Monte Carlo Tree Search
Enjeux et affinités

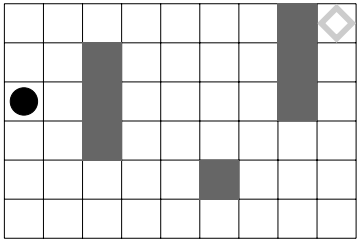
André Fabbri

15 Novembre 2013

Première partie I

Monte Carlo Tree Search

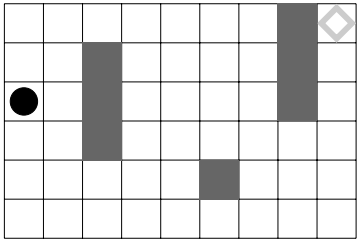
Présentation du problème



Problème du Labyrinthe

- 1 seul agent
- déterministe
- information parfaite

Présentation du problème



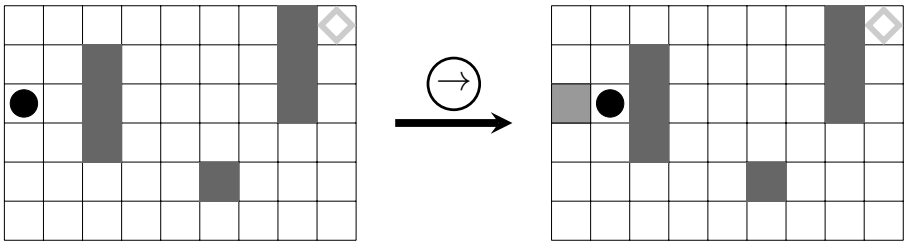
Problème du Labyrinthe

- 1 seul agent
- déterministe
- information parfaite

Description formelle

États : S (initial/final)

Présentation du problème



Problème du Labyrinthe

- 1 seul agent
- déterministe
- information parfaite

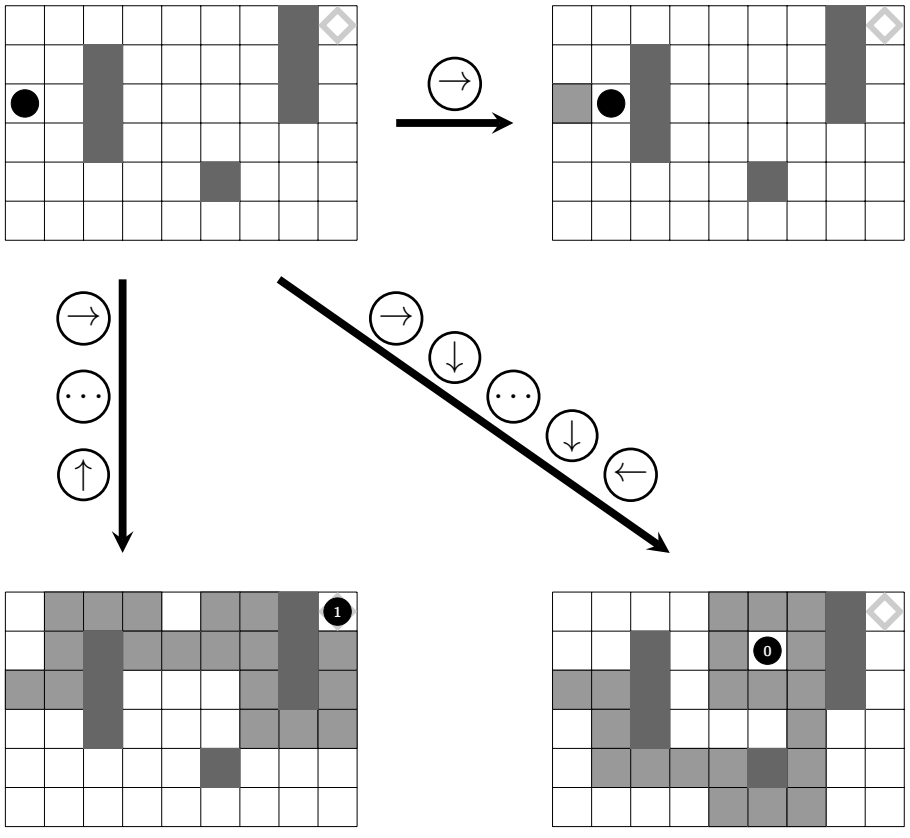
Description formelle

États : S (initial/final)

Actions : $A \uparrow \downarrow \leftarrow \rightarrow$

Objectif : case ◇

Présentation du problème



Problème du Labyrinthe

- 1 seul agent
- déterministe
- information parfaite

Description formelle

États : S (initial/final)

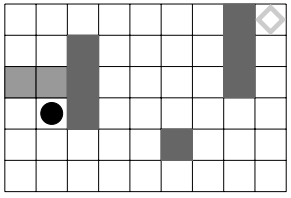
Actions : $A \uparrow \downarrow \leftarrow \rightarrow$

Objectif : case ◇

Récomp : $0/1$

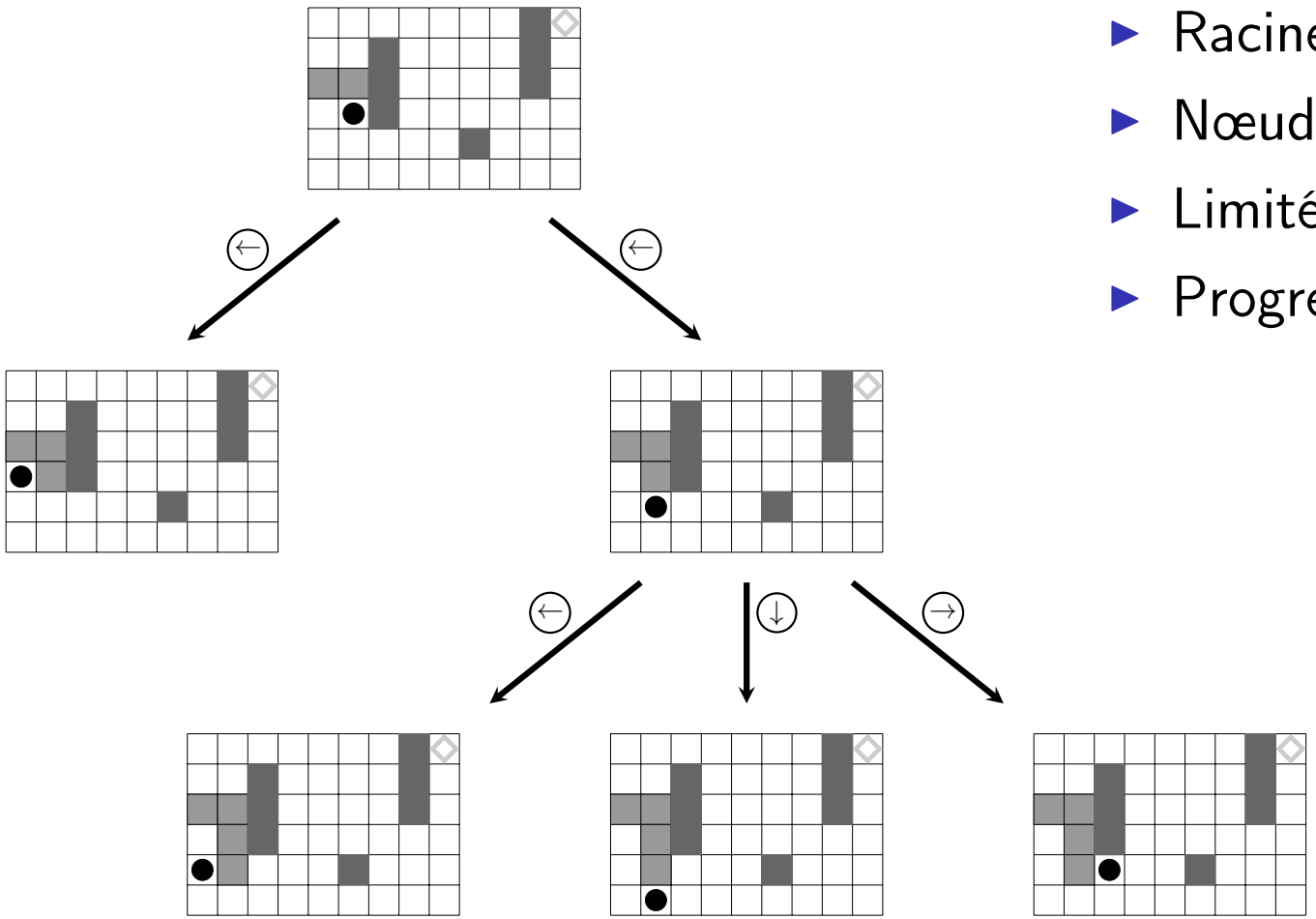
Planning = Trouver une séquence pour atteindre ◇

Arbre de recherche



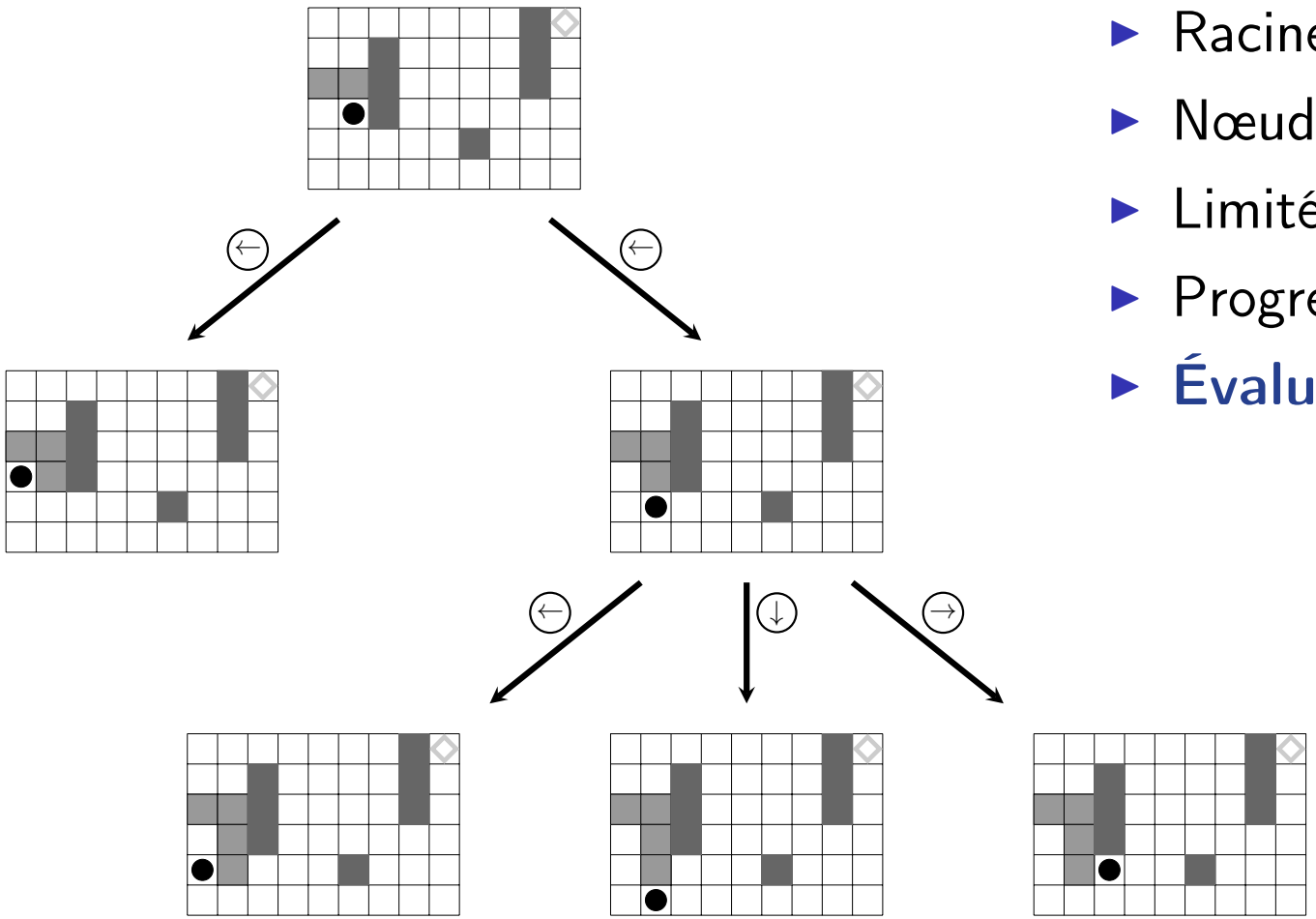
- Racine = État courant

Arbre de recherche



- Racine = État courant
- Nœuds = États possibles
- Limité $|S_{Go}| \simeq 10^{170}$
- Progressif *(a)-symétrique*

