

# Décision dans l'incertain

Cours 6: Réseaux Bayésiens – Inférence par MC

Stéphane Airiau

Université Paris-Dauphine

## Approximation

---

Répondre à une requête peut nécessiter **beaucoup** de calculs dans le pire des cas, le temps de calcul est une fonction exponentielle de la taille du réseau bayésien.

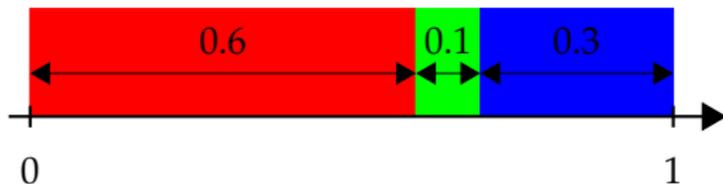
On peut donc approximer en utilisant le réseau bayésien et en tirant des échantillons.

- idée : on répète ces tirages pour obtenir des statistiques
  - comme si on répète la même simulation pour tirer des statistiques.
- tire  $n$  échantillons venant d'une distribution
- on calcule une approximation de la probabilité désirée (fréquence/probabilité empirique)
- on doit montrer que ce processus converge vers la vraie probabilité

## Tirer un échantillon

Supposons qu'on veuille tirer un échantillon de la distribution suivante :

<i>Couleur</i>	$\mathbb{P}(\text{Couleur})$
rouge	0.6
vert	0.1
bleu	0.3



- On tire au sort un réel  $x \in [0, 1)$ 
  - $x \in [0, 0.6)$ , on considère *rouge*
  - $x \in [0.6, 0.7)$ , on considère *vert*
  - $x \in [0.7, 1.0)$ , on considère *bleu*

```
for i in range(10):  
    rnd.random()
```

```
0.3861795824931933  
0.05005345408021611  
0.6190834275288313  
0.30860855431839196  
0.46107245349717174  
0.8640507390726188  
0.008388200283245939  
0.8709112300151461  
0.6171766043667846  
0.5173569165302764
```

## Tirer un échantillon avec un réseau bayésien

---

On va voir 4 méthodes :

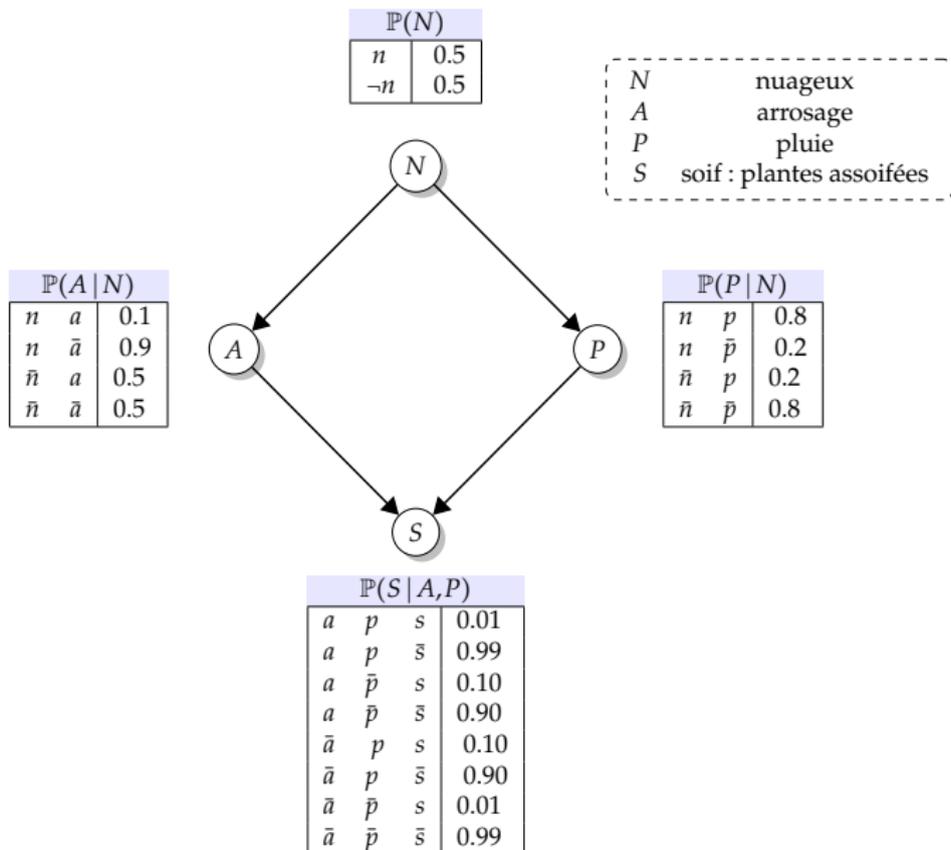
- Echantillonnage simple
- Méthode de rejet
- Correction avec les poids
- Echantillonnage de Gibbs

## "Prior Sampling"

---

- on va faire un parcours de tous les noeuds du réseau bayésien en partant de la racine jusqu'aux feuilles (c'est un graphe orienté acyclique !)
- on va choisir une valeur pour chaque noeud du réseau
  - on sélectionne la distribution  $\mathbb{P}$  par rapport aux valeurs des noeuds parents
  - on tire un échantillon de cette distribution  $\mathbb{P}$  qui fixe la valeur du noeud

