Chaînes de Markov une introduction

De l'analyse de texte à la physique des particules

Jean-Marc.Vincent@imag.fr
Laboratoire d'Informatique de Grenoble

Equipe Projet Mescal

Plan

1. **Chaînes de Markov**
   - Historique
   - Différentes approches
   - Formalisation
   - Applications en informatique
   - Bibliographie
2. **Comportement asymptotique**
   - Classification des états
   - Théorèmes de convergence
   - Echantillonage

Plan

Historique (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.

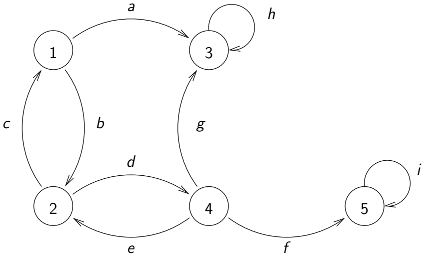
The indices follow the same system that I introduced in my paper "On a Case of Samples Connected in Complex Chain" [Markov 1911b]; with reference to my other paper, "Investigation of a Remarkable Case of Dependent Samples" [Markov 1907a], however, $p_0 = p_2$. We denote the opposite probabilities for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

Approche graphes et chemins



Marche aléatoire

Chemin dans un graphe :
X_n n-ième nœud visité
chemin : $i_0, i_1, \cdots, i_n$
poids normalisé : arc $(i,j) \longrightarrow p_{i,j}$

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

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

automate : états/transitions randomisées (langage)

Approche systèmes dynamiques

Figure 3. A fern drawn by a Markov chain

Diaconis-Freedman 99

Opérateur d'évolution

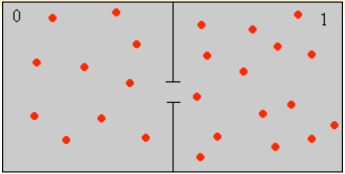
Valeur initiale : X_0
Récurrence : $X_{n+1} = \Phi(X_n, \xi_{n+1})$

Innovation à l'étape $n+1$: ξ_{n+1}
Ensemble fini d'innovations :
$\{\phi_1, \phi_2, \cdots, \phi_K\}$

Choix aléatoire de la fonction appliquée (suivant une probabilité)

Système itéré de fonctions randomisées

Approche mesure



Urne de Ehrenfest (1907)

Paul Ehrenfest (1880-1933)

Distribution des K particules

Etat initial $X_0 = 0$
Etat = nb de particules en 0
Dynamique : choix d'une particule et changement de coté

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

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

Produit itéré de matrices

Approche information

```
int minimum (T,K)
min= +∞
cpt=0;
pour (k=0; k < K; k++)
faire
   si (T[i]< min) alors
      min = T[k];
      Traiter(min);
      cpt++;
   fin si
fin pour
retourne(cpt)
```

Coût au pire K ;
au mieux 1 ;
et en moyenne ?

Nombre d'affectations

Etat : $X_n =$ rang de la $n^{\text{ième}}$ affectation

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

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

Toute l'information nécessaire pour l'étape $n+1$ est contenue dans la valeur de l'état à l'étape k

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

Corrélation au rang 1

Définition formelle

Soit $\{X_n\}_{n\in\mathbb{N}}$ une suite de variables aléatoires à valeurs dans un espace discret $\mathcal{X}$

$\{X_n\}_{n\in\mathbb{N}}$ est une **chaîne de Markov** de loi initiale $\pi(0)$ ssi

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

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

$\{X_n\}_{n\in\mathbb{N}}$ est une chaîne de Markov **homogène** ssi

- pour tout $n \in \mathbb{N}$ et pour tout $(j, i) \in \mathcal{X}^2$

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

(invariance des probabilités de transitions dans le temps)

Définition formelle

Soit $\{X_n\}_{n\in\mathbb{N}}$ une suite de variables aléatoires à valeurs dans un espace discret $\mathcal{X}$

$\{X_n\}_{n\in\mathbb{N}}$ est une **chaîne de Markov** de loi initiale $\pi(0)$ ssi

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

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

$\{X_n\}_{n\in\mathbb{N}}$ est une chaîne de Markov **homogène** ssi

- pour tout $n \in \mathbb{N}$ et pour tout $(j, i) \in \mathcal{X}^2$

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

(invariance des probabilités de transitions dans le temps)

Aspect algébrique

$P = ((p_{i,j}))$ est la **matrice de transition** de la chaîne

- P est une **matrice stochastique**

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

  Equation de récurrence linéaire $\pi_i(n) = \mathbb{P}(X_n = i)$

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

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

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

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

  Interprétation en décomposition de l'ensemble des chemins de longueur $n + m$ de i à j.

Problématiques

Horizon fini

- Estimation de $\pi(n)$
- Estimation de temps d'arrêt

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

- $\cdots$

Horizon infini

- Propriétés de convergence
- Estimation des asymptotique
- Estimation de la vitesse de convergence
- $\cdots$

Problématiques

Horizon fini

- Estimation de $\pi(n)$
- Estimation de temps d'arrêt

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

- $\cdots$

Horizon infini

- Propriétés de convergence
- Estimation des asymptotique
- Estimation de la vitesse de convergence
- $\cdots$

Applications en informatique

Applications dans la plupart des domaines scientifiques ...
En informatique :

Chaîne de Markov : outil algorithmique

- Méthodes numériques (méthodes de Monte-Carlo)
- Algorithmes randomisés (ex : TCP, recherche, pageRank...)
- Apprentissage (chaînes de Markov cachées)
- . . .