Arbre de recherche



- Racine = État courant
- Nœuds = États possibles
- Limité $|S_{Go}| \simeq 10^{170}$
- Progressif *(a)-symétrique*
- **Évaluation ?** *heuristiques ?*

Monte Carlo = Évaluation par échantillonage aléatoire

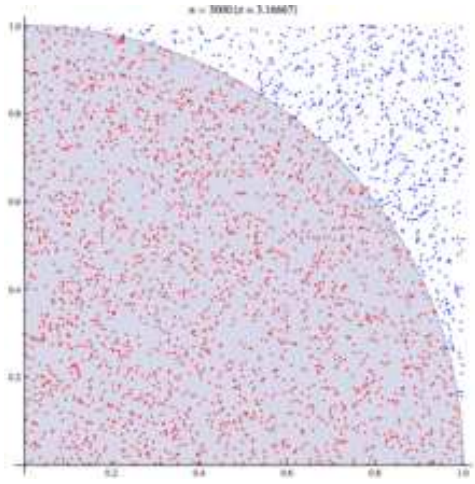


Figure: Approximation de π

Monte Carlo = Évaluation par échantillonage aléatoire

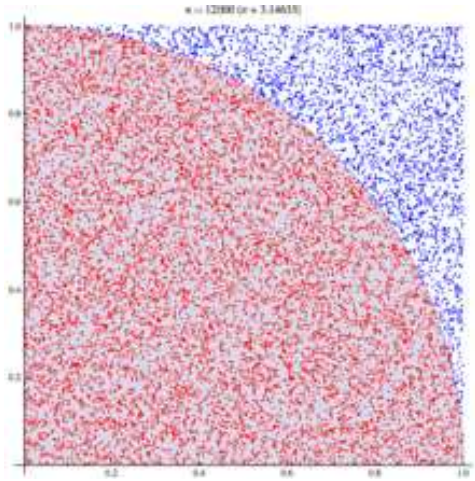


Figure: Approximation de π

Monte Carlo = Évaluation par échantillonage aléatoire

Intérêt de Monte Carlo

- Générique *(faible apport en K)*
- Parallèlisable *(puissance de calcul)*
- « *anytime* » *(temps réel)*

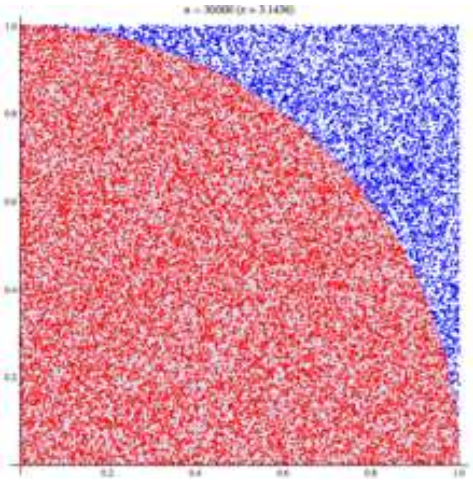


Figure: Approximation de π

Méthodes de Monte Carlo

Monte Carlo = Évaluation par échantillonage aléatoire

Intérêt de Monte Carlo

- Générique *(faible apport en K)*
- Parallèlisable *(puissance de calcul)*
- « *anytime* » *(temps réel)*

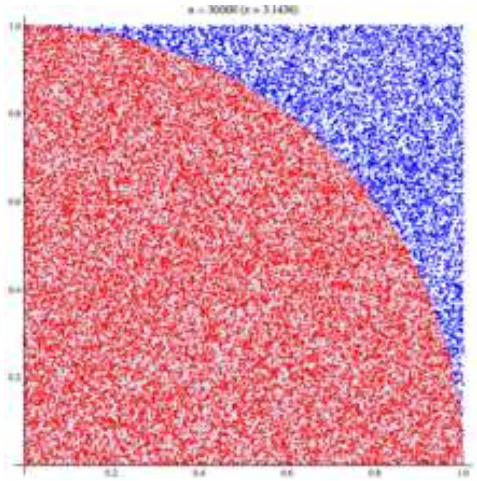
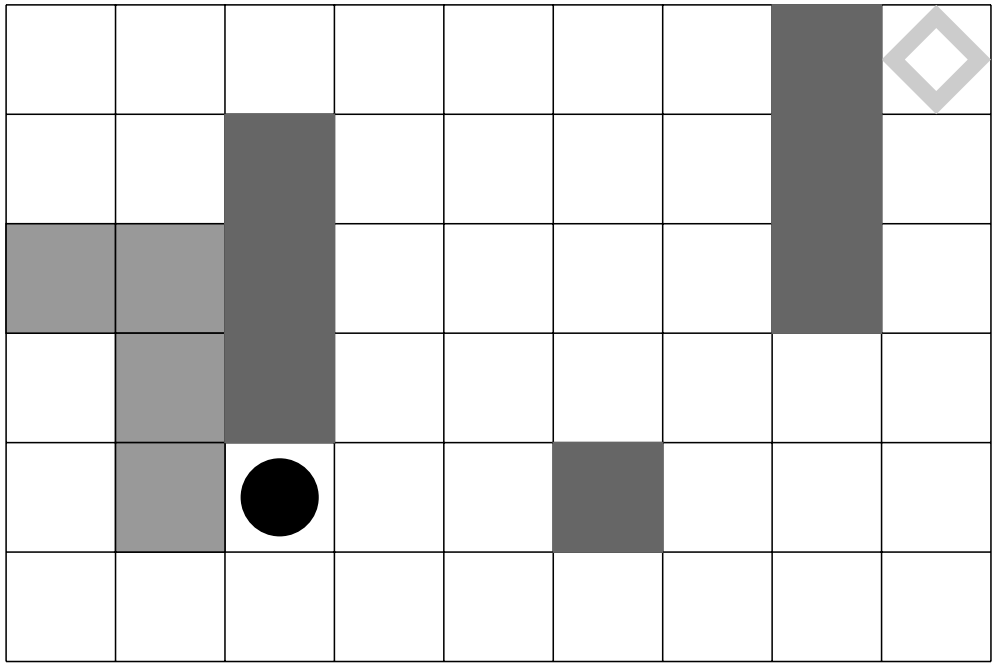


Figure: Approximation de π

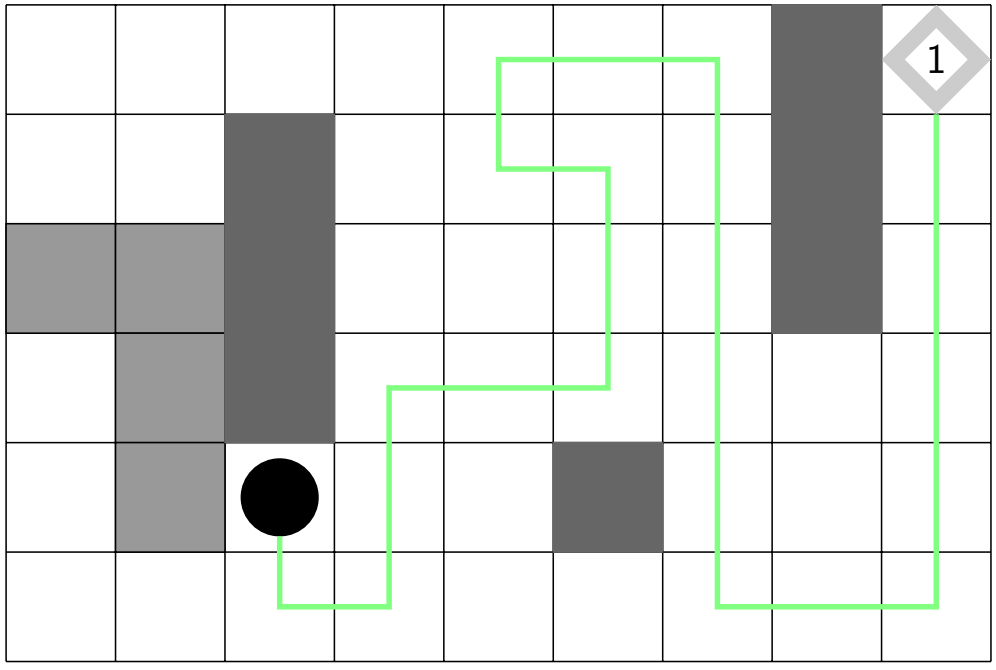
Aléatoire au service du déterministe

Applications

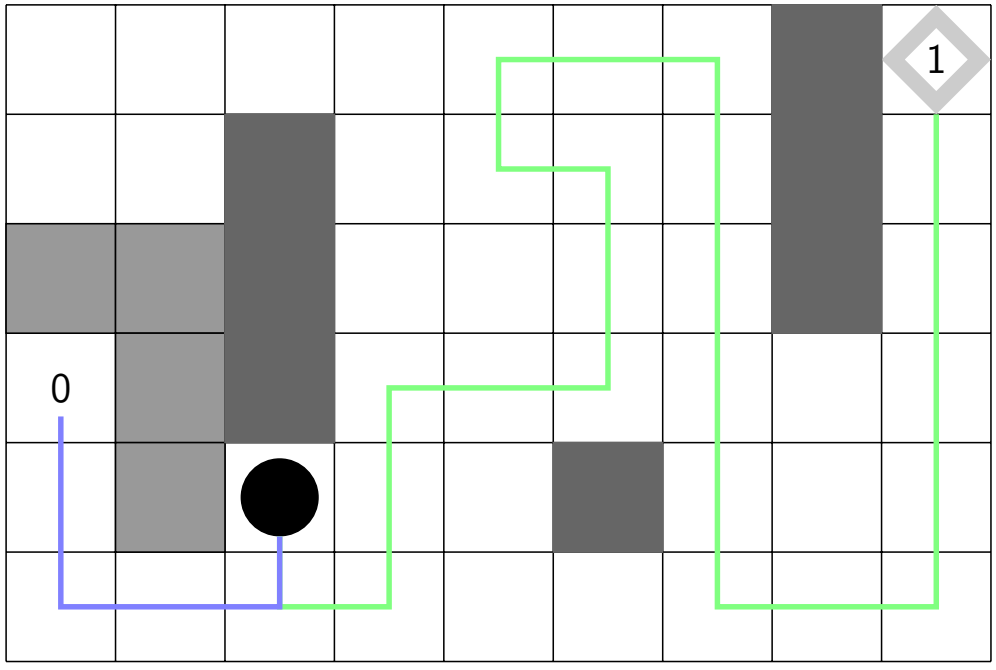
Maths, physique, chimie, IA ... *(intégration, optimisation, météo ... AR)*



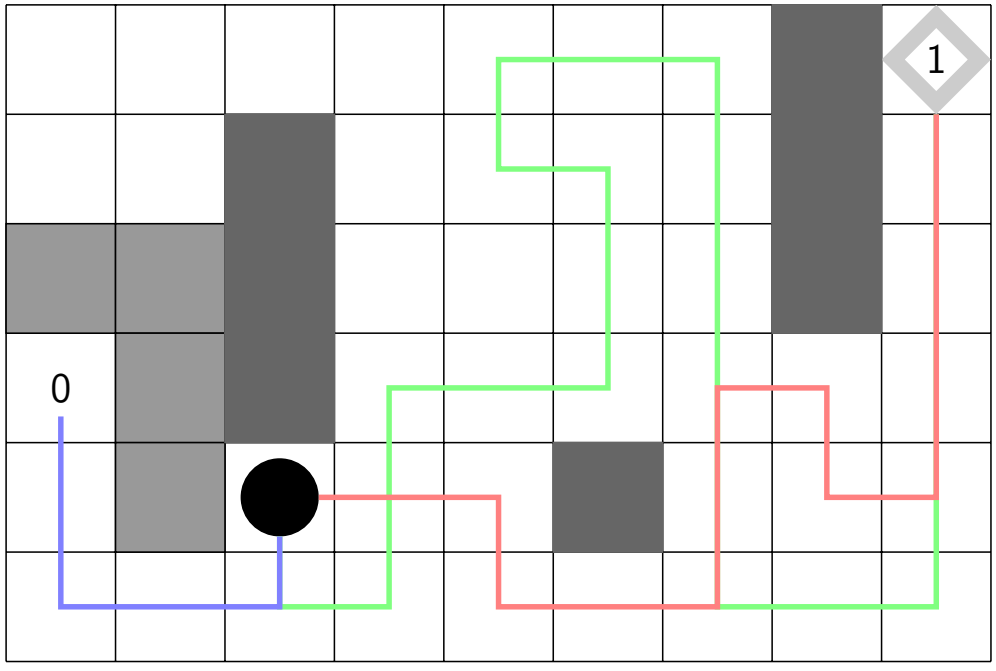
Monte Carlo & Apprentissage par Renforcement