# "Prior Sampling" – exemple



# "Prior Sampling" – exemple

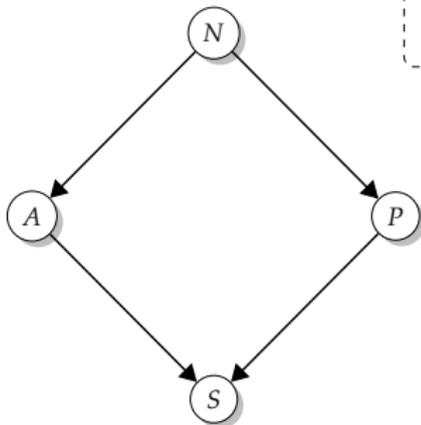
tirage au sort  $\mathbb{P}(N)$   
 (0.5,0.5)  
 résultat :  $n$

$n$	0.5
$-n$	0.5

$N$	nuageux
$A$	arrosage
$P$	pluie
$S$	soif : plantes assoiffées

$\mathbb{P}(A | N)$

$n$	$a$	0.1
$n$	$\bar{a}$	0.9
$\bar{n}$	$a$	0.5
$\bar{n}$	$\bar{a}$	0.5



$\mathbb{P}(P | N)$

$n$	$p$	0.8
$n$	$\bar{p}$	0.2
$\bar{n}$	$p$	0.2
$\bar{n}$	$\bar{p}$	0.8

$\mathbb{P}(S | A, P)$

$a$	$p$	$s$	0.01
$a$	$p$	$\bar{s}$	0.99
$a$	$\bar{p}$	$s$	0.10
$a$	$\bar{p}$	$\bar{s}$	0.90
$\bar{a}$	$p$	$s$	0.10
$\bar{a}$	$p$	$\bar{s}$	0.90
$\bar{a}$	$\bar{p}$	$s$	0.01
$\bar{a}$	$\bar{p}$	$\bar{s}$	0.99

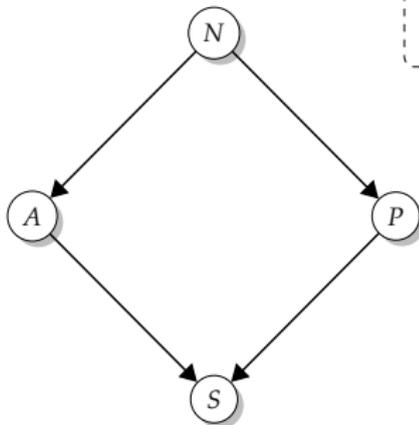
# "Prior Sampling" – exemple

$\mathbb{P}(N)$	
$n$	0.5
$\bar{n}$	0.5

$N$	nuageux
$A$	arrosage
$P$	pluie
$S$	soif : plantes assoiffées

$\mathbb{P}(A N)$		
$n$	$a$	0.1
$n$	$\bar{a}$	0.9
$\bar{n}$	$a$	0.5
$\bar{n}$	$\bar{a}$	0.5

tirage au sort  
(0.1,0.9)  
résultat :  $\bar{a}$



$\mathbb{P}(P N)$		
$n$	$p$	0.8
$n$	$\bar{p}$	0.2
$\bar{n}$	$p$	0.2
$\bar{n}$	$\bar{p}$	0.8

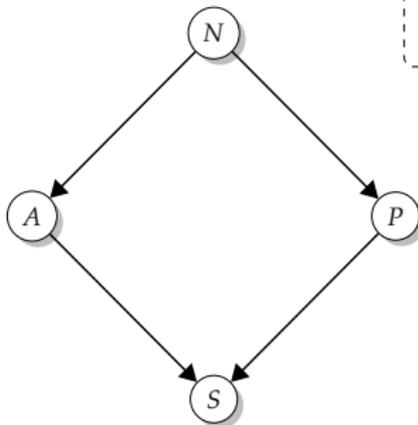
$\mathbb{P}(S A,P)$			
$a$	$p$	$s$	0.01
$a$	$p$	$\bar{s}$	0.99
$a$	$\bar{p}$	$s$	0.10
$a$	$\bar{p}$	$\bar{s}$	0.90
$\bar{a}$	$p$	$s$	0.10
$\bar{a}$	$p$	$\bar{s}$	0.90
$\bar{a}$	$\bar{p}$	$s$	0.01
$\bar{a}$	$\bar{p}$	$\bar{s}$	0.99

# "Prior Sampling" – exemple

$\mathbb{P}(N)$	
$n$	0.5
$-n$	0.5

$N$	nuageux
$A$	arrosage
$P$	pluie
$S$	soif : plantes assoiffées

$\mathbb{P}(A   N)$		
$n$	$a$	0.1
$n$	$\bar{a}$	0.9
$\bar{n}$	$a$	0.5
$\bar{n}$	$\bar{a}$	0.5