Chaîne de Markov : outil de modélisation

- Evaluation de performances (quantification et dimensionnement)
- Contrôle stochastique
- Vérification de programmes
- . . .

Applications en informatique

Applications dans la plupart des domaines scientifiques ...
En informatique :

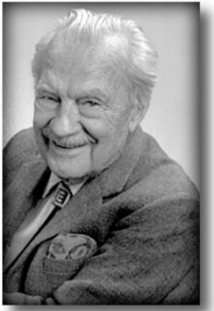
Chaîne de Markov : outil algorithmique

- Méthodes numériques (méthodes de Monte-Carlo)
- Algorithmes randomisés (ex : TCP, recherche, pageRank...)
- Apprentissage (chaînes de Markov cachées)
- . . .

Chaîne de Markov : outil de modélisation

- Evaluation de performances (quantification et dimensionnement)
- Contrôle stochastique
- Vérification de programmes
- . . .

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.

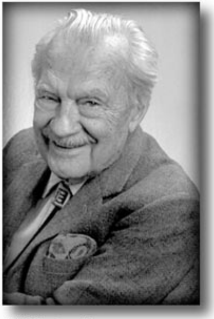
Recuit simulé

Convergence vers un minimum global par une descente de gradient stochastique.

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

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

Nicholas Metropolis (1915-1999)



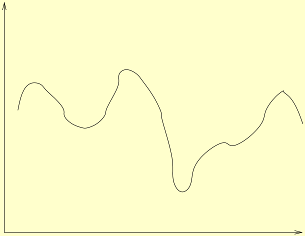
Nick Metropolis

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.

Recuit simulé

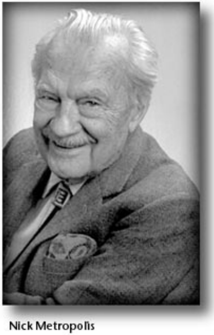
Convergence vers un minimum global par une descente de gradient stochastique.

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



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

Nicholas Metropolis (1915-1999)

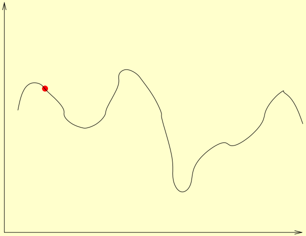


Metropolis contributed several original ideas to mathematics and physics. Perhaps the most widely known is the Monte Carlo method. Also, in 1953 Metropolis co-authored the first paper on a technique that was central to the method known now as simulated annealing. He also developed an algorithm (the Metropolis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

Recuit simulé

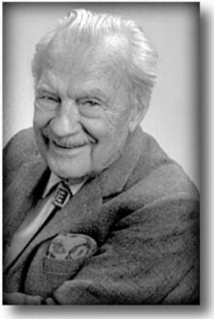
Convergence vers un minimum global par une descente de gradient stochastique.

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



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

Nicholas Metropolis (1915-1999)



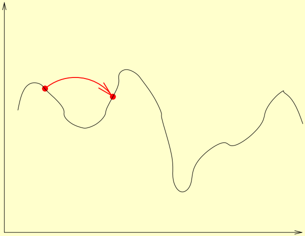
Nick Metropolis

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.

Recuit simulé

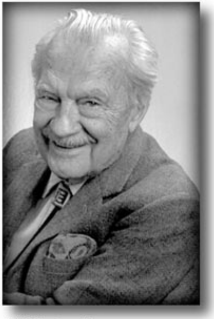
Convergence vers un minimum global par une descente de gradient stochastique.

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



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

Nicholas Metropolis (1915-1999)



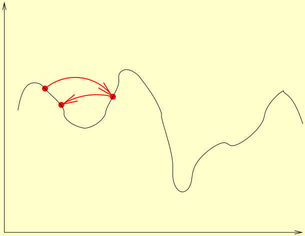
Nick Metropolis

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.

Recuit simulé

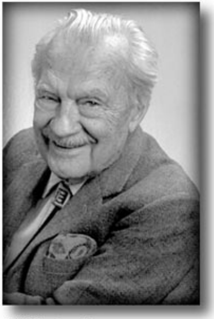
Convergence vers un minimum global par une descente de gradient stochastique.

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



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

Nicholas Metropolis (1915-1999)



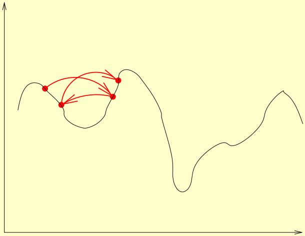
Nick Metropolis

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.

Recuit simulé

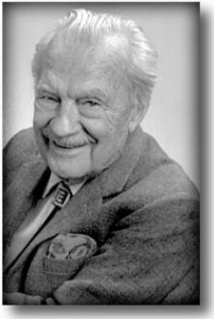
Convergence vers un minimum global par une descente de gradient stochastique.

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



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

Nicholas Metropolis (1915-1999)



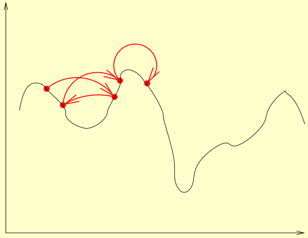
Nick Metropolis

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.