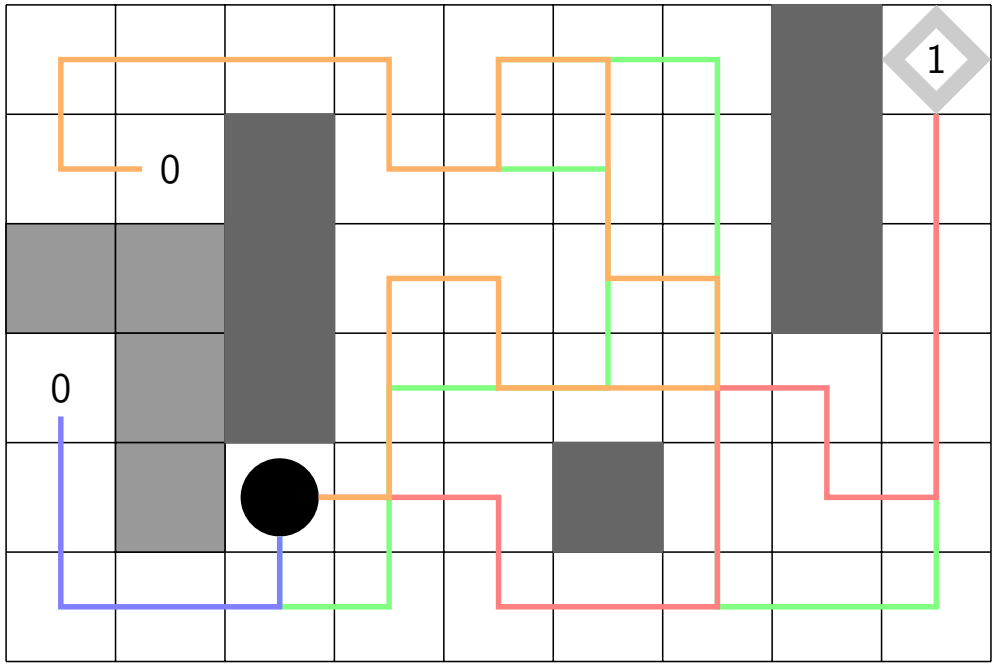
Monte Carlo & Apprentissage par Renforcement



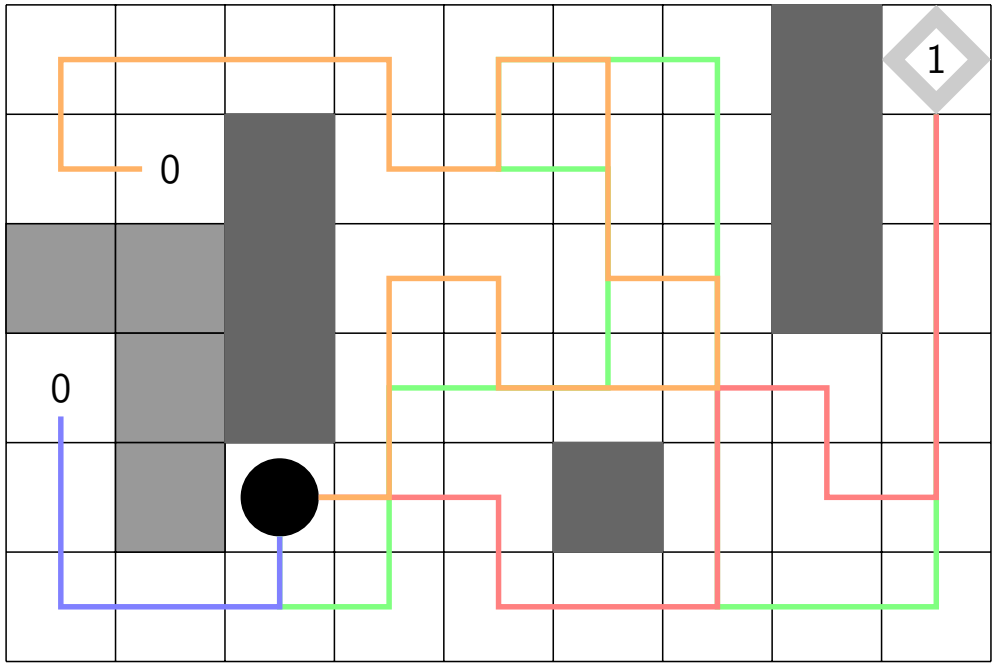
Monte Carlo & Apprentissage par Renforcement



Monte Carlo & Apprentissage par Renforcement

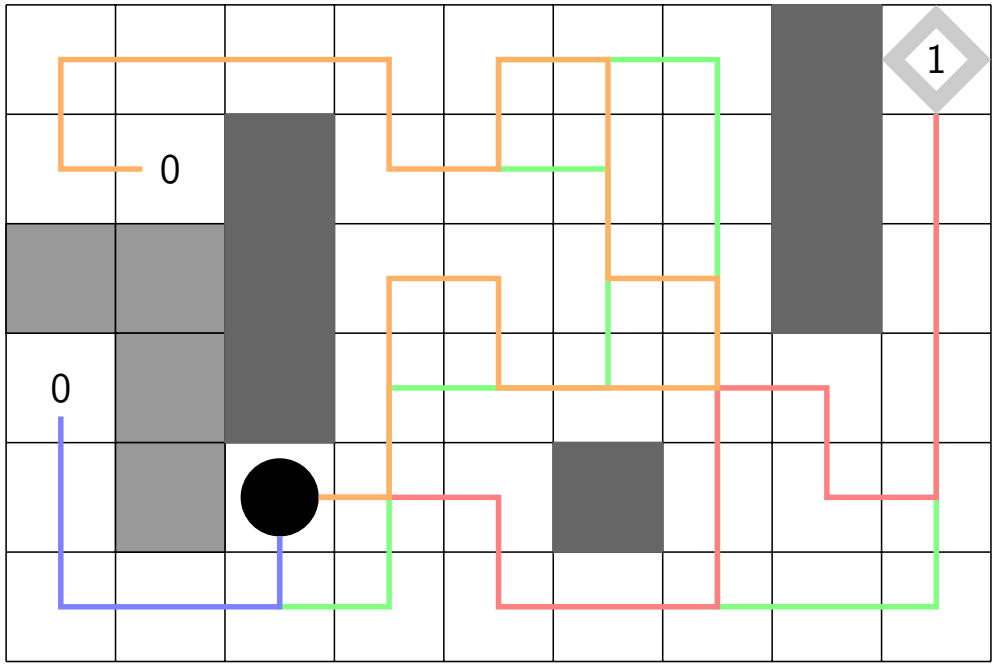


Monte Carlo & Apprentissage par Renforcement

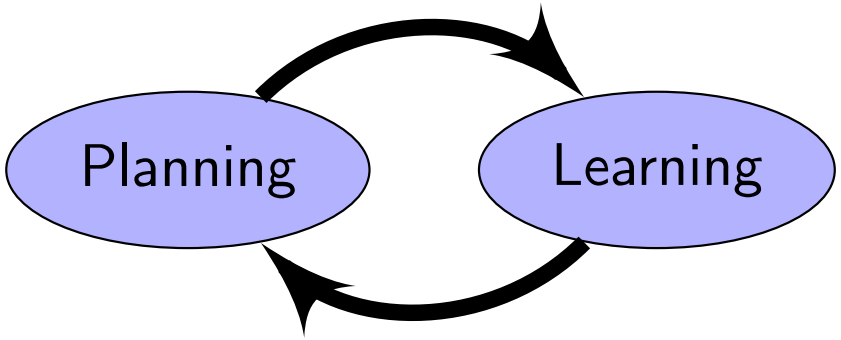


Séquence aléatoire d'action = expérience simulée

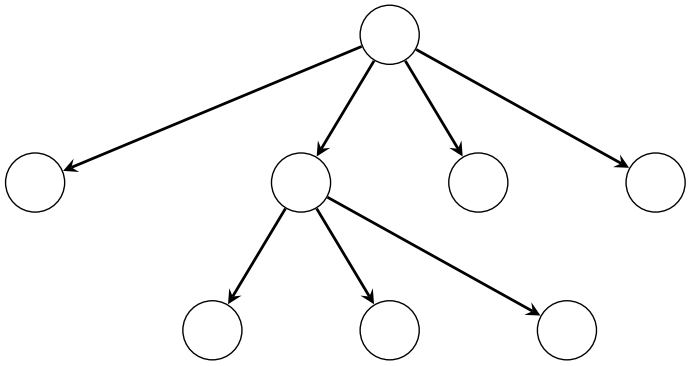
Monte Carlo & Apprentissage par Renforcement