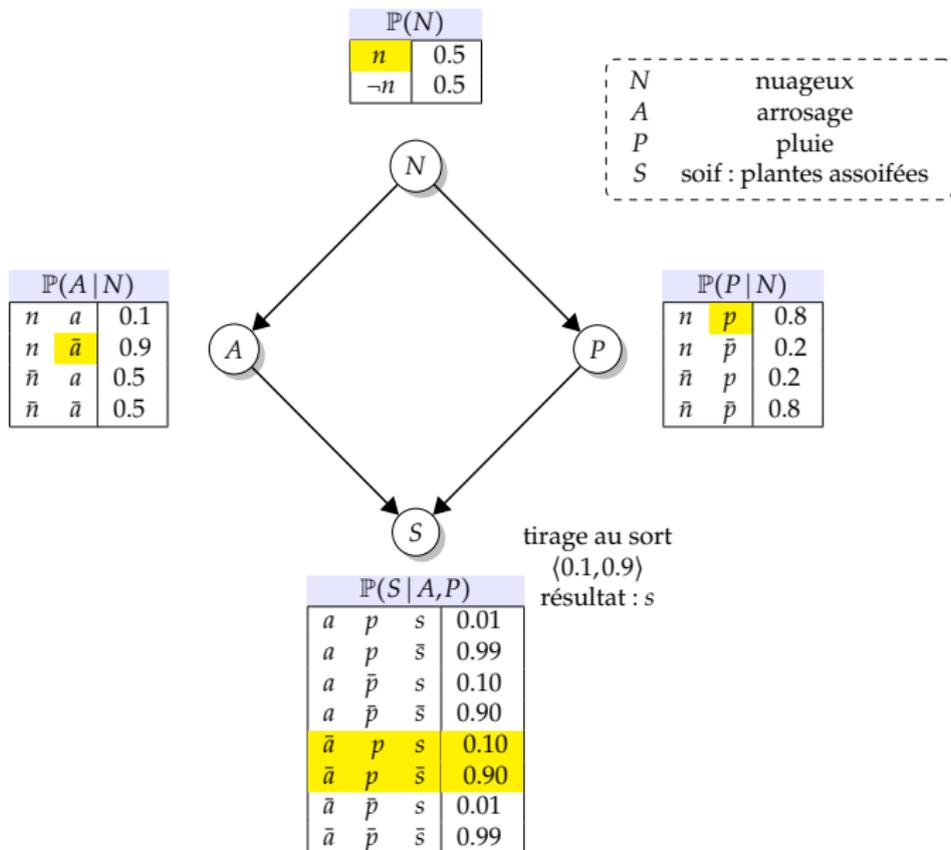


$\mathbb{P}(P   N)$		
$n$	$p$	0.8
$n$	$\bar{p}$	0.2
$\bar{n}$	$p$	0.2
$\bar{n}$	$\bar{p}$	0.8

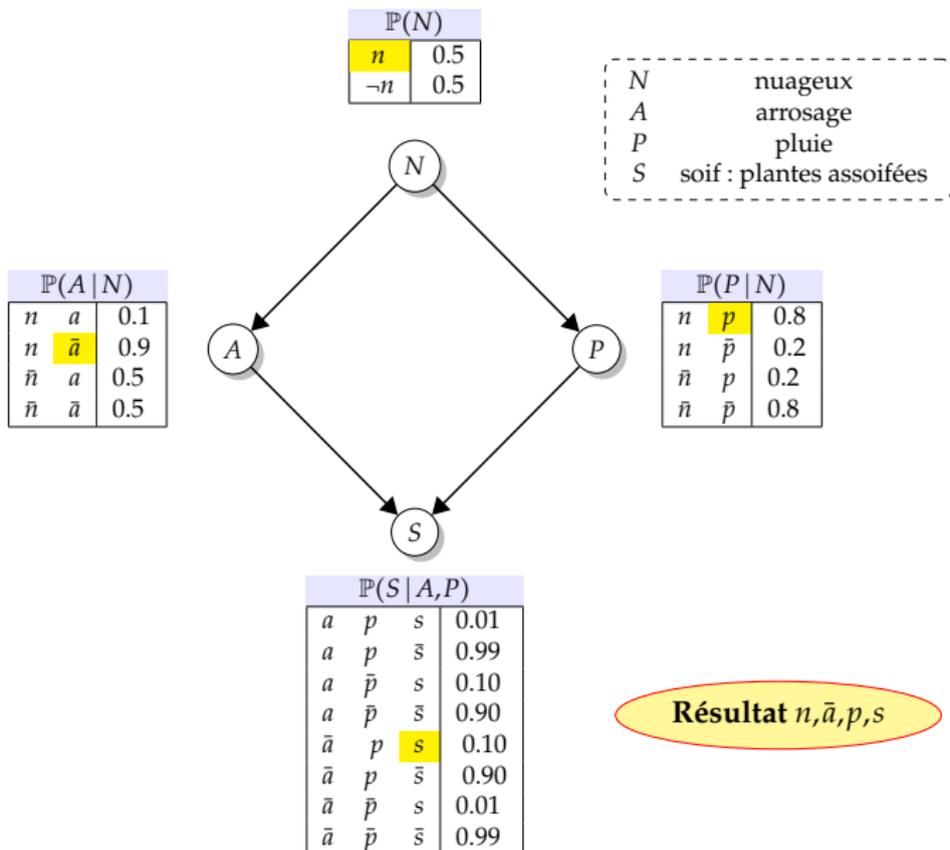
tirage au sort  
(0.8,0.2)  
résultat :  $p$

$\mathbb{P}(S   A, P)$			
$a$	$p$	$s$	0.01
$a$	$p$	$\bar{s}$	0.99
$a$	$\bar{p}$	$s$	0.10
$a$	$\bar{p}$	$\bar{s}$	0.90
$\bar{a}$	$p$	$s$	0.10
$\bar{a}$	$p$	$\bar{s}$	0.90
$\bar{a}$	$\bar{p}$	$s$	0.01
$\bar{a}$	$\bar{p}$	$\bar{s}$	0.99

