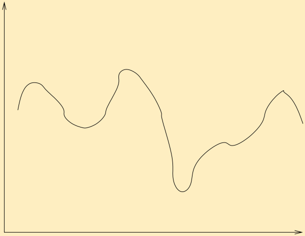
Nick Metropolis

Metropolis contributed several original ideas to mathematics and physics. Perhaps the most widely known is the Monte Carlo method. Also, in 1953 Metropolis co-authored the first paper on a technique that was central to the method known now as simulated annealing. He also developed an algorithm (the Metropolis algorithm or Metropolis-Hastings algorithm) for generating samples from the Boltzmann distribution, later generalized by W.K. Hastings.

Simulated annealing

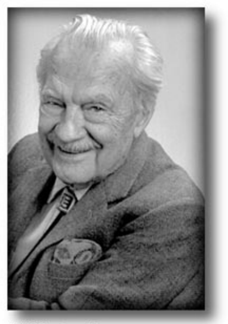
Convergence to a global minimum by a stochastic gradient scheme.

$$X_{n+1} = X_n - \vec{grad}\Phi(X_n)\Delta_n(Random).$$



$$\Delta_n(random) \xrightarrow{n\to\infty} 0.$$

Nicholas Metropolis (1915-1999)



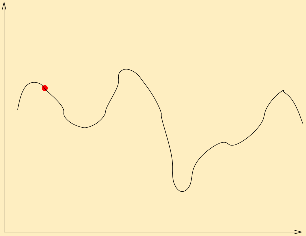
Nick Metropolis

Metropolis contributed several original ideas to mathematics and physics. Perhaps the most widely known is the Monte Carlo method. Also, in 1953 Metropolis co-authored the first paper on a technique that was central to the method known now as simulated annealing. He also developed an algorithm (the Metropolis algorithm or Metropolis-Hastings algorithm) for generating samples from the Boltzmann distribution, later generalized by W.K. Hastings.

Simulated annealing

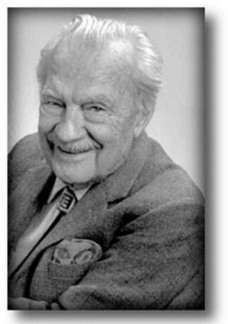
Convergence to a global minimum by a stochastic gradient scheme.

$$X_{n+1} = X_n - \vec{grad}\Phi(X_n)\Delta_n(Random).$$



$$\Delta_n(random) \xrightarrow{n\to\infty} 0.$$

Nicholas Metropolis (1915-1999)



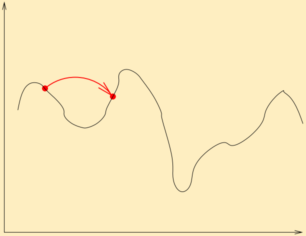
Nick Metropolis

Metropolis contributed several original ideas to mathematics and physics. Perhaps the most widely known is the Monte Carlo method. Also, in 1953 Metropolis co-authored the first paper on a technique that was central to the method known now as simulated annealing. He also developed an algorithm (the Metropolis algorithm or Metropolis-Hastings algorithm) for generating samples from the Boltzmann distribution, later generalized by W.K. Hastings.

Simulated annealing

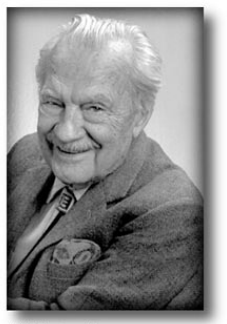
Convergence to a global minimum by a stochastic gradient scheme.

$$X_{n+1} = X_n - \vec{grad}\Phi(X_n)\Delta_n(Random).$$



$$\Delta_n(random) \overset{n\to\infty}{\longrightarrow} 0.$$

Nicholas Metropolis (1915-1999)



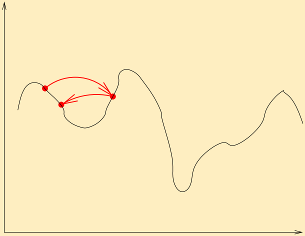
Nick Metropolis

Metropolis contributed several original ideas to mathematics and physics. Perhaps the most widely known is the Monte Carlo method. Also, in 1953 Metropolis co-authored the first paper on a technique that was central to the method known now as simulated annealing. He also developed an algorithm (the Metropolis algorithm or Metropolis-Hastings algorithm) for generating samples from the Boltzmann distribution, later generalized by W.K. Hastings.

Simulated annealing

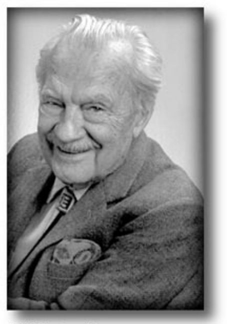
Convergence to a global minimum by a stochastic gradient scheme.

$$X_{n+1} = X_n - \vec{grad}\Phi(X_n)\Delta_n(Random).$$



$$\Delta_n(random) \xrightarrow{n\to\infty} 0.$$

Nicholas Metropolis (1915-1999)



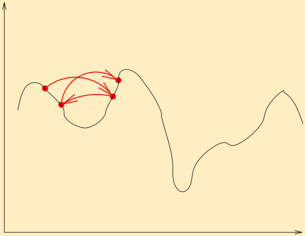
Nick Metropolis

Metropolis contributed several original ideas to mathematics and physics. Perhaps the most widely known is the Monte Carlo method. Also, in 1953 Metropolis co-authored the first paper on a technique that was central to the method known now as simulated annealing. He also developed an algorithm (the Metropolis algorithm or Metropolis-Hastings algorithm) for generating samples from the Boltzmann distribution, later generalized by W.K. Hastings.