Recuit simulé

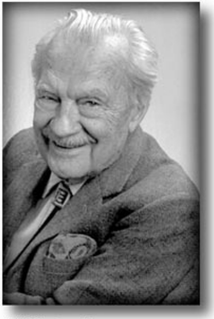
Convergence vers un minimum global par une descente de gradient stochastique.

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



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

Nicholas Metropolis (1915-1999)



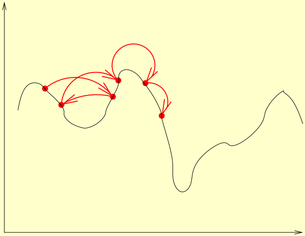
Nick Metropolis

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.

Recuit simulé

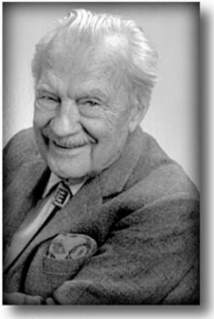
Convergence vers un minimum global par une descente de gradient stochastique.

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



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

Nicholas Metropolis (1915-1999)



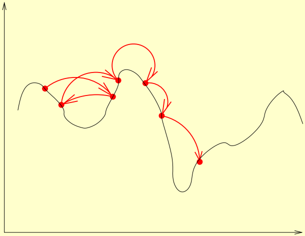
Nick Metropolis

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.

Recuit simulé

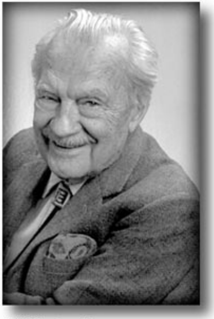
Convergence vers un minimum global par une descente de gradient stochastique.

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



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

Nicholas Metropolis (1915-1999)



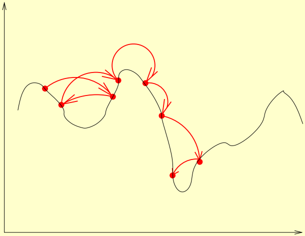
Nick Metropolis

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.

Recuit simulé

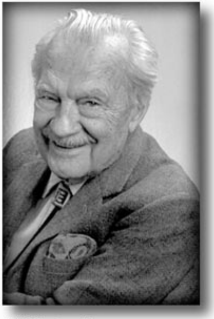
Convergence vers un minimum global par une descente de gradient stochastique.

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



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

Nicholas Metropolis (1915-1999)



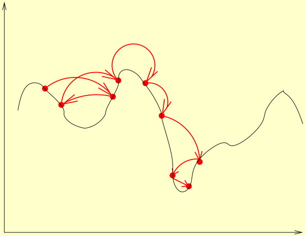
Nick Metropolis

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.

Recuit simulé

Convergence vers un minimum global par une descente de gradient stochastique.

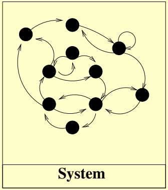
$$X_{n+1} = X_n - \vec{grad}\Phi(X_n)\Delta_n(Random).$$



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

Modeling and Analysis of Computer Systems

Complex system



Basic model assumptions

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

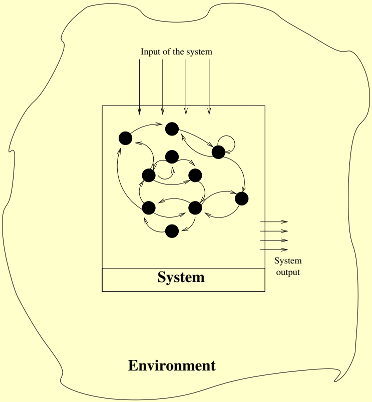
Environment : non deterministic
- time homogeneous
- stochastically regular

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

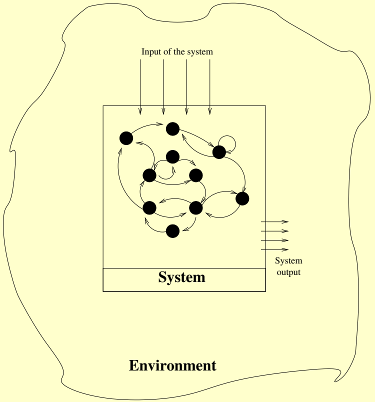
Environment : non deterministic
- time homogeneous
- stochastically regular

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

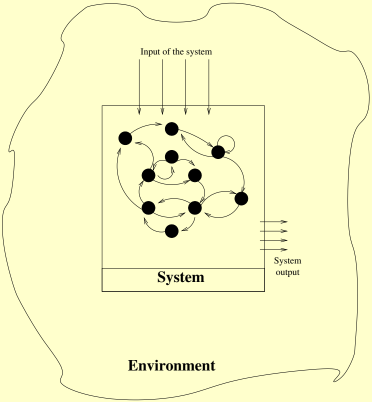
Environment : non deterministic
- time homogeneous
- stochastically regular