# "Prior Sampling" – exemple



# "Prior Sampling" – exemple



## "Prior Sampling"

---

- ce processus génère des échantillons avec la probabilité du réseau bayésien !
- Soit  $n_{(x_1, \dots, x_n)}$  le nombre d'échantillons de l'évènement  $(x_1, \dots, x_n)$ .
- avec la loi des grands nombre

$$\begin{aligned}\lim_{n \rightarrow \infty} \hat{\mathbb{P}}(x_1, \dots, x_n) &= \lim_{n \rightarrow \infty} \frac{n_{(x_1, \dots, x_n)}}{n} \\ &= \mathbb{P}(x_1, \dots, x_n)\end{aligned}$$

⇒ la procédure est cohérente

## "Prior Sampling" – exemple

---

Supposons qu'on ait  $n = 5$  échantillons :

$$n, \bar{a}, p, \bar{s}$$

$$n, a, p, \bar{s}$$

$$\bar{n}, a, p, s$$

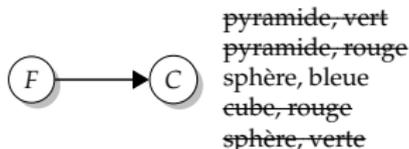
$$n, \bar{a}, p, \bar{s}$$

$$\bar{n}, \bar{a}, \bar{p}, \bar{s}$$

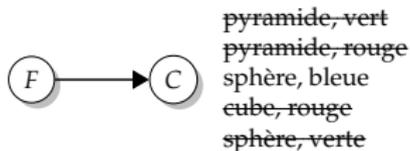
- requête :  $\mathbb{P}(S) : 4 \bar{s} 1 s \Rightarrow$  on normalise  $\hat{\mathbb{P}}(S) = \langle 0.8, 0.2 \rangle$
- avec plus d'échantillon, on va de mieux en mieux approcher  $\mathbb{P}(S)$
- peut-on dire quelque chose de
  - $\mathbb{P}(N | \bar{s})$  ?
  - $\mathbb{P}(N | p, \bar{s})$  ?
  - $\mathbb{P}(N | \bar{p}, s)$  ?
- moins de temps ?  $\Rightarrow$  moins d'échantillons  $\Rightarrow$  moins de précision !

- supposons qu'on soit simplement intéressé à  $\mathbb{P}(N)$
- pas besoin de conserver toute l'information, on peut maintenir le compte pour les valeurs de  $N$  seulement.
- requête  $\mathbb{P}(N | \bar{s})$
- même idée, on compte les valeurs de  $N$ , mais en ignorant tout échantillon qui ne vérifie pas  $S = \bar{s}$
- ➡ méthode de rejet
- c'est aussi cohérent !
- L'écart type de l'erreur de chaque probabilité est proportionnelle à  $\frac{1}{\sqrt{n}}$  où  $n$  est le nombre d'échantillons.

- si les observations sont peu probables  $\Rightarrow$  rejets de beaucoup d'échantillons !
- certaines observations ne sont pas exploitées durant l'élaboration de l'échantillon complet
- requête  $\mathbb{P}(\text{sphère} \mid \text{bleue})$

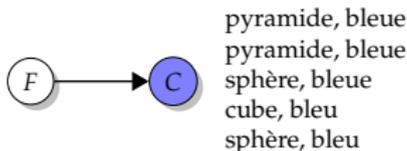


- si les observations sont peu probables  $\Rightarrow$  rejets de beaucoup d'échantillons !
- certaines observations ne sont pas exploitées durant l'élaboration de l'échantillon complet
- requête  $\mathbb{P}(\text{sphère} \mid \text{bleue})$



Idée : fixe les variables observées, et on tire des échantillons sur le reste !

- problème si on ne fait rien, on risque de perdre la cohérence !
- solution : on ré-équilibre en mettant des poids sur la probabilités des observations selon la valeur des parents.



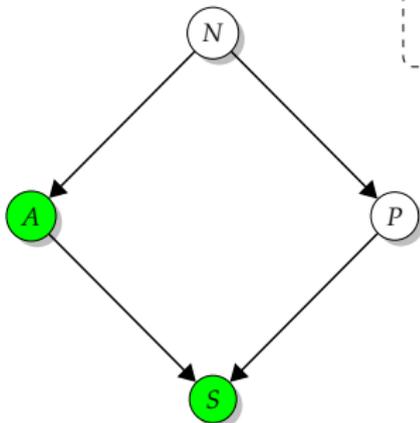
# "Likelihood weighting" – exemple

on observe  $a, \bar{s}$

$\mathbb{P}(N)$	
$n$	0.5
$\bar{n}$	0.5

$N$  nuageux  
 $A$  arrosage  
 $P$  pluie  
 $S$  soif : plantes assoiffées

$\mathbb{P}(A   N)$		
$n$	$a$	0.1
$n$	$\bar{a}$	0.9
$\bar{n}$	$a$	0.5
$\bar{n}$	$\bar{a}$	0.5



$\mathbb{P}(P   N)$		
$n$	$p$	0.8
$n$	$\bar{p}$	0.2
$\bar{n}$	$p$	0.2
$\bar{n}$	$\bar{p}$	0.8