Séquence aléatoire d'action = expérience simulée

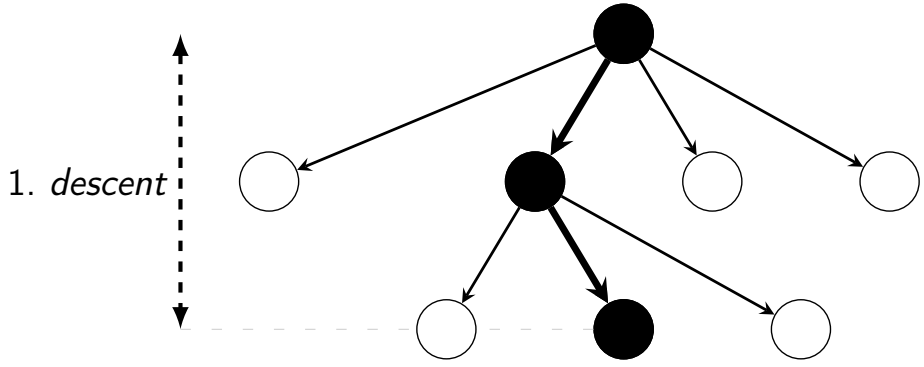


Monte Carlo Tree Search



Boucle d'apprentissage

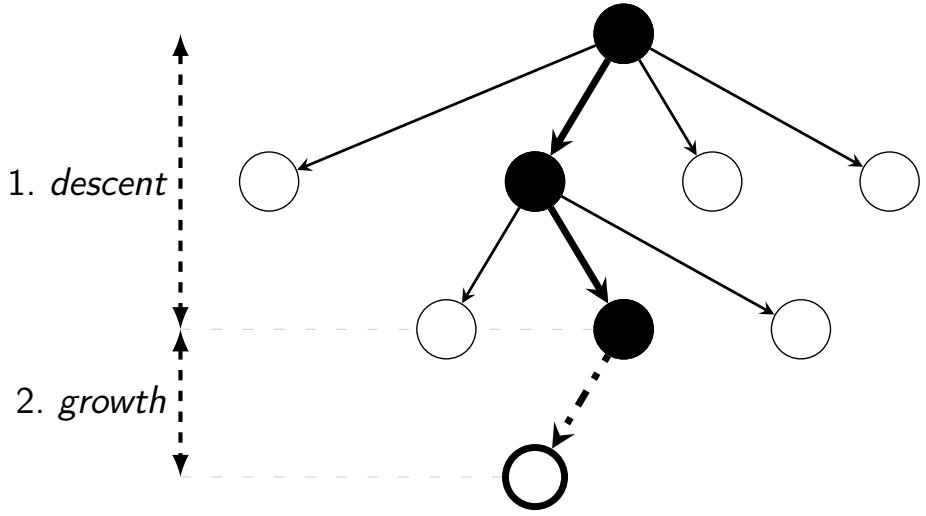
Monte Carlo Tree Search



Boucle d'apprentissage

1. déterministe

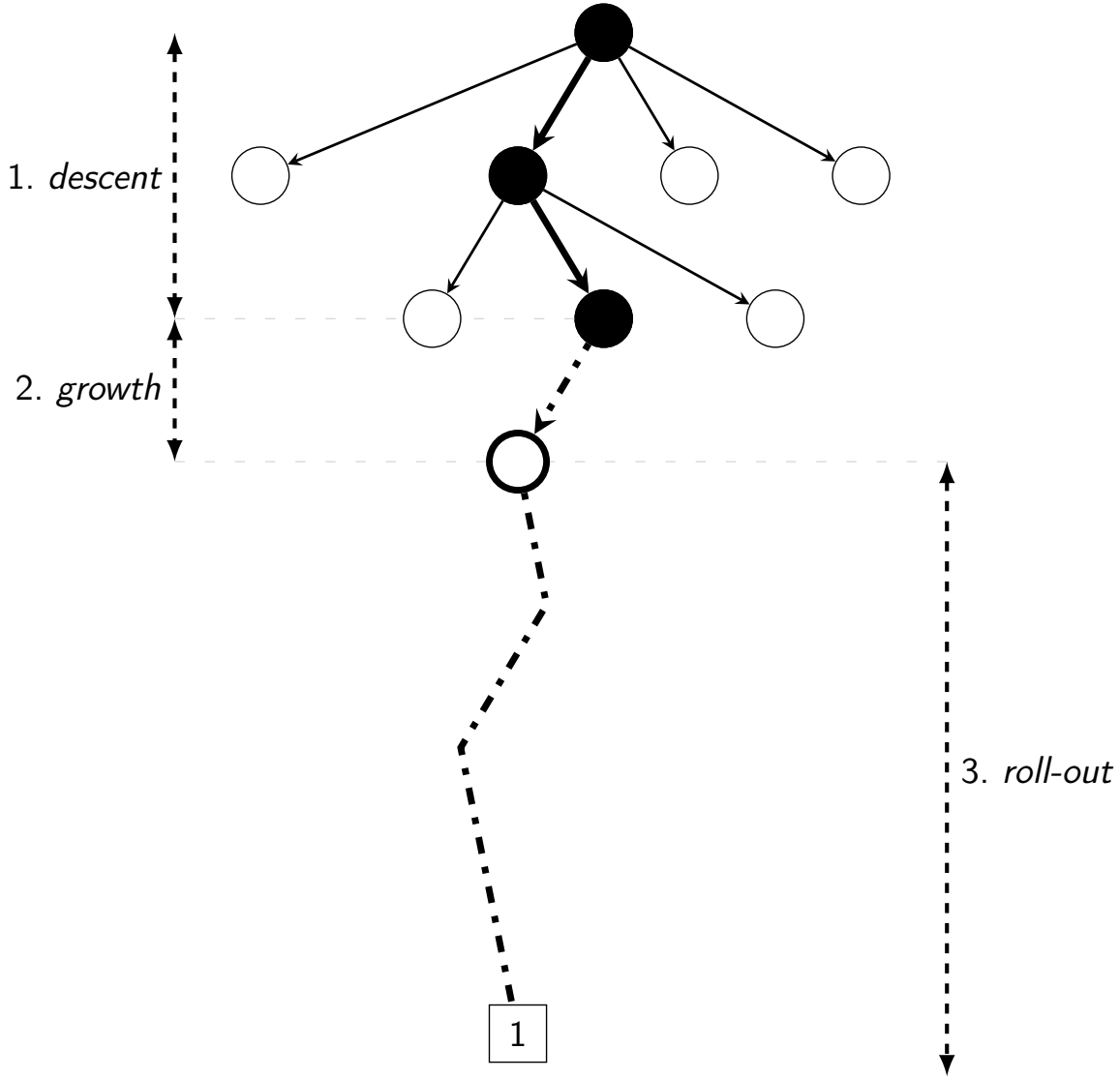
Monte Carlo Tree Search



Boucle d'apprentissage

1. déterministe
2. 1 seul nœud

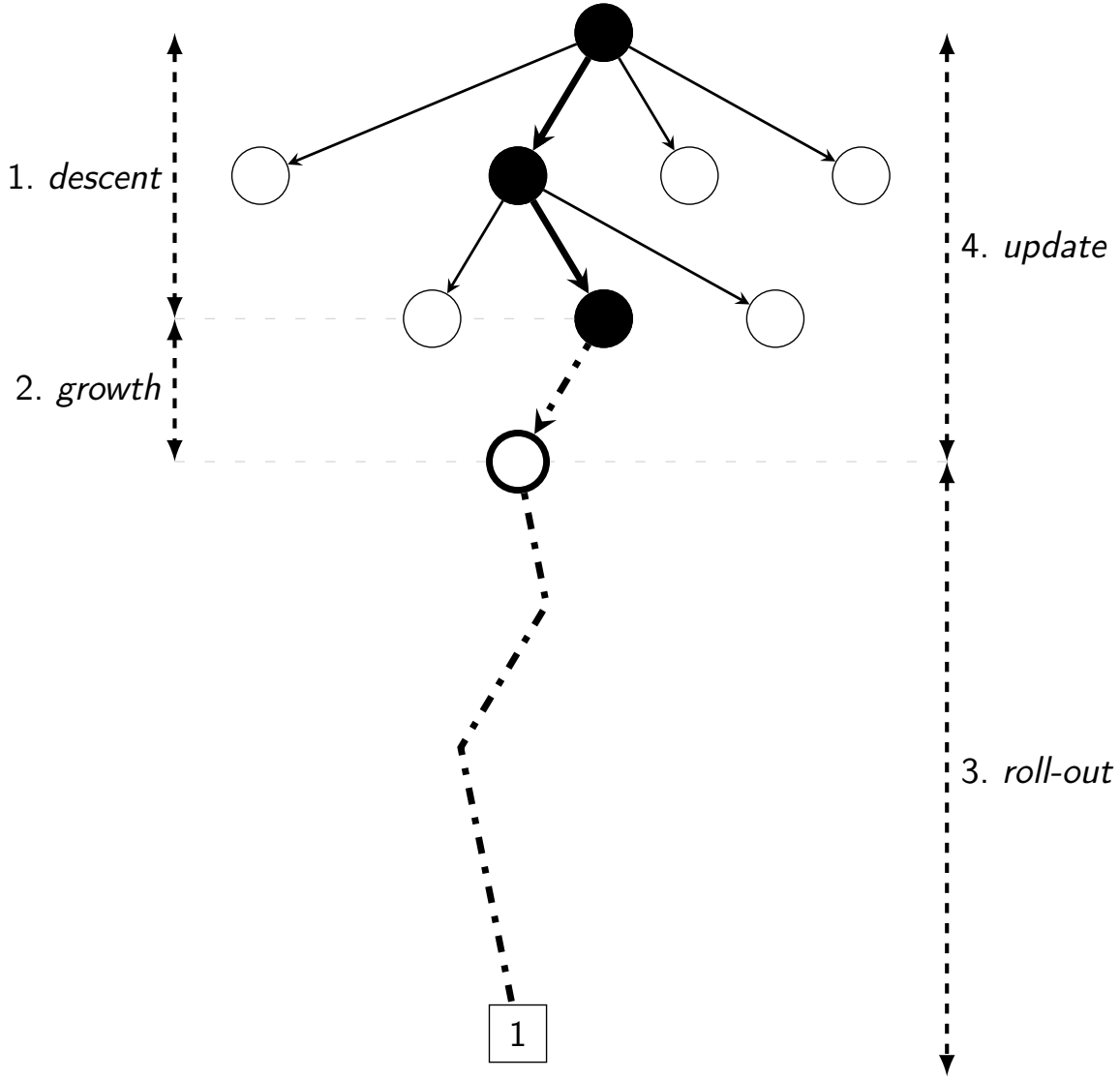
Monte Carlo Tree Search



Boucle d'apprentissage

1. déterministe
2. 1 seul nœud
3. aléatoire

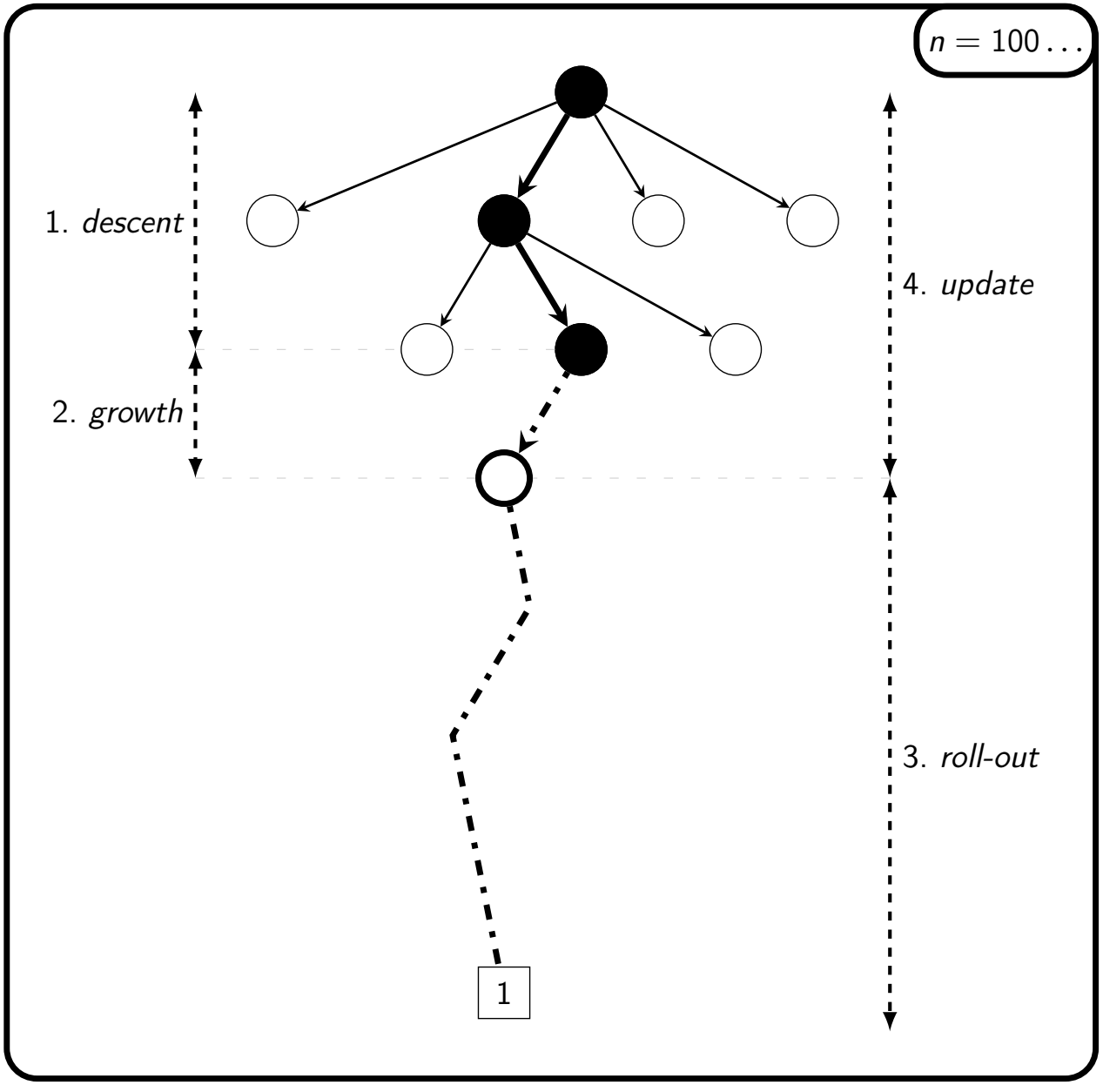
Monte Carlo Tree Search



Boucle d'apprentissage

1. déterministe
2. 1 seul nœud
3. aléatoire
4. 1 simulation

Monte Carlo Tree Search

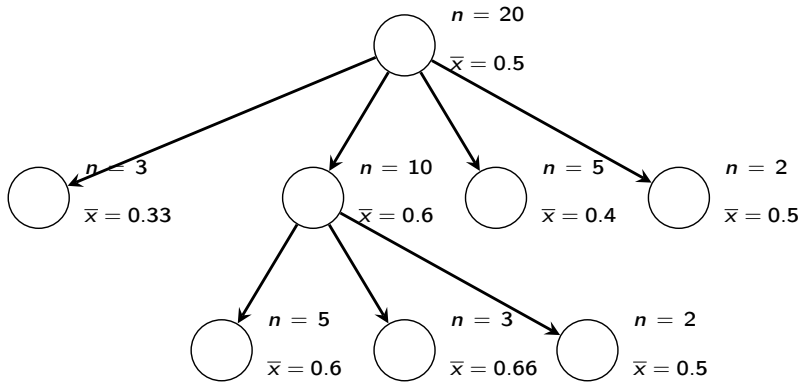


Boucle d'apprentissage

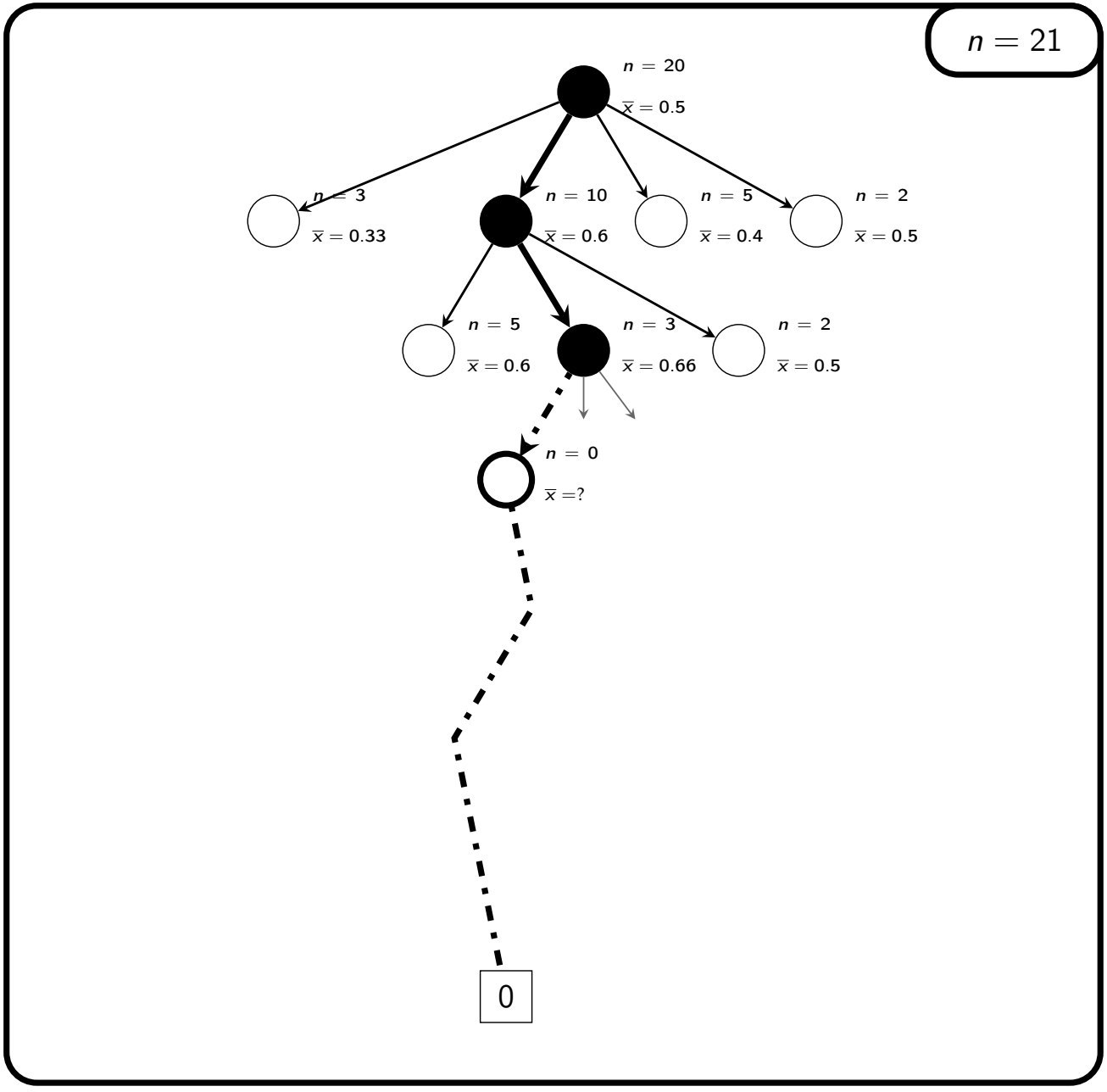
1. déterministe
2. 1 seul nœud
3. aléatoire
4. 1 simulation