Simulated annealing

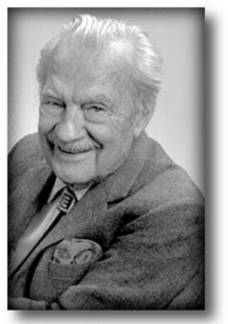
Convergence to a global minimum by a stochastic gradient scheme.

$$X_{n+1} = X_n - \vec{grad}\Phi(X_n)\Delta_n(Random).$$



$$\Delta_n(random) \xrightarrow{n\to\infty} 0.$$

Nicholas Metropolis (1915-1999)



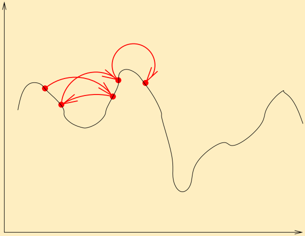
Nick Metropolis

Metropolis contributed several original ideas to mathematics and physics. Perhaps the most widely known is the Monte Carlo method. Also, in 1953 Metropolis co-authored the first paper on a technique that was central to the method known now as simulated annealing. He also developed an algorithm (the Metropolis algorithm or Metropolis-Hastings algorithm) for generating samples from the Boltzmann distribution, later generalized by W.K. Hastings.

Simulated annealing

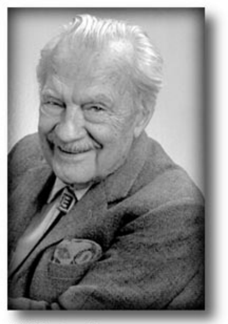
Convergence to a global minimum by a stochastic gradient scheme.

$$X_{n+1} = X_n - \vec{grad}\Phi(X_n)\Delta_n(Random).$$



$$\Delta_n(random) \xrightarrow{n\to\infty} 0.$$

Nicholas Metropolis (1915-1999)



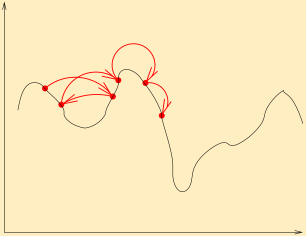
Nick Metropolis

Metropolis contributed several original ideas to mathematics and physics. Perhaps the most widely known is the Monte Carlo method. Also, in 1953 Metropolis co-authored the first paper on a technique that was central to the method known now as simulated annealing. He also developed an algorithm (the Metropolis algorithm or Metropolis-Hastings algorithm) for generating samples from the Boltzmann distribution, later generalized by W.K. Hastings.

Simulated annealing

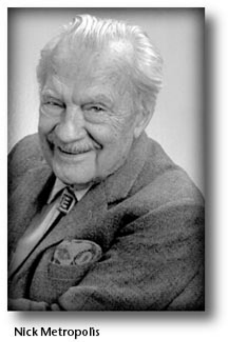
Convergence to a global minimum by a stochastic gradient scheme.

$$X_{n+1} = X_n - \vec{grad}\Phi(X_n)\Delta_n(Random).$$



$$\Delta_n(random) \xrightarrow{n\to\infty} 0.$$

Nicholas Metropolis (1915-1999)

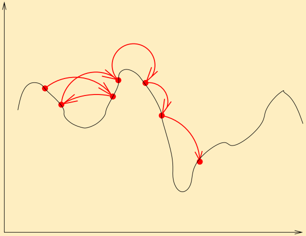


Metropolis contributed several original ideas to mathematics and physics. Perhaps the most widely known is the Monte Carlo method. Also, in 1953 Metropolis co-authored the first paper on a technique that was central to the method known now as simulated annealing. He also developed an algorithm (the Metropolis algorithm or Metropolis-Hastings algorithm) for generating samples from the Boltzmann distribution, later generalized by W.K. Hastings.

Simulated annealing

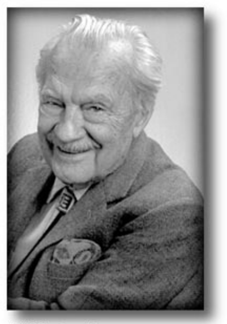
Convergence to a global minimum by a stochastic gradient scheme.

$$X_{n+1} = X_n - \vec{grad}\,\Phi(X_n)\Delta_n(Random).$$



$$\Delta_n(random) \xrightarrow{n\to\infty} 0.$$

Nicholas Metropolis (1915-1999)



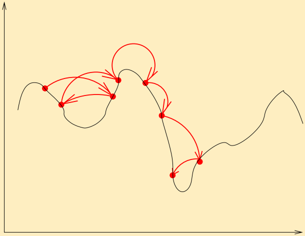
Nick Metropolis

Metropolis contributed several original ideas to mathematics and physics. Perhaps the most widely known is the Monte Carlo method. Also, in 1953 Metropolis co-authored the first paper on a technique that was central to the method known now as simulated annealing. He also developed an algorithm (the Metropolis algorithm or Metropolis-Hastings algorithm) for generating samples from the Boltzmann distribution, later generalized by W.K. Hastings.