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

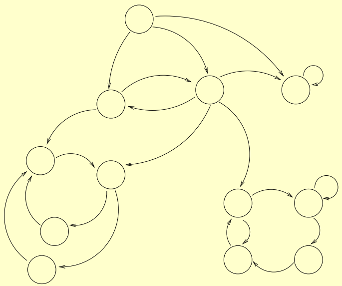
Problem

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

Plan

States classification

Graph analysis



Irreducible class

Strongly connected components
i and j are in the same component if there exist a path from i to j and a path from j to i with a positive probability
Leaves of the tree of strongly connected components are **irreducible** classes
States in irreducible classes are called **recurrent**
Other states are called **transient**

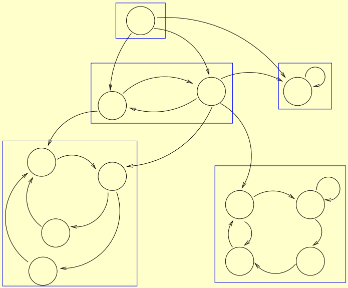
Periodicity

An irreducible class is **aperiodic** if the gcd of length of all cycles is 1

A Markov chain is **irreducible** if there is only one class.
Each state is reachable from any other state with a positive probability path.

States classification

Graph analysis



Irreducible class

Strongly connected components
i and j are in the same component if there exist a path from i to j and a path from j to i with a positive probability
Leaves of the tree of strongly connected components are **irreducible** classes
States in irreducible classes are called **recurrent**
Other states are called **transient**

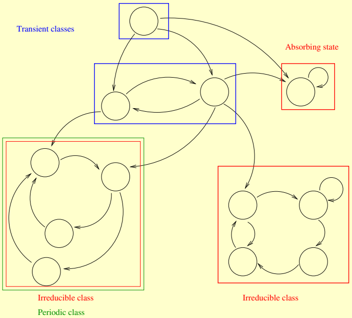
Periodicity

An irreducible class is **aperiodic** if the gcd of length of all cycles is 1

A Markov chain is **irreducible** if there is only one class.
Each state is reachable from any other state with a positive probability path.

States classification

Graph analysis



Irreducible class

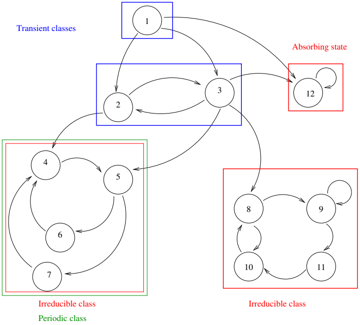
Strongly connected components
i and j are in the same component if there exist a path from i to j and a path from j to i with a positive probability
Leaves of the tree of strongly connected components are **irreducible** classes
States in irreducible classes are called **recurrent**
Other states are called **transient**

Periodicity

An irreducible class is **aperiodic** if the gcd of length of all cycles is 1

A Markov chain is **irreducible** if there is only one class.
Each state is reachable from any other state with a positive probability path.

States classification : matrix form

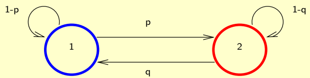


	1	2	3	4	5	6	7	8	9	10	11	12
1		*	*									*
2			*	*								
3					*			*				*
4					*							
5						*	*					
6				*								
7				*								
8									*	*		
9									*		*	
10								*				
11										*		
12												1

Automaton Flip-flop

ON-OFF system

Two states model :
- communication line
- processor activity
- ...



Parameters :
- proportion of transitions : p, q
- mean sojourn time in state 1 : $\frac{1}{p}$
- mean sojourn time in state 2 : $\frac{1}{q}$

Trajectory

X_n state of the automaton at time n.

Transient distribution

$$\pi_n(1) = \mathbb{P}(X_n = 1);$$

$$\pi_n(2) = \mathbb{P}(X_n = 2)$$

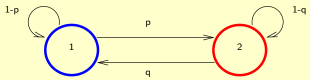
Problem

Estimation of π_n : state prevision, resource utilization

Automaton Flip-flop

ON-OFF system

Two states model :
- communication line
- processor activity
- ...



Parameters :
- proportion of transitions : p, q
- mean sojourn time in state 1 : $\frac{1}{p}$
- mean sojourn time in state 2 : $\frac{1}{q}$

Trajectory

X_n state of the automaton at time n.

Transient distribution

$$\pi_n(1) = \mathbb{P}(X_n = 1);$$

$$\pi_n(2) = \mathbb{P}(X_n = 2)$$

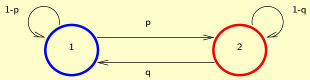
Problem

Estimation of π_n : state prevision, resource utilization

Automaton Flip-flop

ON-OFF system

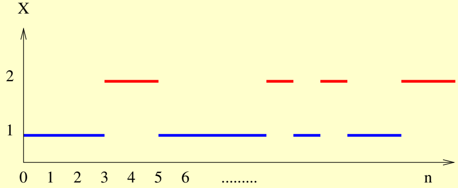
Two states model :
- communication line
- processor activity
- ...



Parameters :
- proportion of transitions : p, q
- mean sojourn time in state 1 : $\frac{1}{p}$
- mean sojourn time in state 2 : $\frac{1}{q}$

Trajectory

X_n state of the automaton at time n.