Éxécution de l'algorithme

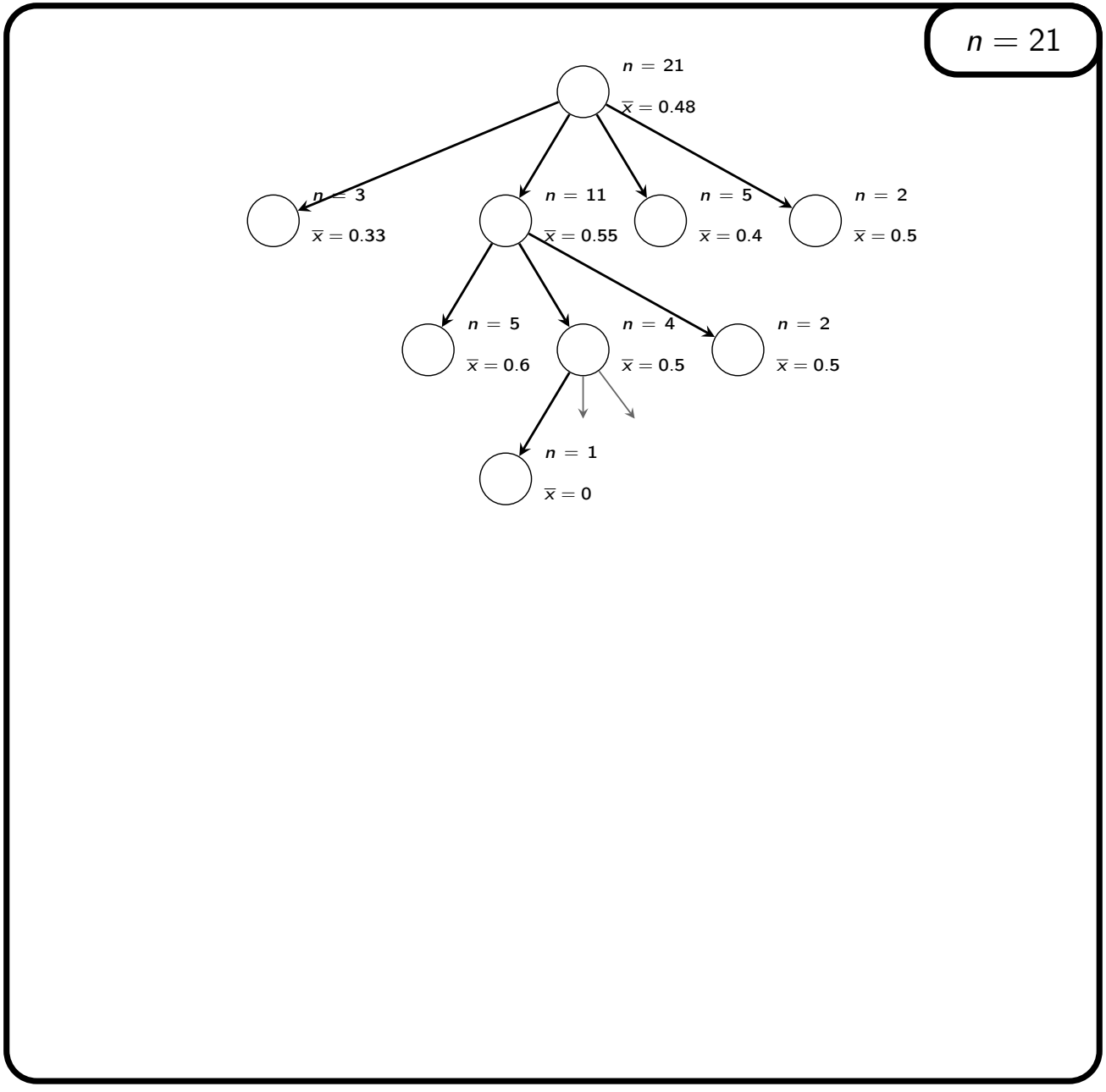
$n = 20$



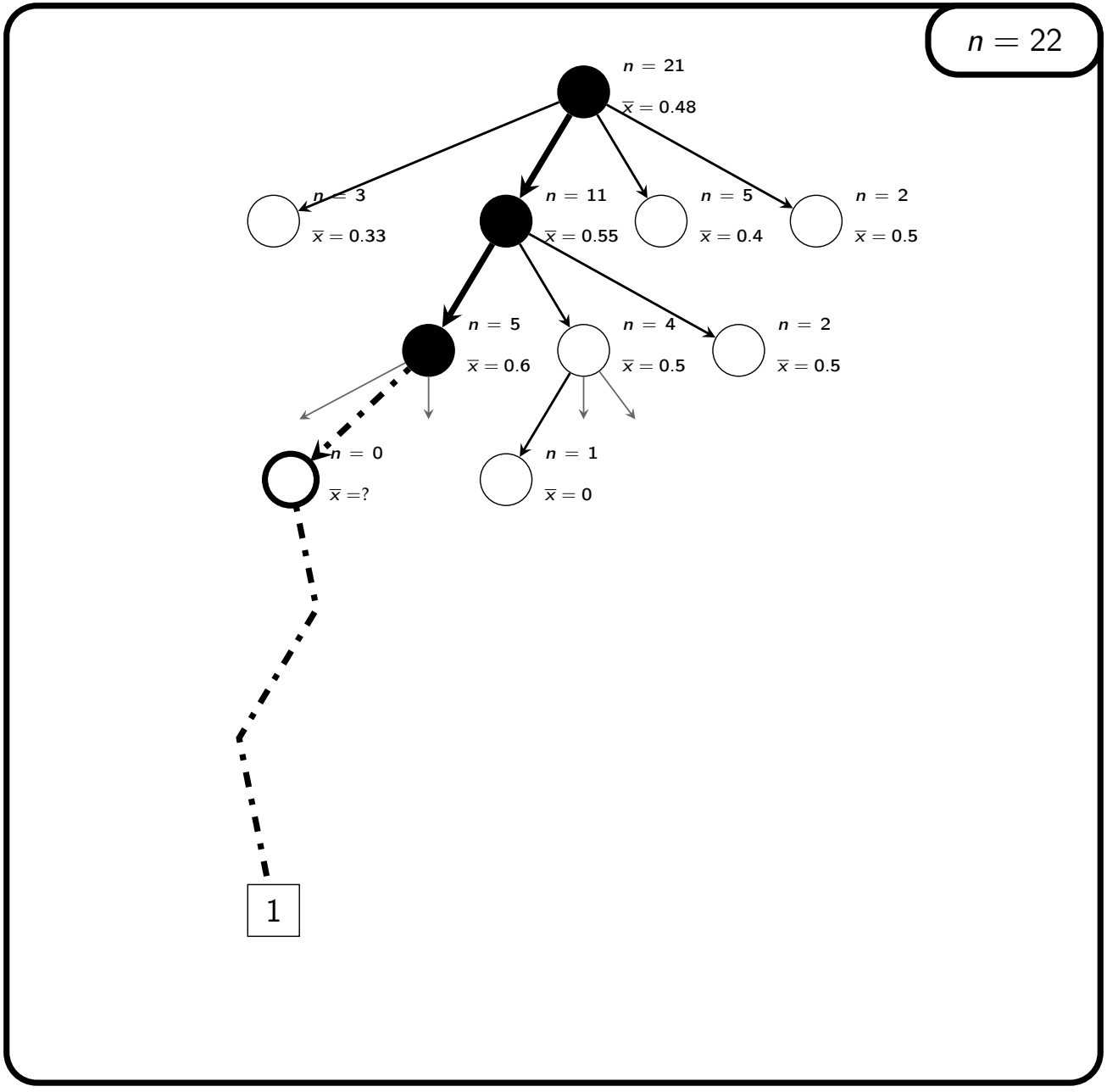
Éxécution de l'algorithme



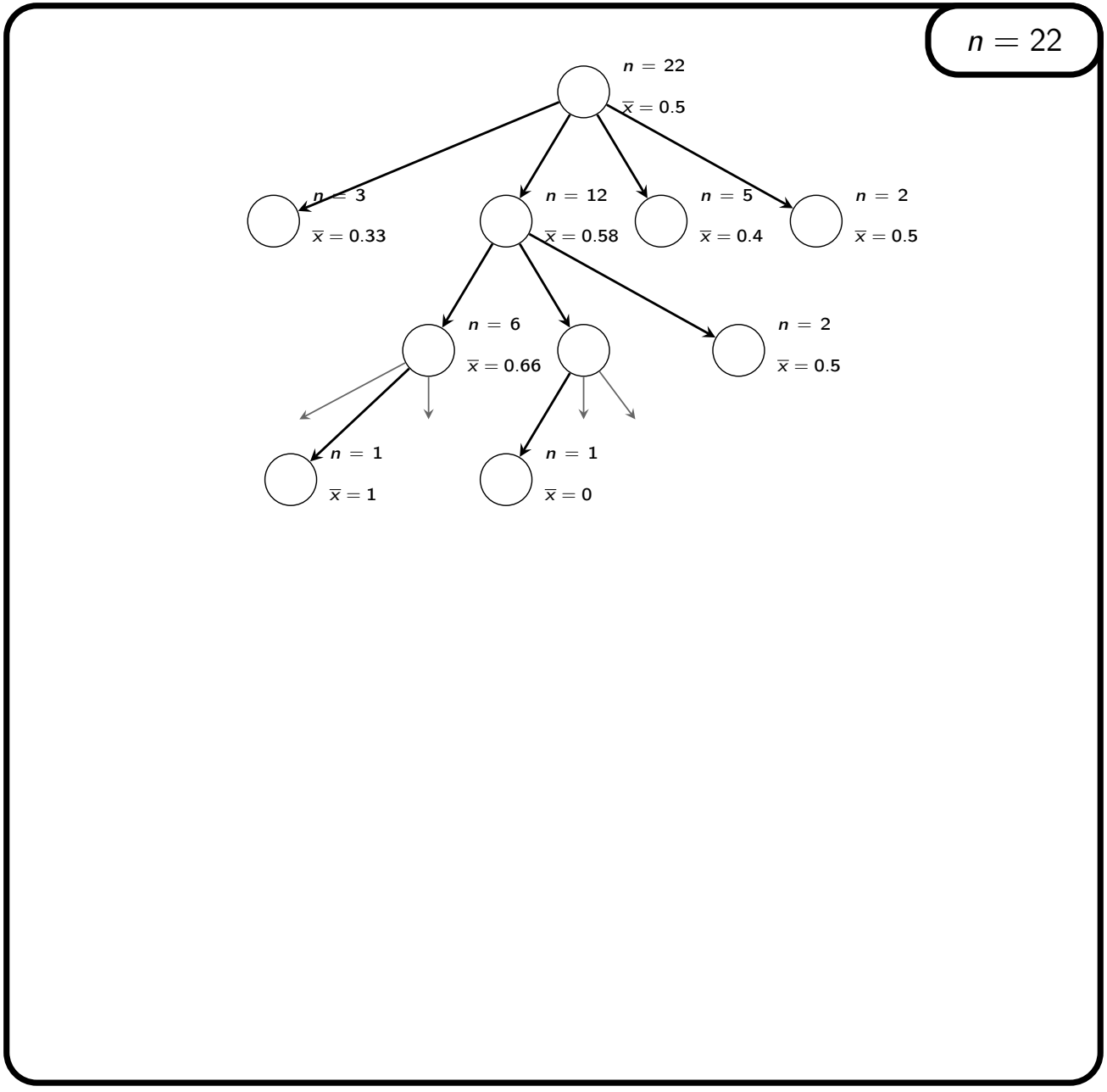
Éxécution de l'algorithme



Éxécution de l'algorithme



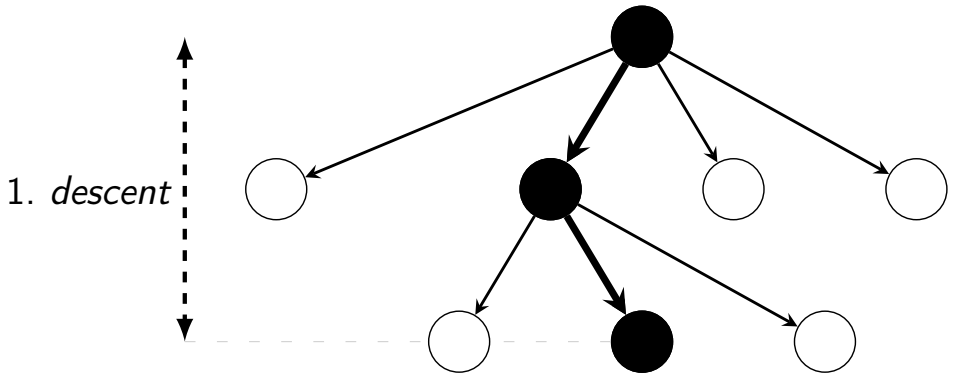
Éxécution de l'algorithme



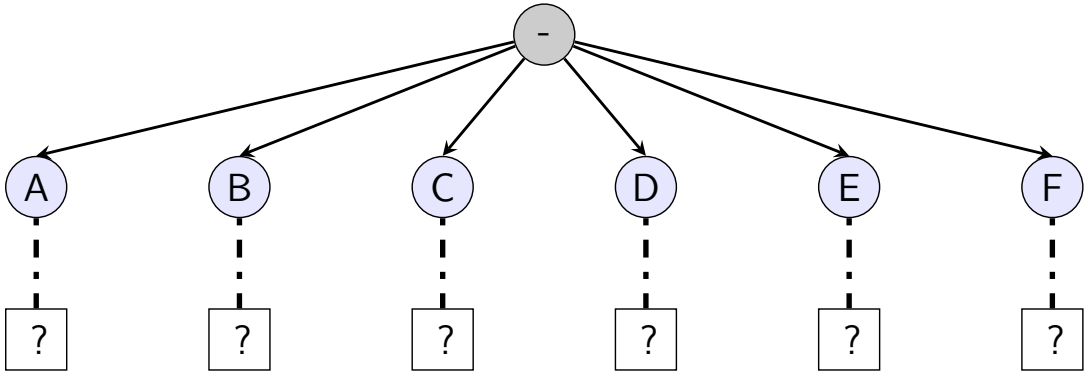
Deuxième partie II

Enjeux

1. *descent* phase



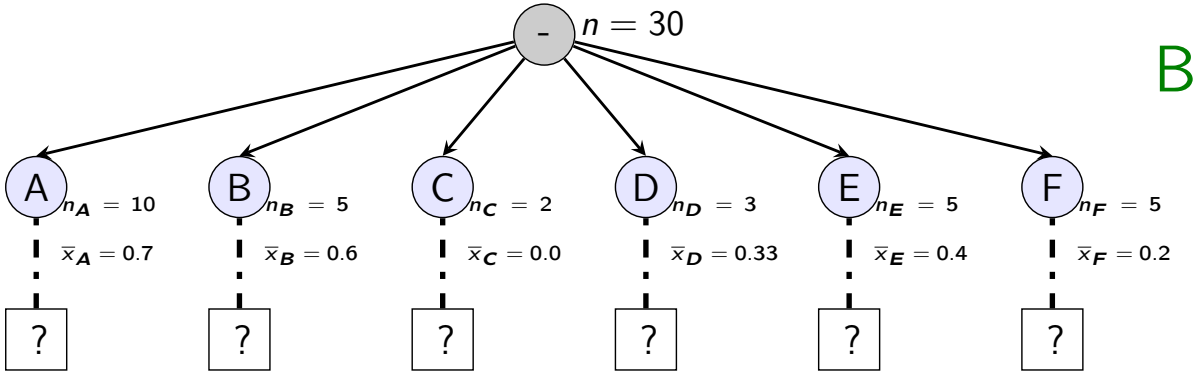
Exploration VS Exploitation



Bandit à *n*-bras

- *n* actions
- récompense inconnue
- minimiser le regret

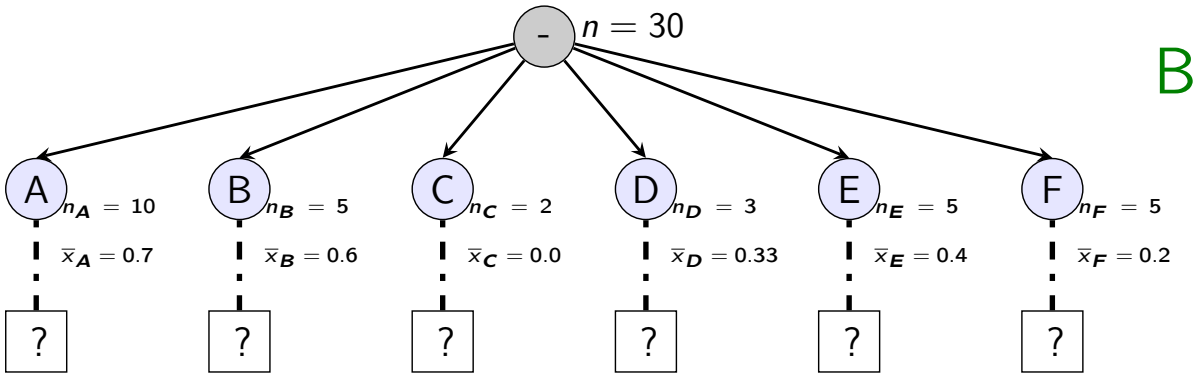
Exploration VS Exploitation



Bandit à *n*-bras