Simulated annealing

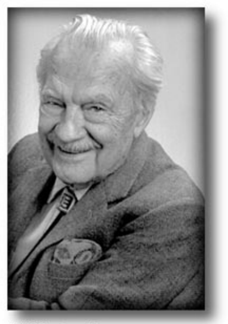
Convergence to a global minimum by a stochastic gradient scheme.

$$X_{n+1} = X_n - \vec{grad}\Phi(X_n)\Delta_n(Random).$$



$$\Delta_n(random) \overset{n\to\infty}{\longrightarrow} 0.$$

Nicholas Metropolis (1915-1999)



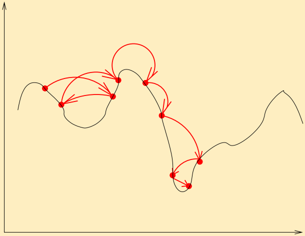
Nick Metropolis

Metropolis contributed several original ideas to mathematics and physics. Perhaps the most widely known is the Monte Carlo method. Also, in 1953 Metropolis co-authored the first paper on a technique that was central to the method known now as simulated annealing. He also developed an algorithm (the Metropolis algorithm or Metropolis-Hastings algorithm) for generating samples from the Boltzmann distribution, later generalized by W.K. Hastings.

Simulated annealing

Convergence to a global minimum by a stochastic gradient scheme.

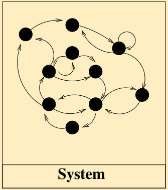
$$X_{n+1} = X_n - \vec{grad}\Phi(X_n)\Delta_n(Random).$$



$$\Delta_n(random) \overset{n\to\infty}{\longrightarrow} 0.$$

Modeling and Analysis of Computer Systems

Complex system



Basic model assumptions

System :
- automaton (discrete state space)
- **discrete** or continuous time

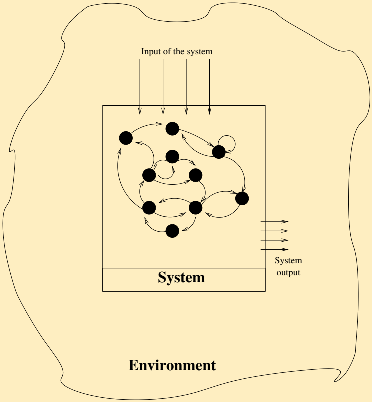
Environment : non deterministic
- time homogeneous
- stochastically regular

Problem

Understand "typical" states
- steady-state estimation
- ergodic simulation
- state space exploring techniques

Modeling and Analysis of Computer Systems

Complex system



Basic model assumptions

System :
- automaton (discrete state space)
- **discrete** or continuous time

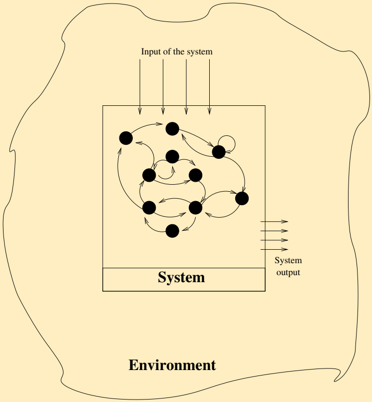
Environment : non deterministic
- time homogeneous
- stochastically regular

Problem

Understand "typical" states
- steady-state estimation
- ergodic simulation
- state space exploring techniques

Modeling and Analysis of Computer Systems

Complex system



Basic model assumptions

System :
- automaton (discrete state space)
- **discrete** or continuous time

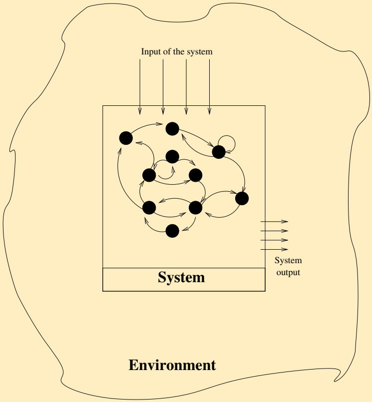
Environment : non deterministic
- time homogeneous
- stochastically regular

Problem

Understand "typical" states
- steady-state estimation
- ergodic simulation
- state space exploring techniques

Modeling and Analysis of Computer Systems

Complex system



Basic model assumptions

System :
- automaton (discrete state space)
- **discrete** or continuous time

Environment : non deterministic
- time homogeneous
- stochastically regular

Problem

Understand "typical" states
- steady-state estimation
- ergodic simulation
- state space exploring techniques

Outline