$\mathbb{P}(S   A, P)$			
$a$	$p$	$s$	0.01
$a$	$p$	$\bar{s}$	0.99
$a$	$\bar{p}$	$s$	0.10
$a$	$\bar{p}$	$\bar{s}$	0.90
$\bar{a}$	$p$	$s$	0.10
$\bar{a}$	$p$	$\bar{s}$	0.90
$\bar{a}$	$\bar{p}$	$s$	0.01
$\bar{a}$	$\bar{p}$	$\bar{s}$	0.99

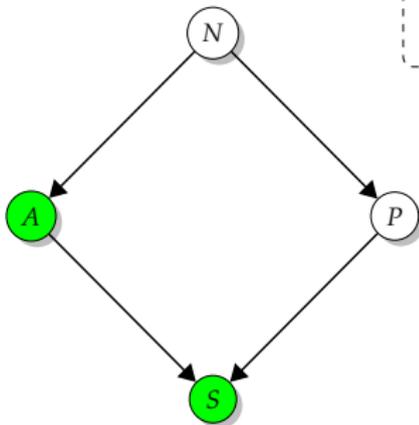
# "Likelihood weighting" – exemple

tirage au sort  
on observe  $a, \bar{s} (0.5, 0.5)$   
résultat :  $n$

$\mathbb{P}(N)$	
$n$	0.5
$\bar{n}$	0.5

$N$  nuageux  
 $A$  arrosage  
 $P$  pluie  
 $S$  soif : plantes assoiffées

$\mathbb{P}(A   N)$		
$n$	$a$	0.1
$n$	$\bar{a}$	0.9
$\bar{n}$	$a$	0.5
$\bar{n}$	$\bar{a}$	0.5



$\mathbb{P}(P   N)$		
$n$	$p$	0.8
$n$	$\bar{p}$	0.2
$\bar{n}$	$p$	0.2
$\bar{n}$	$\bar{p}$	0.8

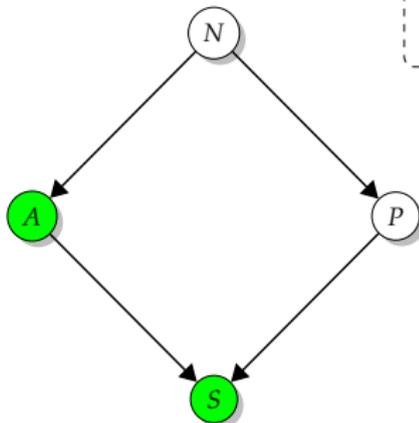
$\mathbb{P}(S   A, P)$			
$a$	$p$	$s$	0.01
$a$	$p$	$\bar{s}$	0.99
$a$	$\bar{p}$	$s$	0.10
$a$	$\bar{p}$	$\bar{s}$	0.90
$\bar{a}$	$p$	$s$	0.10
$\bar{a}$	$p$	$\bar{s}$	0.90
$\bar{a}$	$\bar{p}$	$s$	0.01
$\bar{a}$	$\bar{p}$	$\bar{s}$	0.99

# "Likelihood weighting" – exemple

on observe  $a, \bar{s}$

$\mathbb{P}(N)$	
$n$	0.5
$\bar{n}$	0.5

$N$  nuageux  
 $A$  arrosage  
 $P$  pluie  
 $S$  soif : plantes assoiffées



$\mathbb{P}(A N)$		
$n$	$a$	0.1
$n$	$\bar{a}$	0.9
$\bar{n}$	$a$	0.5
$\bar{n}$	$\bar{a}$	0.5

$\mathbb{P}(P N)$		
$n$	$p$	0.8
$n$	$\bar{p}$	0.2
$\bar{n}$	$p$	0.2
$\bar{n}$	$\bar{p}$	0.8

tirage au sort  
 $\langle 0.8, 0.2 \rangle$   
 résultat :  $p$

$\mathbb{P}(S A,P)$			
$a$	$p$	$s$	0.01
$a$	$p$	$\bar{s}$	0.99
$a$	$\bar{p}$	$s$	0.10
$a$	$\bar{p}$	$\bar{s}$	0.90
$\bar{a}$	$p$	$s$	0.10
$\bar{a}$	$p$	$\bar{s}$	0.90
$\bar{a}$	$\bar{p}$	$s$	0.01
$\bar{a}$	$\bar{p}$	$\bar{s}$	0.99

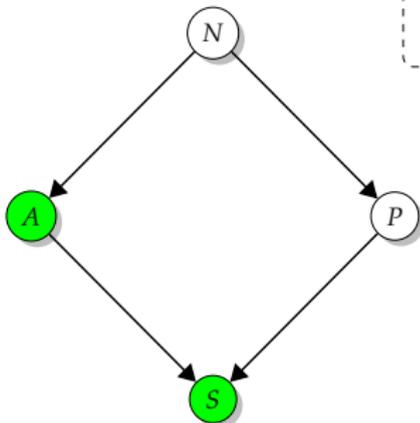
# "Likelihood weighting" – exemple

on observe  $a, \bar{s}$

$\mathbb{P}(N)$	
$n$	0.5
$\bar{n}$	0.5

$N$  nuageux  
 $A$  arrosage  
 $P$  pluie  
 $S$  soif : plantes assoiffées

$\mathbb{P}(A   N)$		
$n$	$a$	0.1
$n$	$\bar{a}$	0.9
$\bar{n}$	$a$	0.5
$\bar{n}$	$\bar{a}$	0.5