- *n* actions
- récompense inconnue
- minimiser le regret

Exploration VS Exploitation



Bandit à n-bras

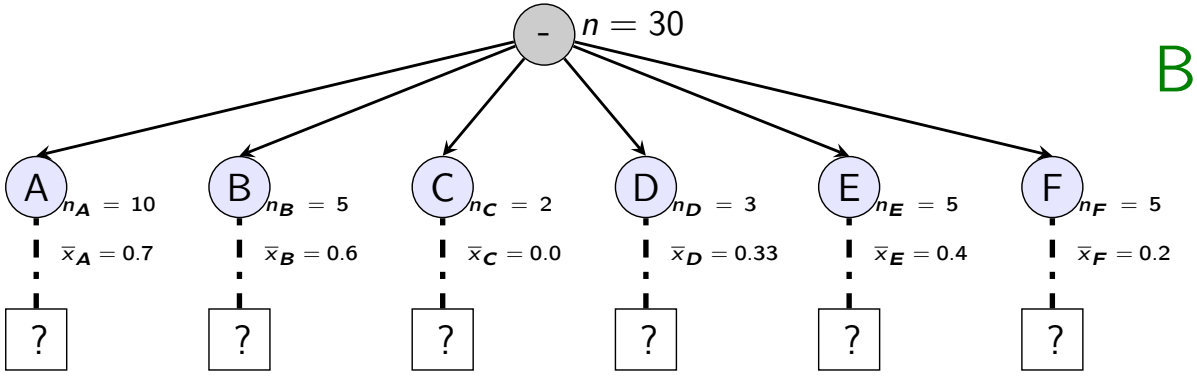
- n actions
- récompense inconnue
- minimiser le regret

Upper Confidence bound

$$\mathrm{UCB}_i = \overline{x}_i + c \times \sqrt{\frac{\ln(n)}{n_i}}$$

c : cœfficient d'exploration

Exploration VS Exploitation



Bandit à *n*-bras

- *n* actions
- récompense inconnue
- minimiser le regret

Upper Confidence bound applied to Trees

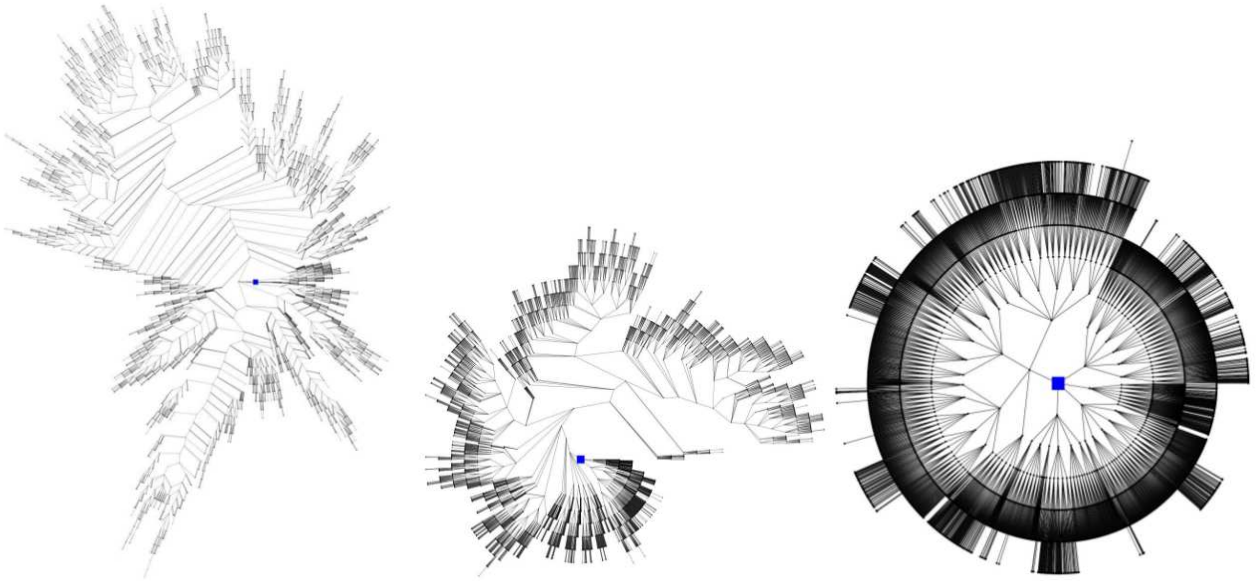
$$\mathrm{UCB}_i = \overline{x}_i + c \times \sqrt{\frac{\ln(n)}{n_i}} \qquad c : \text{cœfficient d'exploration}$$