Transient distribution

$$\pi_n(1) = \mathbb{P}(X_n = 1);$$

$$\pi_n(2) = \mathbb{P}(X_n = 2)$$

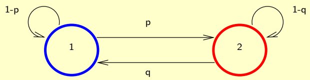
Problem

Estimation of π_n : state prevision, resource utilization

Automaton Flip-flop

ON-OFF system

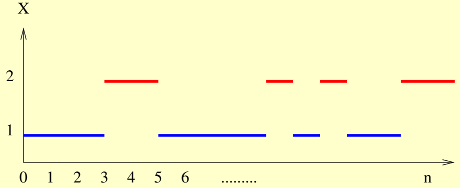
Two states model :
- communication line
- processor activity
- ...



Parameters :
- proportion of transitions : p, q
- mean sojourn time in state 1 : $\frac{1}{p}$
- mean sojourn time in state 2 : $\frac{1}{q}$

Trajectory

X_n state of the automaton at time n.



Transient distribution

$$\pi_n(1) = \mathbb{P}(X_n = 1);$$

$$\pi_n(2) = \mathbb{P}(X_n = 2)$$

Problem

Estimation of π_n : state prevision, resource utilization

Mathematical model

Transition probabilities

$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$$

$$\begin{aligned} \mathbb{P}(X_{n+1} = 1 | X_n = 1) &= 1-p; \\ \mathbb{P}(X_{n+1} = 2 | X_n = 1) &= p; \\ \mathbb{P}(X_{n+1} = 1 | X_n = 2) &= q; \\ \mathbb{P}(X_{n+1} = 2 | X_n = 2) &= 1-q. \end{aligned}$$

$$\begin{cases} \pi_{n+1}(1) = \pi_n(1)(1-p) + \pi_n(2)q; \\ \pi_{n+1}(2) = \pi_n(1)p + \pi_n(2)(1-q); \end{cases}$$

$$\pi_{n+1} = \pi_n P$$

Linear iterations

Spectrum of P (eigenvalues)

$\mathcal{S}p = \{1, 1-p-q\}$

System resolution

$|1-p-q| < 1$ Non pathologic case

$$\begin{cases} \pi_n(1) = \frac{q}{p+q} + \left(\pi_1(0) - \frac{q}{p+q}\right)(1-p-q)^n; \\ \pi_n(2) = \frac{p}{p+q} + \left(\pi_2(0) - \frac{p}{p+q}\right)(1-p-q)^n; \end{cases}$$

$1-p-q = 1$ $p = q = 0$ Reducible behavior

$1-p-q = -1$ $p = q = -1$ Periodic behavior

Mathematical model

Transition probabilities

$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$$

$$\begin{aligned} \mathbb{P}(X_{n+1} = 1 | X_n = 1) &= 1-p; \\ \mathbb{P}(X_{n+1} = 2 | X_n = 1) &= p; \\ \mathbb{P}(X_{n+1} = 1 | X_n = 2) &= q; \\ \mathbb{P}(X_{n+1} = 2 | X_n = 2) &= 1-q. \end{aligned}$$

$$\begin{cases} \pi_{n+1}(1) = \pi_n(1)(1-p) + \pi_n(2)q; \\ \pi_{n+1}(2) = \pi_n(1)p + \pi_n(2)(1-q); \end{cases}$$

$$\pi_{n+1} = \pi_n P$$

Linear iterations

Spectrum of P (eigenvalues)

$\mathcal{S}p = \{1, 1-p-q\}$

System resolution

$|1-p-q| < 1$ Non pathologic case

$$\begin{cases} \pi_n(1) = \frac{q}{p+q} + \left(\pi_1(0) - \frac{q}{p+q}\right)(1-p-q)^n; \\ \pi_n(2) = \frac{p}{p+q} + \left(\pi_2(0) - \frac{p}{p+q}\right)(1-p-q)^n; \end{cases}$$

$1-p-q = 1$ $p = q = 0$ Reducible behavior

$1-p-q = -1$ $p = q = -1$ Periodic behavior

Mathematical model

Transition probabilities

$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$$

$$\begin{aligned} \mathbb{P}(X_{n+1} = 1 | X_n = 1) &= 1-p; \\ \mathbb{P}(X_{n+1} = 2 | X_n = 1) &= p; \\ \mathbb{P}(X_{n+1} = 1 | X_n = 2) &= q; \\ \mathbb{P}(X_{n+1} = 2 | X_n = 2) &= 1-q. \end{aligned}$$

$$\left\{ \begin{array}{l} \pi_{n+1}(1) = \pi_n(1)(1-p) + \pi_n(2)q; \\ \pi_{n+1}(2) = \pi_n(1)p + \pi_n(2)(1-q); \end{array} \right.$$

$$\pi_{n+1} = \pi_n P$$

Linear iterations

Spectrum of P (eigenvalues)

$\mathcal{S}p = \{1, 1-p-q\}$

System resolution

$|1-p-q| < 1$ Non pathologic case

$$\left\{ \begin{array}{l} \pi_n(1) = \frac{q}{p+q} + \left(\pi_1(0) - \frac{q}{p+q}\right)(1-p-q)^n; \\ \pi_n(2) = \frac{p}{p+q} + \left(\pi_2(0) - \frac{p}{p+q}\right)(1-p-q)^n; \end{array} \right.$$