$\mathbb{P}(P   N)$		
$n$	$p$	0.8
$n$	$\bar{p}$	0.2
$\bar{n}$	$p$	0.2
$\bar{n}$	$\bar{p}$	0.8

$\mathbb{P}(S   A, P)$			
$a$	$p$	$s$	0.01
$a$	$p$	$\bar{s}$	0.99
$a$	$\bar{p}$	$s$	0.10
$a$	$\bar{p}$	$\bar{s}$	0.90
$\bar{a}$	$p$	$s$	0.10
$\bar{a}$	$p$	$\bar{s}$	0.90
$\bar{a}$	$\bar{p}$	$s$	0.01
$\bar{a}$	$\bar{p}$	$\bar{s}$	0.99

**Résultat  $n, a, p, s$**   
**Poids  $w = 0.1 \times 0.99$**

## "Likelihood weighting" – algorithme

---

```
1  instancie les observations
2   $w = 1.0$ 
3  for  $i \in \{1, 2, \dots, n\}$ 
4      if  $X_i$  est une variable observée
5           $X_i = x_i$  fixée
6           $w \leftarrow w \cdot \mathbb{P}(x_i | Parents(X_i))$ 
7      else
8          tirer un échantillon de  $\mathbb{P}(X_i | Parents(X_i))$ 
9  return  $(x_1, x_2, \dots, x_n), w$ 
```

On peut démontrer que le processus est aussi cohérent !

## "Likelihood weighting" – pour et contre

---

- on prend en compte les observations lorsqu'on tire nos échantillons
- le poids  $w$  est bien calculé en fonction de l'échantillon tiré
- nos échantillons seront tous utilisables !

### Mais il reste des problèmes

- si les observations sont loin dans le réseau,
- les observations influencent le choix de variables placées "dessous"
- ➡ on voudrait avoir des observation pour n'importe quelle variable

- approche vraiment différente :  
on part d'une "monde possible" puis on va le perturber de manière répétée.
- fait partie de la famille dite de Monte-Carlo par chaînes de Markov

Procédure :

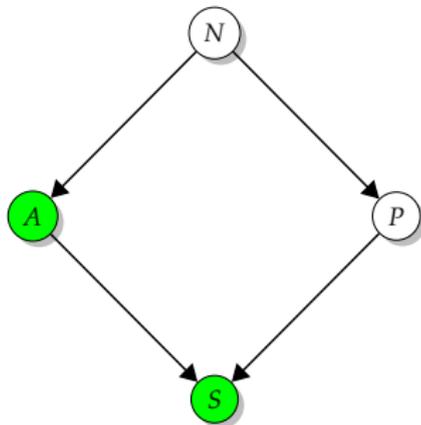
- On part d'une affectation complète et arbitraire  $x_1, \dots, x_n$  qui respecte les observations
- On répète :
  - on tire au hasard une variable non observée
  - on tire un échantillon de cette variable étant donnée la valeur courante des variables dans la couverture de Markov
  - on maintient un compte
- on normalise le décompte.

# Echantillonnage de Gibbs – exemple

on observe  $a, \bar{s}$

$\mathbb{P}(N)$	
$n$	0.5
$\bar{n}$	0.5

$N$  nuageux  
 $A$  arrosage  
 $P$  pluie  
 $S$  soif : plantes assoiffées



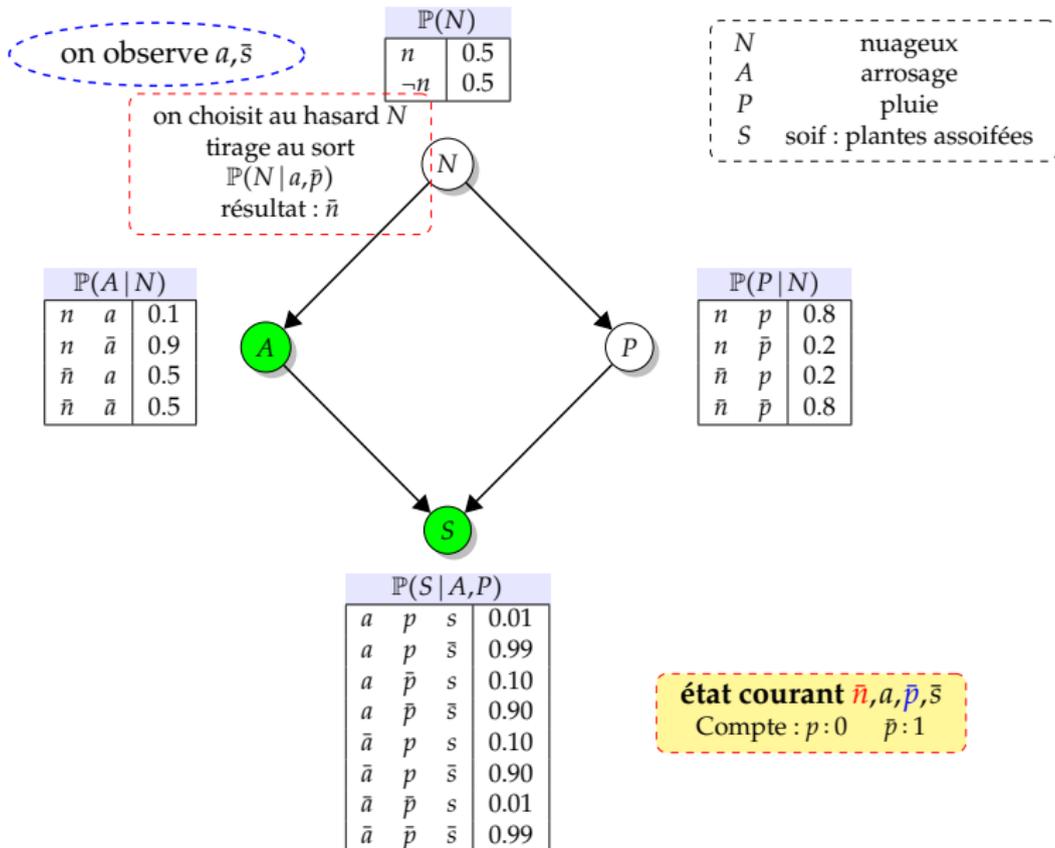
$\mathbb{P}(A N)$		
$n$	$a$	0.1
$n$	$\bar{a}$	0.9
$\bar{n}$	$a$	0.5
$\bar{n}$	$\bar{a}$	0.5