UCT = chaque nœud est un bandit à *n*-bras *[Kocsis06]*

Construction de l'arbre

Faible c : confiance aveugle *arbre asymétrique*

Fort c : circonspection *arbre symétrique (type MinMax)*



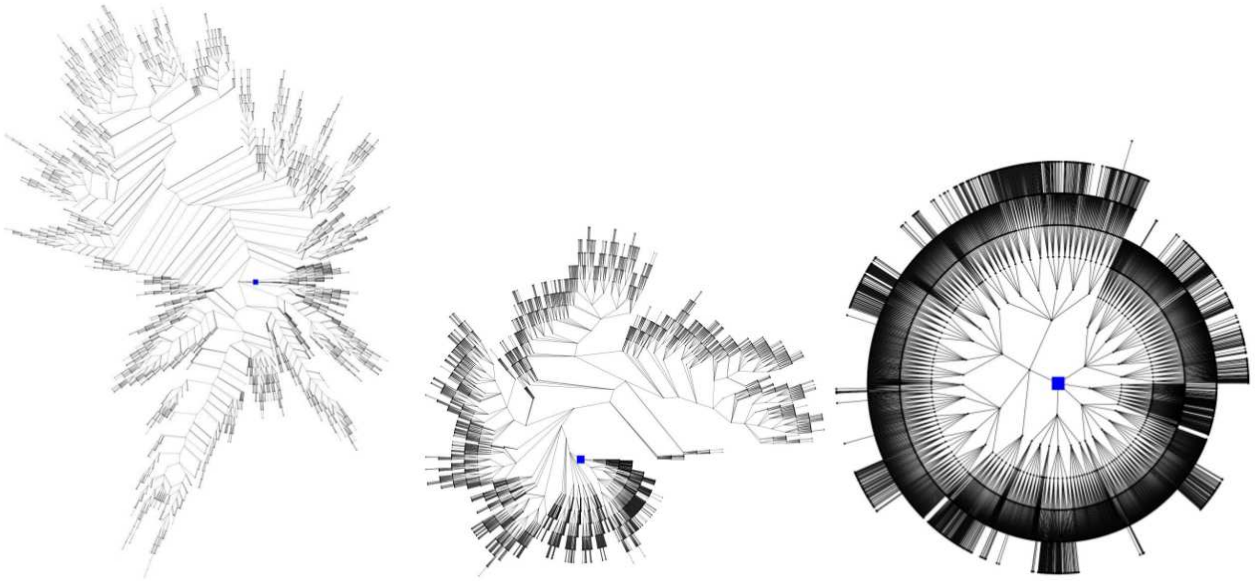
(a) $c = 0$ (b) $c = 2.5$ (c) $c = 20$

Figure: Arbre de recherche UCT avec c croissant *[Ramanujan13]*

Construction de l'arbre

Faible c : confiance aveugle *arbre asymétrique*

Fort c : circonspection *arbre symétrique (type MinMax)*

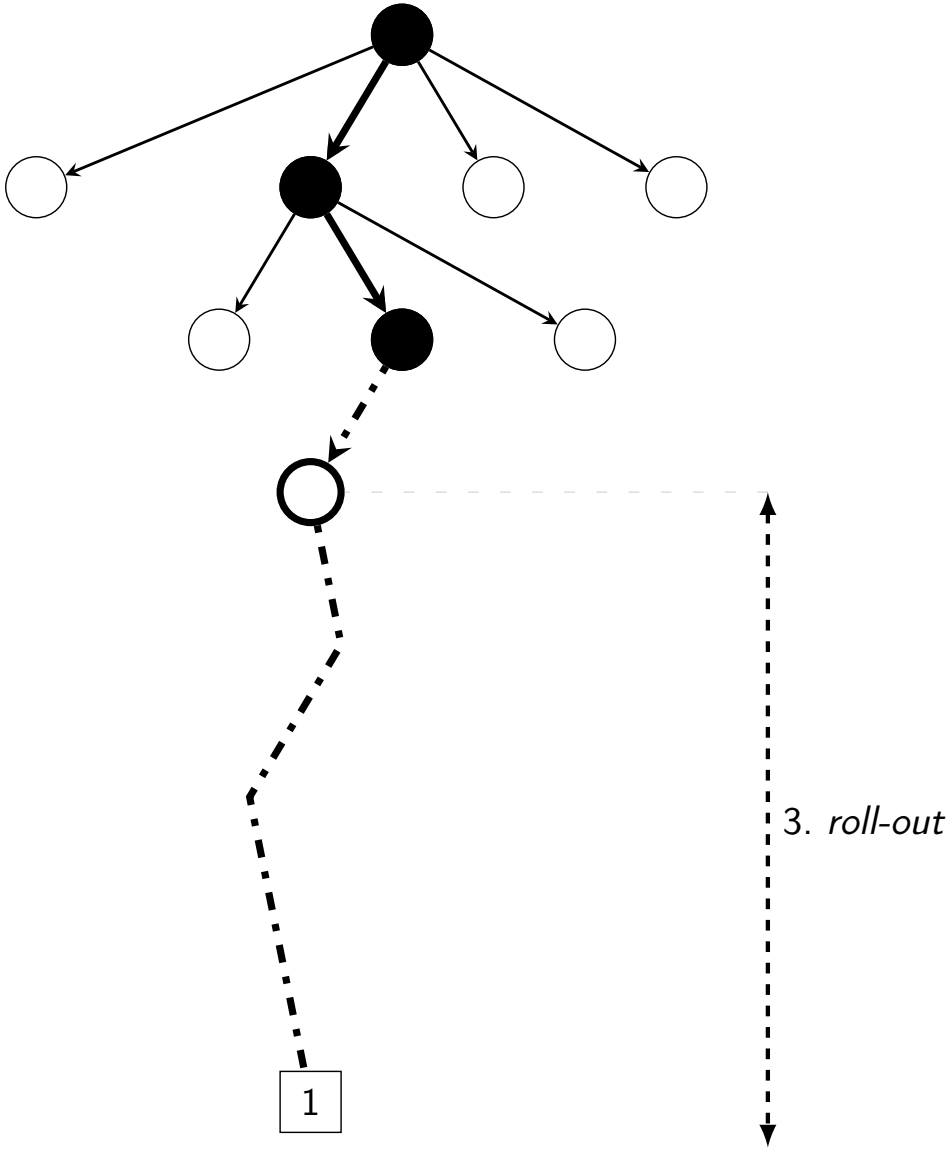


(a) $c = 0$ (b) $c = 2.5$ (c) $c = 20$

Figure: Arbre de recherche UCT avec c croissant *[Ramanujan13]*

L'estimateur *(UCB,RAVE,...)* détermine la structure de l'arbre

3. *roll-out* phase



Dangers de Monte Carlo

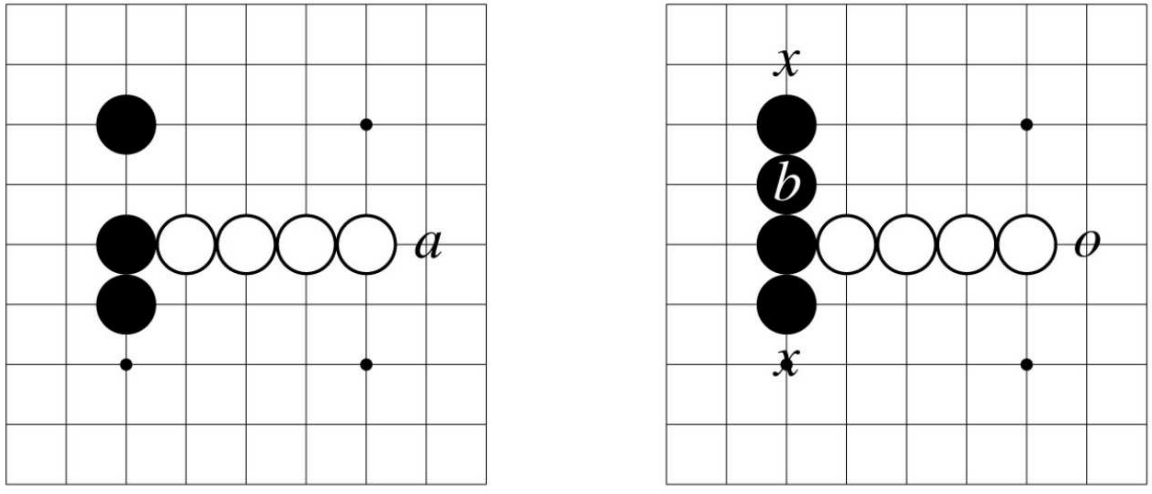


Figure: Situation critique pour Monte Carlo *[Browne11]*

Dangers de Monte Carlo

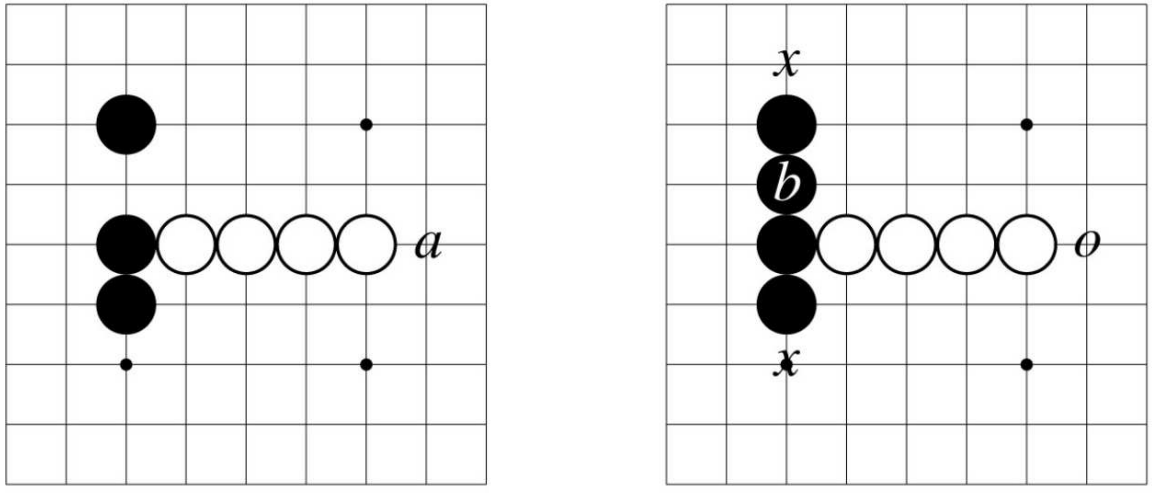


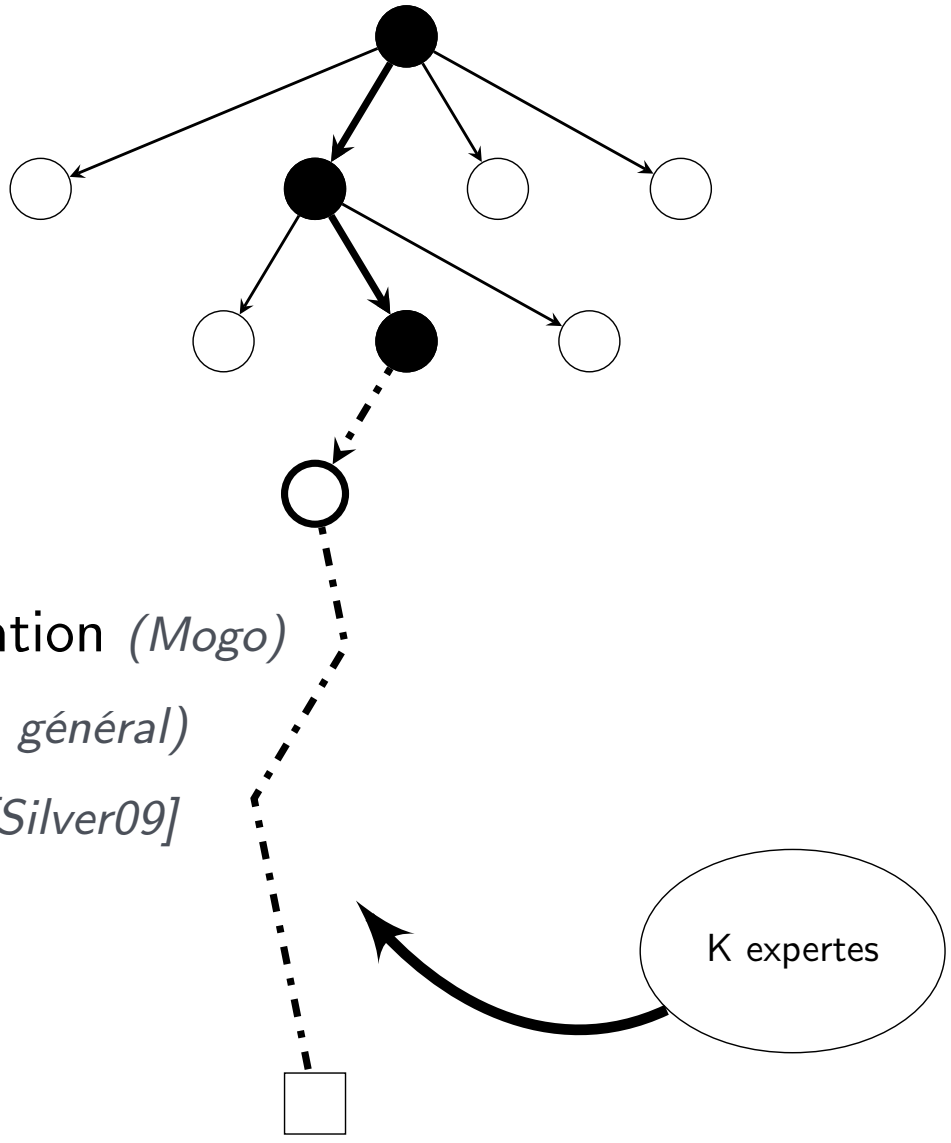
Figure: Situation critique pour Monte Carlo *[Browne11]*