$1-p-q = 1$ $p = q = 0$ Reducible behavior

$1-p-q = -1$ $p = q = -1$ Periodic behavior

Mathematical model

Transition probabilities

$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$$

$$\begin{aligned} \mathbb{P}(X_{n+1}=1|X_n=1) &= 1-p; \\ \mathbb{P}(X_{n+1}=2|X_n=1) &= p; \\ \mathbb{P}(X_{n+1}=1|X_n=2) &= q; \\ \mathbb{P}(X_{n+1}=2|X_n=2) &= 1-q. \end{aligned}$$

$$\begin{cases} \pi_{n+1}(1) = \pi_n(1)(1-p) + \pi_n(2)q; \\ \pi_{n+1}(2) = \pi_n(1)p + \pi_n(2)(1-q); \end{cases}$$

$$\pi_{n+1} = \pi_n P$$

Linear iterations

Spectrum of P (eigenvalues)

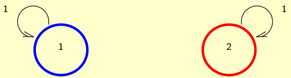
$\mathcal{S}p = \{1, 1-p-q\}$

System resolution

$|1-p-q| < 1$ Non pathologic case

$$\begin{cases} \pi_n(1) = \frac{q}{p+q} + \left(\pi_1(0) - \frac{q}{p+q}\right)(1-p-q)^n; \\ \pi_n(2) = \frac{p}{p+q} + \left(\pi_2(0) - \frac{p}{p+q}\right)(1-p-q)^n; \end{cases}$$

$1-p-q = 1$ $p = q = 0$ Reducible behavior



$1-p-q = -1$ $p = q = -1$ Periodic behavior

Mathematical model

Transition probabilities

$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$$

$$\begin{array}{l} \mathbb{P}(X_{n+1}=1|X_n=1) = 1-p; \\ \mathbb{P}(X_{n+1}=2|X_n=1) = p; \\ \mathbb{P}(X_{n+1}=1|X_n=2) = q; \\ \mathbb{P}(X_{n+1}=2|X_n=2) = 1-q. \end{array}$$

$$\left\{ \begin{array}{l} \pi_{n+1}(1) = \pi_n(1)(1-p) + \pi_n(2)q; \\ \pi_{n+1}(2) = \pi_n(1)p + \pi_n(2)(1-q); \end{array} \right.$$

$$\pi_{n+1} = \pi_n P$$

Linear iterations

Spectrum of P (eigenvalues)

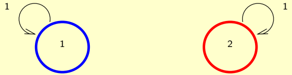
$\mathcal{S}p = \{1, 1-p-q\}$

System resolution

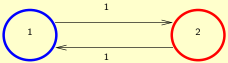
$|1-p-q| < 1$ Non pathologic case

$$\left\{ \begin{array}{l} \pi_n(1) = \frac{q}{p+q} + \left(\pi_1(0) - \frac{q}{p+q}\right)(1-p-q)^n; \\ \pi_n(2) = \frac{p}{p+q} + \left(\pi_2(0) - \frac{p}{p+q}\right)(1-p-q)^n; \end{array} \right.$$

$1-p-q = 1$ $p = q = 0$ Reducible behavior



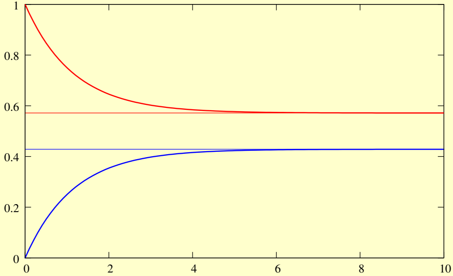
$1-p-q = -1$ $p = q = -1$ Periodic behavior



Recurrent behavior

Numerical example

$p = \frac{1}{4}$, $q = \frac{1}{3}$



Rapid convergence
(exponential rate)

Steady state behavior

$$\begin{cases} \pi_\infty(1) = \frac{q}{p+q}; \\ \pi_\infty(2) = \frac{p}{p+q}. \end{cases}$$

π_∞ unique probability vector solution

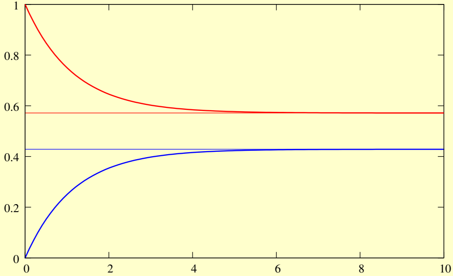
$$\pi_\infty = \pi_\infty P.$$

If $\pi_0 = \pi_\infty$ then $\pi_n = \pi_\infty$ for all n
stationary behavior

Recurrent behavior

Numerical example

$p = \frac{1}{4}$, $q = \frac{1}{3}$



Rapid convergence
(exponential rate)

Steady state behavior

$$\begin{cases} \pi_\infty(1) = \frac{q}{p+q}; \\ \pi_\infty(2) = \frac{p}{p+q}. \end{cases}$$

π_∞ unique probability vector solution

$$\pi_\infty = \pi_\infty P.$$

If $\pi_0 = \pi_\infty$ then $\pi_n = \pi_\infty$ for all n
stationary behavior

Théorème de convergence en loi

Soit $\{X_n\}_{n\in\mathbb{N}}$ une chaîne de Markov homogène, apériodique et irréductible à valeurs dans un espace discret $\mathcal{X}$ alors