$\mathbb{P}(P N)$		
$n$	$p$	0.8
$n$	$\bar{p}$	0.2
$\bar{n}$	$p$	0.2
$\bar{n}$	$\bar{p}$	0.8

$\mathbb{P}(S A,P)$			
$a$	$p$	$s$	0.01
$a$	$p$	$\bar{s}$	0.99
$a$	$\bar{p}$	$s$	0.10
$a$	$\bar{p}$	$\bar{s}$	0.90
$\bar{a}$	$p$	$s$	0.10
$\bar{a}$	$p$	$\bar{s}$	0.90
$\bar{a}$	$\bar{p}$	$s$	0.01
$\bar{a}$	$\bar{p}$	$\bar{s}$	0.99

état initial  $n, a, \bar{p}, \bar{s}$   
 Compte :  $p : 0$   $\bar{p} : 0$

# Echantillonnage de Gibbs – exemple

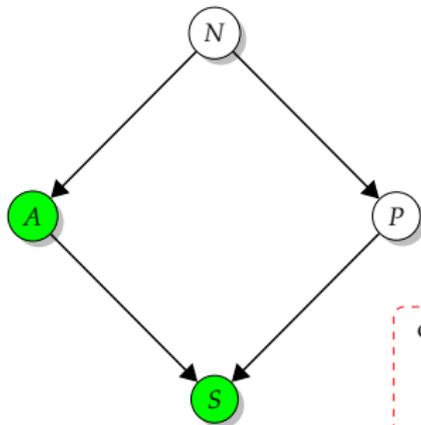


# Echantillonnage de Gibbs – exemple

on observe  $a, \bar{s}$

$\mathbb{P}(N)$	
$n$	0.5
$\bar{n}$	0.5

$N$  nuageux  
 $A$  arrosage  
 $P$  pluie  
 $S$  soif : plantes assoiffées



$\mathbb{P}(A N)$		
$n$	$a$	0.1
$n$	$\bar{a}$	0.9
$\bar{n}$	$a$	0.5
$\bar{n}$	$\bar{a}$	0.5

$\mathbb{P}(P N)$		
$n$	$p$	0.8
$n$	$\bar{p}$	0.2
$\bar{n}$	$p$	0.2
$\bar{n}$	$\bar{p}$	0.8

on choisit au hasard  $P$   
 tirage au sort  
 $\mathbb{P}(P|\bar{n}, a, \bar{s})$   
 résultat :  $p$

$\mathbb{P}(S A,P)$			
$a$	$p$	$s$	0.01
$a$	$p$	$\bar{s}$	0.99
$a$	$\bar{p}$	$s$	0.10
$a$	$\bar{p}$	$\bar{s}$	0.90
$\bar{a}$	$p$	$s$	0.10
$\bar{a}$	$p$	$\bar{s}$	0.90
$\bar{a}$	$\bar{p}$	$s$	0.01
$\bar{a}$	$\bar{p}$	$\bar{s}$	0.99

état courant  $\bar{n}, a, p, \bar{s}$   
 Compte :  $p : 1$   $\bar{p} : 1$

## Echantillonnage de Gibbs - pourquoi cela fonctionne ?

---

idée : la fraction du temps passée dans chaque état est exactement proportionnelle à la probabilité cherchée.

- $\vec{x} = (x_1, \dots, x_n)$  est un monde possible, i.e. une affectation de chaque valeur  $X_i$  à une valeur  $x_i$ .
- $\vec{e}$  sont les observations, i.e. pour les variables observées on a  $X_j = e_j$ .
- $q(\vec{x} \rightarrow \vec{x}')$  est la probabilité que la chaîne passe de l'état  $\vec{x}$  à l'état  $\vec{x}'$ .
- $\pi_t(\vec{x})$  la probabilité que la chaîne soit à l'état  $\vec{x}$  à l'étape  $t \in \mathbb{N}$ .
- on a

$$\pi_{t+1}(\vec{x}) = \sum_{\vec{x}'} \pi_t(\vec{x}') q(\vec{x}' \rightarrow \vec{x}) \text{ pour tout } \vec{x}'.$$

### Definition (Stationarité)

---

La chaîne atteint une probabilité **stationnaire** ssi  $\pi_t(\vec{x}) = \pi_{t+1}(\vec{x})$ .