Myopie des simulations aléatoires

- coups « inutiles » *(invraisemblables)*
- situations pièges non couvertes par l'arbre *(Échecs/Go)*

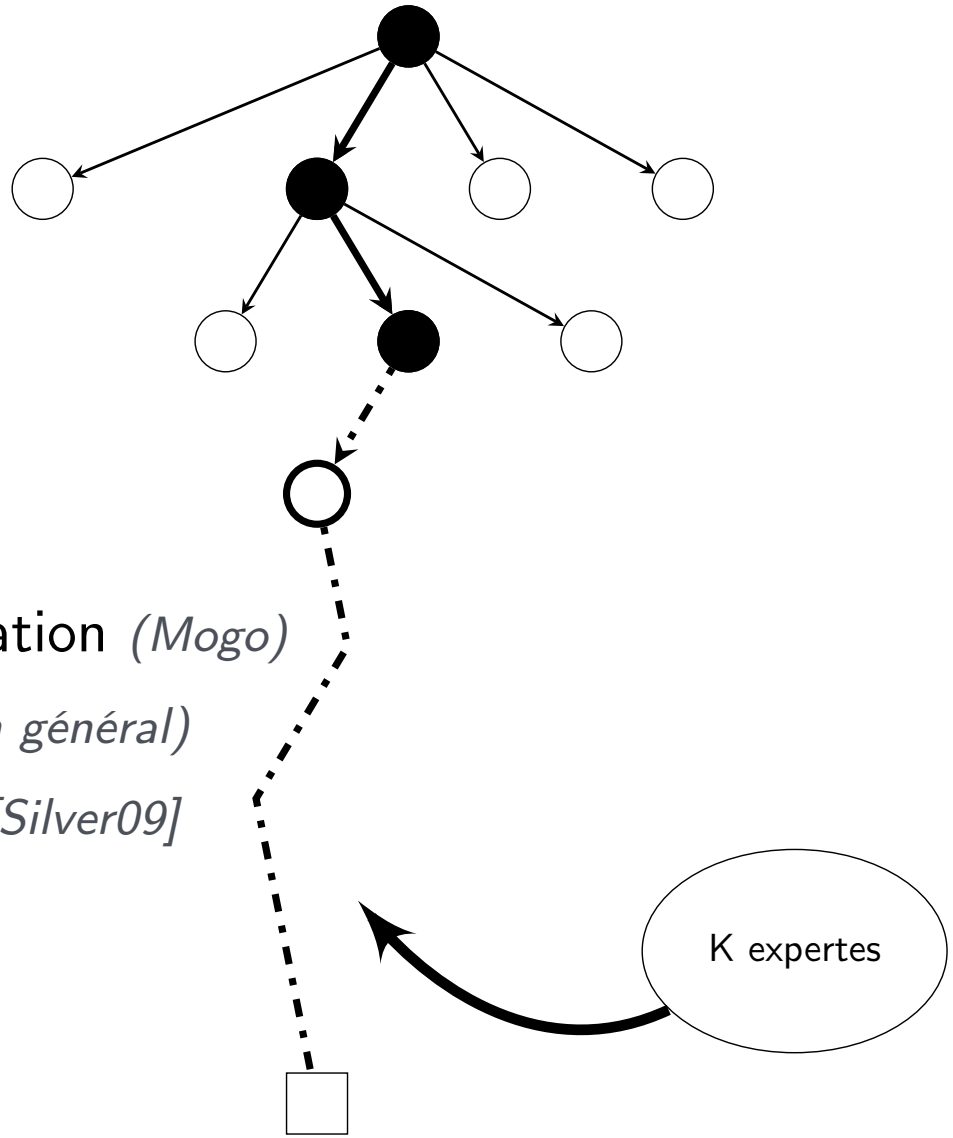
Contrôle de l'aléatoire

- Meilleure représentation *(Mogo)*
- K spécifiques *(... en général)*
- « Magic formula » *[Silver09]*



Contrôle de l'aléatoire

Couplage complexe entre simulations & arbre de recherche



- Meilleure représentation *(Mogo)*
- K spécifiques *(. . . en général)*
- « Magic formula » *[Silver09]*

État d'avancement de la compréhension

Forces

- $\emptyset$ connaissances
- « *anytime* »
- robuste aux bruits

État d'avancement de la compréhension

Forces

- ∅ connaissances
- « *anytime* »
- robuste aux bruits

Faiblesses

- ∅ assimilation *(perte de K)*
- situations pièges *(changement de phase)*
- couplages mal-compris
  (peu de méthodes d'évaluation)

État d'avancement de la compréhension

Forces

- ∅ connaissances
- « *anytime* »
- robuste aux bruits

Faiblesses

- ∅ assimilation *(perte de K)*
- situations pièges *(changement de phase)*
- couplages mal-compris
  (peu de méthodes d'évaluation)

Applications

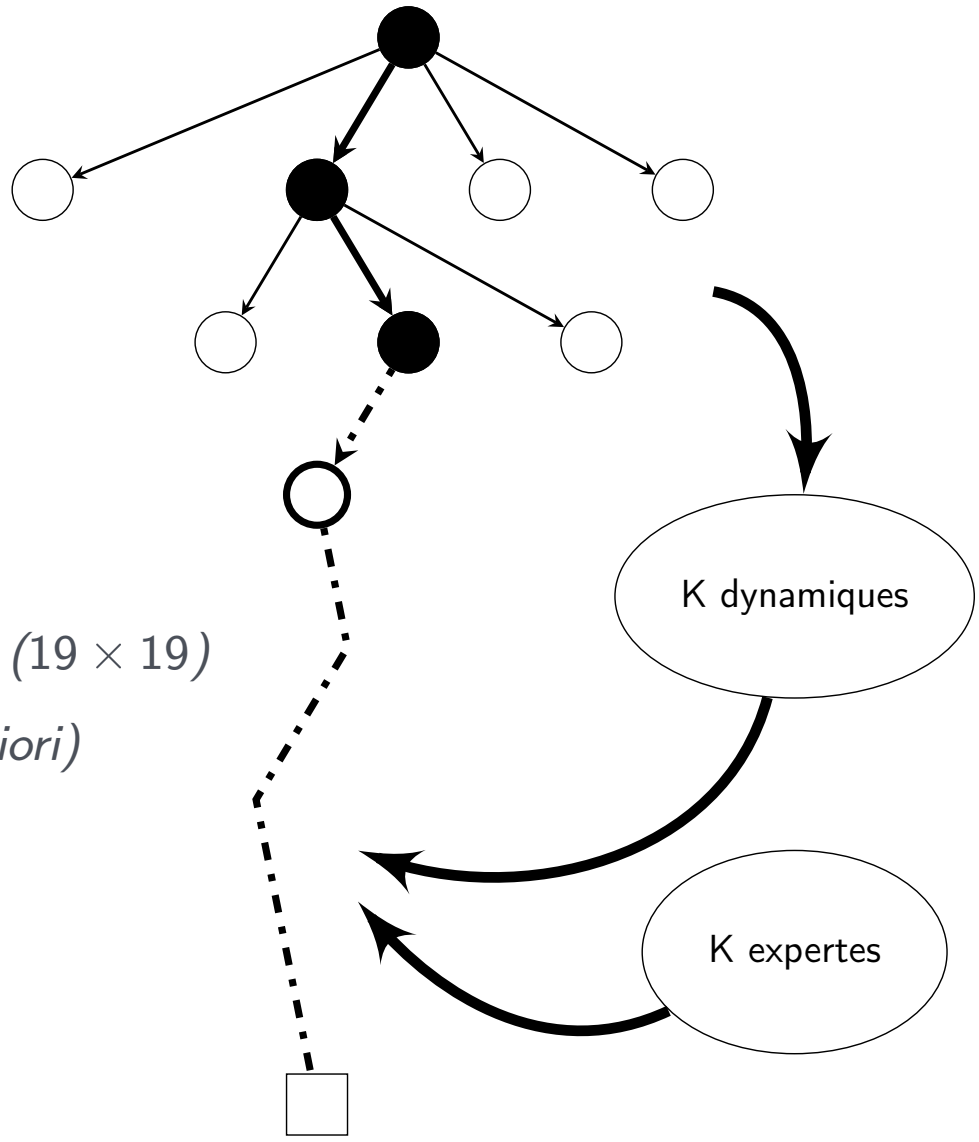
- Jeux combinatoires *(Go,Hex)*
- Jeux non-déterministes *(Poker,Backgammon)*
- Optimisation combinatoire *(TSP, approximation de fonction)*
- Planfication
- ...

Troisième partie III

. . . et affinités

Contrôle par des connaissances dynamiques

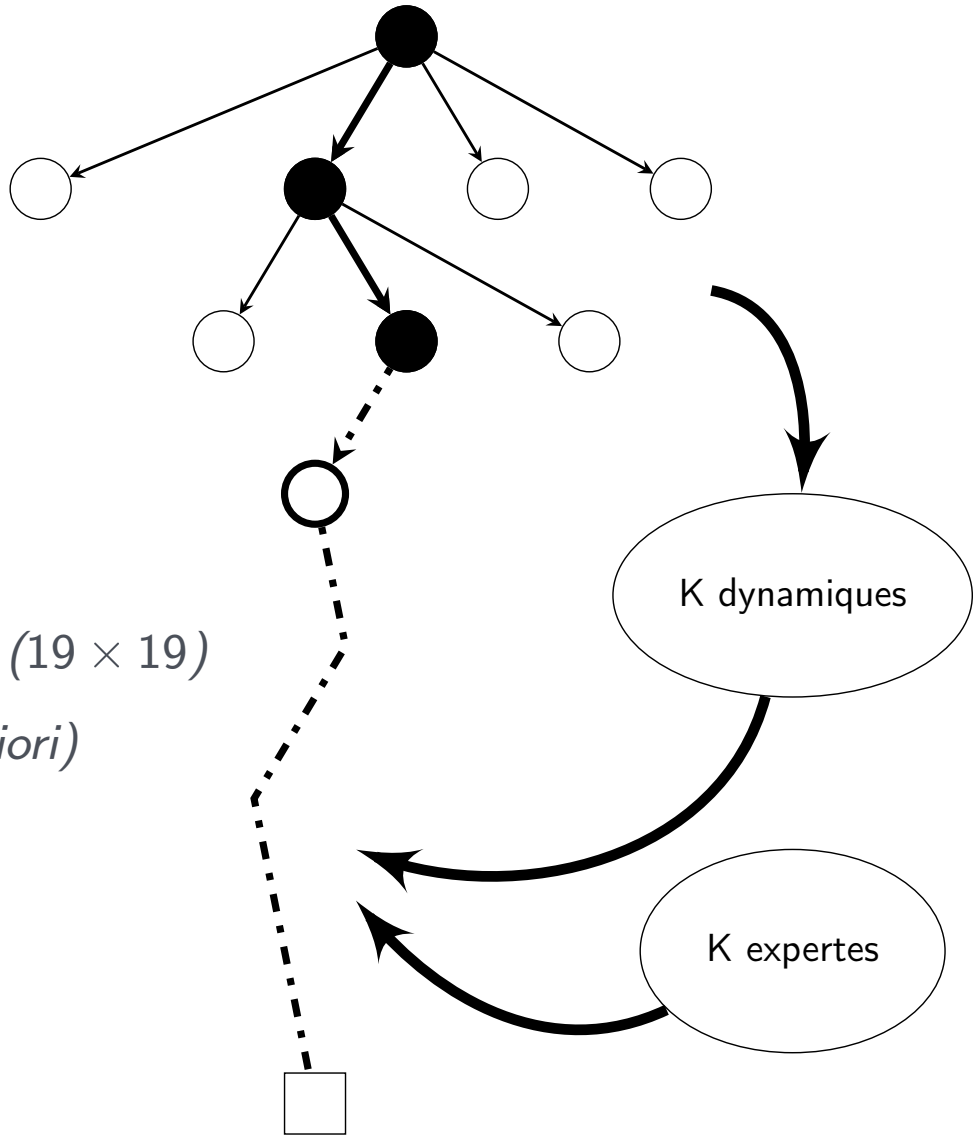
- K complémentaires *(19 × 19)*
- K génériques *(∅ apriori)*
- accumulation



Contrôle par des connaissances dynamiques

Intérêt de réexploiter K de l'arbre pour les simulations

- K complémentaires *(19 × 19)*
- K génériques *(∅ apriori)*
- accumulation



Représentation complémentaire émergente

Réprésentation plus abstraite au service de l'arbre