- Les limites suivantes existent (et ne dépendent pas de i)

$$\lim_{n\to+\infty} \mathbb{P}(X_n = j | X_0 = i) = \pi_j;$$

- π est l'unique vecteur de probabilité invariant par P

$$\pi P = \pi;$$

- La convergence est rapide ; il existe $C > 0$ et $0 < \alpha < 1$ tels que

$$||\mathbb{P}(X_n = j | X_0 = i) - \pi_j|| \leqslant C.\alpha^n.$$

On note

$$X_n \xrightarrow{\mathcal{L}} X_\infty;$$

avec X_∞ de loi π
π est appelée **probabilité stationnaire** associée à la chaîne

Preuve 1 : Cas fini approche matricielle

Calculatoire $P > 0$

schéma contractant sur $\max_i p_{i,j}^{(n)} - \min_i p_{i,j}^{(n)}$

Perron-Froebenius $P > 0$

La matrice P est stochastique à coefficients strictement positifs donc son rayon spectral $r = 1$ est valeur propre de multiplicité 1, son vecteur propre est à coefficients strictement positifs, les autres v.p sont de module < 1 cqfd

Cas $P \geqslant 0$

Apériodique et irréductible $\Rightarrow$ il existe k tel que $P^k > 0$ et on applique le résultat ci-dessus

Preuve 1 : détail $P > 0$

Soit x et $y = Px$, $\omega = \min_{i,j} p_{i,j}$

$$\overline{x} = \max_i x_i, \ \underline{x} = \min_i x_i.$$

$$y_i = \sum_j p_{i,j} x_j$$

Propriété du barycentre :

$$(1-\omega)\underline{x} + \omega\overline{x} \leqslant y_i \leqslant (1-\omega)\overline{x} + \omega\underline{x}$$

$$0 \leqslant \overline{y} - \underline{y} \leqslant (1-2\omega)(\overline{x} - \underline{x})$$

$$P^n x \longrightarrow s(x)(1,1,\cdots,1)^t$$

D'où la convergence de P^n vers une matrice ayant toutes ses lignes identiques.

Preuve 2 : Temps de retour

$$\tau_i^+ = \inf\{n \geqslant 1; \;\; X_n = i | X_0 = i\}.$$

alors $\frac{1}{\mathbb{E}\tau_i^+}$ est une mesure invariante (lemme de Kac)

1914-1984

Preuve :

1. $\mathbb{E}\tau_i^+ < \infty$
2. Etude sur un intervalle de régénération (propriété de Markov forte)
3. Unicité par fonctions harmoniques

Preuve 2 : détail

Preuve 3 : Couplage

Soit $\{X_n\}_{n\in\mathbb{N}}$ une chaîne de Markov homogène de loi initiale $\pi(0)$ et de probabilité stationaire π.

Soit $\left\{\tilde{X}_n\right\}_{n\in\mathbb{N}}$ une chaîne de Markov de loi initiale $\tilde{\pi}$ ayant les mêmes probabilités de transition que $\{X_n\}$

$\{X_n\}$ et $\left\{\tilde{X}_n\right\}$ indépendantes

- $Z_n = (X_n, \tilde{X}_n)$ est une chaîne de Markov homogène
- Si $\{X_n\}$ est apériodique et irréductible, il en est de même pour Z_n

Il existe un temps τ d'atteinte de la diagonale, $\tau < \infty$ P-p.s. alors

$$|\mathbb{P}(X_n = i) - \mathbb{P}(\tilde{X}_n = i)| < 2\mathbb{P}(\tau > n)$$

$$|\mathbb{P}(X_n = i) - \pi(i)| < 2\mathbb{P}(\tau > n) \longrightarrow 0.$$

Preuve 3bis : Couplage

Supposons que $\{X_n\}_{n\in\mathbb{N}}$ soit représentée sous la forme

$$X_{n+1} = \Phi(X_n, \xi_{n+1}),$$

avec $\{\xi_n\}_{n\in\mathbb{N}}$ i.i.d.
Hypothèse : "il existe un motif synchronisant"
Soit $\pi(0)$ loi initiale de X_n
Soit $\left\{\tilde{X}_n\right\}_{n\in\mathbb{N}}$ une chaîne de Markov de loi initiale $\tilde{\pi}$ ayant les mêmes innovations que $\{X_n\}$
Il existe un temps τ' d'atteinte de la diagonale, $\tau' < \infty$ P-p.s. alors

$$|\mathbb{P}(X_n = i) - \mathbb{P}(\tilde{X}_n = i)| < 2\mathbb{P}(\tau > n);$$

$$|\mathbb{P}(X_n = i) - \pi(i)| < 2\mathbb{P}(\tau > n) \longrightarrow 0.$$

Théorème ergodique

Soit $\{X_n\}_{n\in\mathbb{N}}$ une chaîne de Markov homogène, apériodique et irréductible à valeurs dans un espace discret $\mathcal{X}$ de probabilité stationnaire π alors

- pour toute fonction f vérifiant $\mathbb{E}_\pi|f| < +\infty$

$$\frac{1}{N}\sum_{n=1}^{N} f(X_n) \stackrel{P-p.s.}{\longrightarrow} \mathbb{E}_\pi f.$$

équivalent de la *loi des grands nombres*

- Si $\mathbb{E}_\pi f = 0$ alors il existe σ tel que

$$\frac{1}{\sigma\sqrt{N}}\sum_{n=1}^{N} f(X_n) \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N}(0,1).$$

équivalent du *théorème central limite*

Problème fondamental

Etant donnée une fonction f (de coût, récompense,...) estimer

$$\mathbb{E}_\pi f$$

et donner la qualité de cette estimation.