On note cette probabilité  $\pi$  et elle vérifie l'équation

$$\pi(\vec{x}) = \sum_{\vec{x}'} \pi(\vec{x}') q(\vec{x}' \rightarrow \vec{x})$$

## Echantillonnage de Gibbs - pourquoi cela fonctionne ?

---

Lorsqu'on tire un échantillon de la variable  $X_i$ , on note

- $q_i$  la probabilité de transition
- $x_{-i}$  le vecteur des valeurs de toutes les variables qui ne sont ni  $X_i$ , ni les variables observées.

Lorsqu'on tire un nouvel échantillon  $x'_i$  de  $X_i$  alors

$$q_i(\vec{x} \rightarrow \vec{x}') = q_i((x_i, x_{-i}) \rightarrow (x'_i, x_{-i})) = \mathbb{P}(x'_i | x_{-i}, \vec{e})$$

$$\begin{aligned}\pi(\vec{x})q_i(\vec{x} \rightarrow \vec{x}') &= \mathbb{P}(\vec{x} | e) \cdot \mathbb{P}(x'_i | x_{-i}, \vec{e}) \\ &= \mathbb{P}(x_i, x_{-i} | e) \cdot \mathbb{P}(x'_i | x_{-i}, \vec{e}) \\ &= \mathbb{P}(x_i | x_{-i}, e) \cdot \mathbb{P}(x_{-i} | e) \cdot \mathbb{P}(x'_i | x_{-i}, \vec{e}) \\ &= \mathbb{P}(x_i | x_{-i}, e) \cdot \mathbb{P}(x'_i, x_{-i} | \vec{e}) \\ &= \mathbb{P}(x_i | x_{-i}, e) \cdot \mathbb{P}(\vec{x}' | \vec{e}) \\ &= q_i(\vec{x}' \rightarrow \vec{x}) \cdot \pi(\vec{x}')\end{aligned}$$

## Echantillonnage de Gibbs - pourquoi cela fonctionne ?

---

On a établi que  $\pi(\vec{x})q_i(\vec{x} \rightarrow \vec{x}') = \pi(\vec{x}') \cdot q_i(\vec{x}' \rightarrow \vec{x})$

Sommons sur tous les états  $\vec{x}$

$$\sum_{\vec{x}} \pi(\vec{x})q_i(\vec{x} \rightarrow \vec{x}') = \sum_{\vec{x}} \pi(\vec{x}') \cdot q_i(\vec{x}' \rightarrow \vec{x}) = \pi(\vec{x}') \underbrace{\sum_{\vec{x}} q_i(\vec{x}' \rightarrow \vec{x})}_{=1} = \pi(\vec{x}')$$

Donc  $q_i$  a pour distribution stationnaire  $\pi$ .

Lorsque l'on prend un échantillon tour à tour de chacune des variables non observées, c'est comme si on avait une transition globale  $q = q_{\sigma(1)} \circ q_{\sigma(1)} \circ \dots \circ q_{\sigma(k)}$  où on a  $k$  variables non observées et  $\sigma$  est une permutation de  $\{1, \dots, k\}$ .

On devrait montrer que si  $q_i$  et  $q_j$  ont  $\pi$  comme distribution stationnaire, alors  $q_i \circ q_j$  aussi.

Et donc on peut conclure que la transition  $q$  a pour distribution stationnaire  $\mathbb{P}(\vec{x} | e)$ .

On rappelle que

- la couverture de Markov est l'ensemble des variables parentes, filles, et parents des filles.
- une variable est indépendante de toute autre variable étant donnée sa couverture de Markov
- $\mathbb{P}(x_i | cm(X_i)) \propto \mathbb{P}(x_i | Parents(X_i)) \cdot \prod_{Y_j \in Enfants(X_i)} \mathbb{P}(y_j | Parents(Y_j))$ .

Donc  $\mathbb{P}(x'_i | x_{-i}, e) = \mathbb{P}(\vec{x} | cm(X_i))$  où  $cm$  est la **couverture de Markov**.

$$\mathbb{P}(x'_i | cm(X_i)) \propto \mathbb{P}(x'_i | Parents(X_i)) \cdot \prod_{Y_j \in Enfants(X_i)} \mathbb{P}(y_j | Parents(Y_j))$$

## Echantillonnage de Gibbs – algorithme

---

### Estimation de $\mathbb{P}(Y|e)$ avec échantillonnage de Gibbs

```
1 | N vecteur de décompte pour chaque valeur de X, initialisé à zéro
2 | Z les variables non observées
3 | x l'état courant
4 |
5 | initialise x avec e et tirage aléatoires (uniforme) pour les variables de Z
6 |
7 | for t = 1 to Tmax
8 |     for Zi ∈ Z
9 |         Tire une valeur z de Zi de la distribution  $\mathbb{P}(Z_i | cm(Z_i))$ 
10 |          $x \leftarrow (x_1, \dots, x_{k-1}, z, x_{k+1}, x_n)$ 
11 |          $N(y) \leftarrow N(y) + 1$  où y est la valeur de Y dans x
12 | return Normalisation de N
```

## Echantillonnage de Gibbs

---

- tirer des échantillons de  $\mathbb{P}(Z_i | cm(Z_i))$  est beaucoup plus facile que de tirer un échantillon de la distribution complète.
- meilleure méthode pour le moment. Mais il existe encore de meilleures méthodes pour approximer (hors programme : approximation variationnelle, loopy propagation)