Solving methods

Solving $\pi = \pi P$

- Analytical/approximation methods
- Formal methods $N \leqslant 50$
  Maple, Sage,...
- Direct numerical methods $N \leqslant 1000$
  Mathematica, Scilab,...
- Iterative methods with preconditioning $N \leqslant 100,000$
  Marca,...
- Adapted methods (structured Markov chains) $N \leqslant 1,000,000$
  PEPS,...
- Monte-Carlo simulation $N \geqslant 10^7$

Postprocessing of the stationary distribution

Computation of rewards (expected stationary functions)
Utilization, response time,...

Ergodic Sampling(1)

Ergodic sampling algorithm

Representation : **transition fonction**

$$X_{n+1} = \Phi(X_n, e_{n+1}).$$

$x \leftarrow x_0$
{choice of the initial state at time =0}
$n = 0$;
répéter
  $n \leftarrow n + 1$;
  $e \leftarrow Random\_event()$;
  $x \leftarrow \Phi(x, e)$;
  Store x
  {computation of the next state X_{n+1}}
jusqu'à some empirical criteria
return the trajectory

Problem : Stopping criteria

Ergodic Sampling(2)

Start-up

Convergence to stationary behavior

$$\lim_{n \to +\infty} \mathbb{P}(X_n = x) = \pi_x.$$

Warm-up period : Avoid initial state dependence
Estimation error :

$$||\mathbb{P}(X_n = x) - \pi_x|| \leqslant C\lambda_2^n.$$

λ_2 second greatest eigenvalue of the transition matrix
- bounds on C and λ_2 (spectral gap)
- cut-off phenomena

λ_2 and C non reachable in practice
(complexity equivalent to the computation of π)
some known results (Birth and Death processes)

Ergodic Sampling(3)

Estimation quality

Ergodic theorem :

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} f(X_i) = \mathbb{E}_\pi f.$$

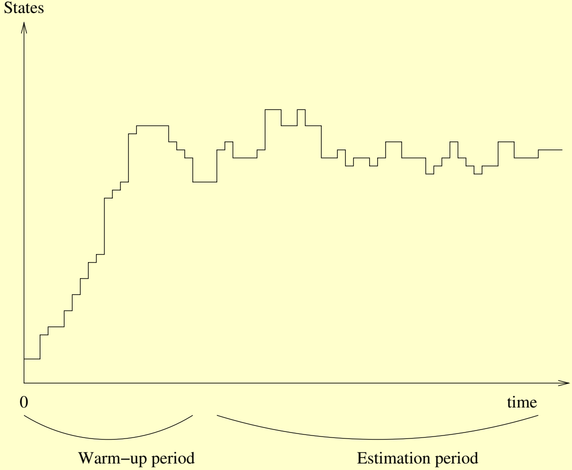
Length of the sampling : Error control (CLT theorem)

Complexity

Complexity of the transition function evaluation (computation of $\Phi(x,.)$)
Related to the stabilization period + Estimation time

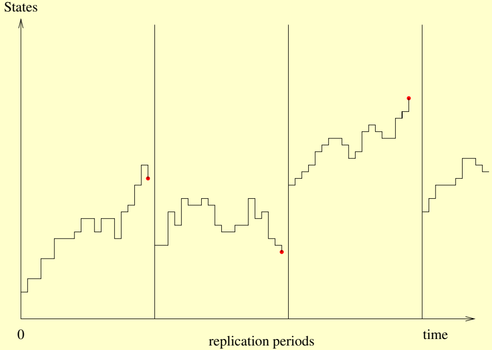
Ergodic sampling(4)

Typical trajectory



Replication Method

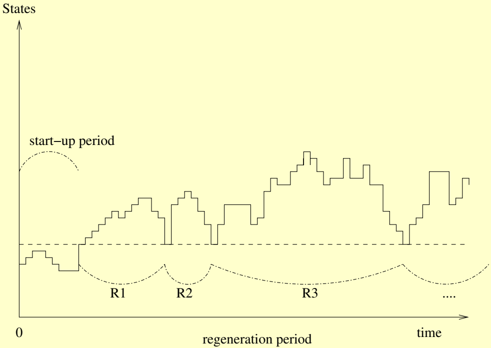
Typical trajectory



Sample of independent states
Drawback : length of the replication period (dependence from initial state)

Regeneration Method

Typical trajectory



Sample of independent trajectories
Drawback : length of the regeneration period (choice of the regenerative state)

Bibliography

Books

- O. Haggstrom. **Finite Markov Chains and algorithmic applications**. Cambridge University Press. 2001.
- P. Brémaud **Markov chains : Gibbs fields, Monte Carlo Simulation and Queues**. Springer-Verlag, 1999
- D. A. Levin, Y. Peres, et E. L. Wilmer **Markov Chains and Mixing Times**, American Mathematical Society,2009.

Websites

- David Wilson's web site
  `http://dimacs.rutgers.edu/~dbwilson/exact`
- Virtual Laboratories in Probability and Statistics
  `http://www.math.uah.edu/stat/index.xhtml`
- The MacTutor History of Mathematics archive (photos)
  `http://www-history.mcs.st-and.ac.uk/`