Lecture notes for *Principles of Artificial Intelligence* (Com S 4720/5720)
Yan-Bin Jia, Iowa State University

Markov Chain Monte Carlo (MCMC) Simulation

Outline

I. Gibbs simulation

II. Posterior distribution under the Markov blanket

III. Markov chains

* Figures are from the textbook site.

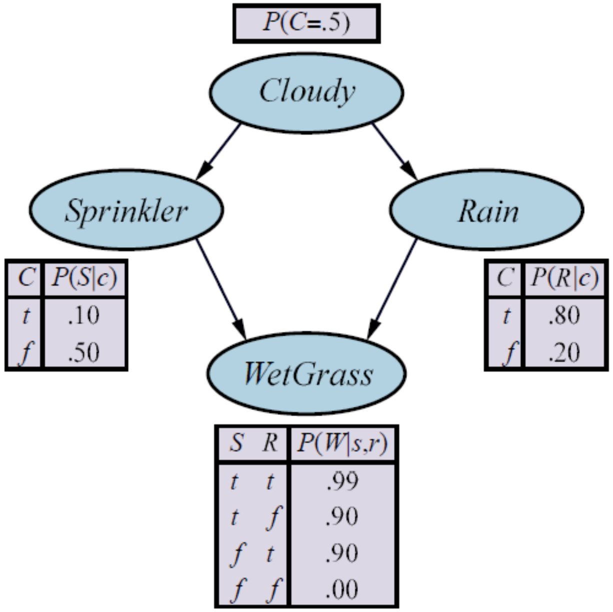
Markov Chain Monte Carlo (MCMC) Simulation

Outline

I. Gibbs simulation

II. Posterior distribution under the Markov blanket

III. Markov chains



* Figures are from the textbook site.

Use of Markov Chains

- A Markov Chain Monte Carlo (MCMC) algorithm
  - specifies a value for every variable at the current *state* (i.e., sample), and
  - generates a next state by making random changes to the current state.

- *Markov chain* is a random process that generates a sequence of states.

Gibbs Sampling

An MCMC algorithm is well suited for Bayes nets that

- starts with an arbitrary state
- fix evidence variables at their observed values
- chose a variable X_i out of the m nonevidence variables with a specified probability:

$$\rho(i) = P(X_i \text{ is chosen among } X_1, \ldots, X_m)$$

- randomly sample a value for the chosen variable X_i

Gibbs Sampling

An MCMC algorithm is well suited for Bayes nets that

- starts with an arbitrary state
- fix evidence variables at their observed values
- chose a variable X_i out of the m nonevidence variables with a specified probability:

$$\rho(i) = P(X_i \text{ is chosen among } X_1, \ldots, X_m \ )$$

- randomly sample a value for the chosen variable X_i

X_i is independent of all the variables outside of its *Markov blanket* $mb(X_i)$.

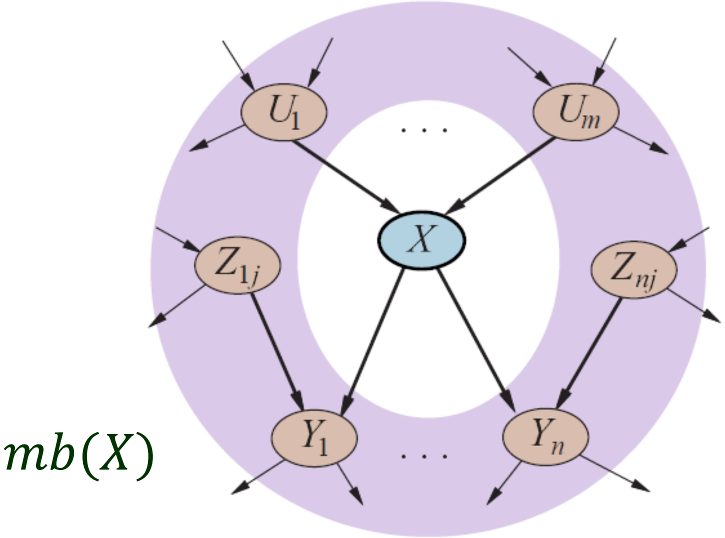
Gibbs Sampling

An MCMC algorithm is well suited for Bayes nets that

- starts with an arbitrary state
- fix evidence variables at their observed values
- chose a variable X_i out of the m nonevidence variables with a specified probability:

$$\rho(i) = P(X_i \text{ is chosen among } X_1, \ldots, X_m\ )$$

- randomly sample a value for the chosen variable X_i



X_i is independent of all the variables outside of its *Markov blanket* $mb(X_i)$.

↑

X_i's parents, children, and children's other parents

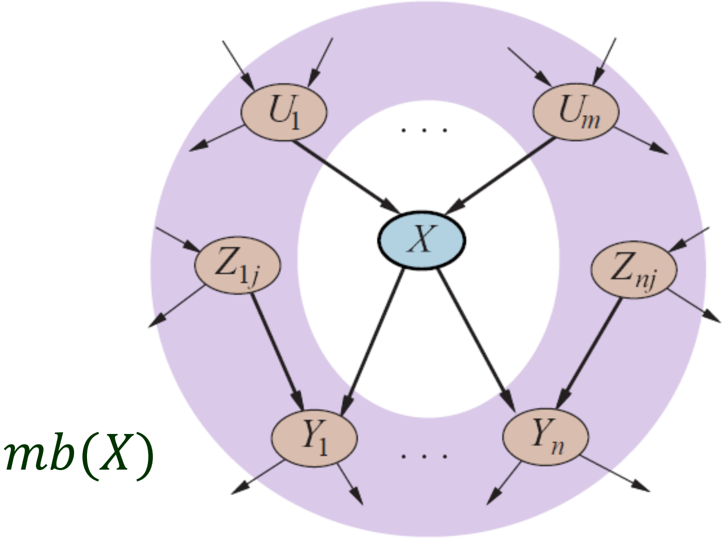
Gibbs Sampling

An MCMC algorithm is well suited for Bayes nets that

- starts with an arbitrary state
- fix evidence variables at their observed values
- chose a variable X_i out of the m nonevidence variables with a specified probability:

$$\rho(i) = P(X_i \text{ is chosen among } X_1, \ldots, X_m\ )$$

- randomly sample a value for the chosen variable X_i according to

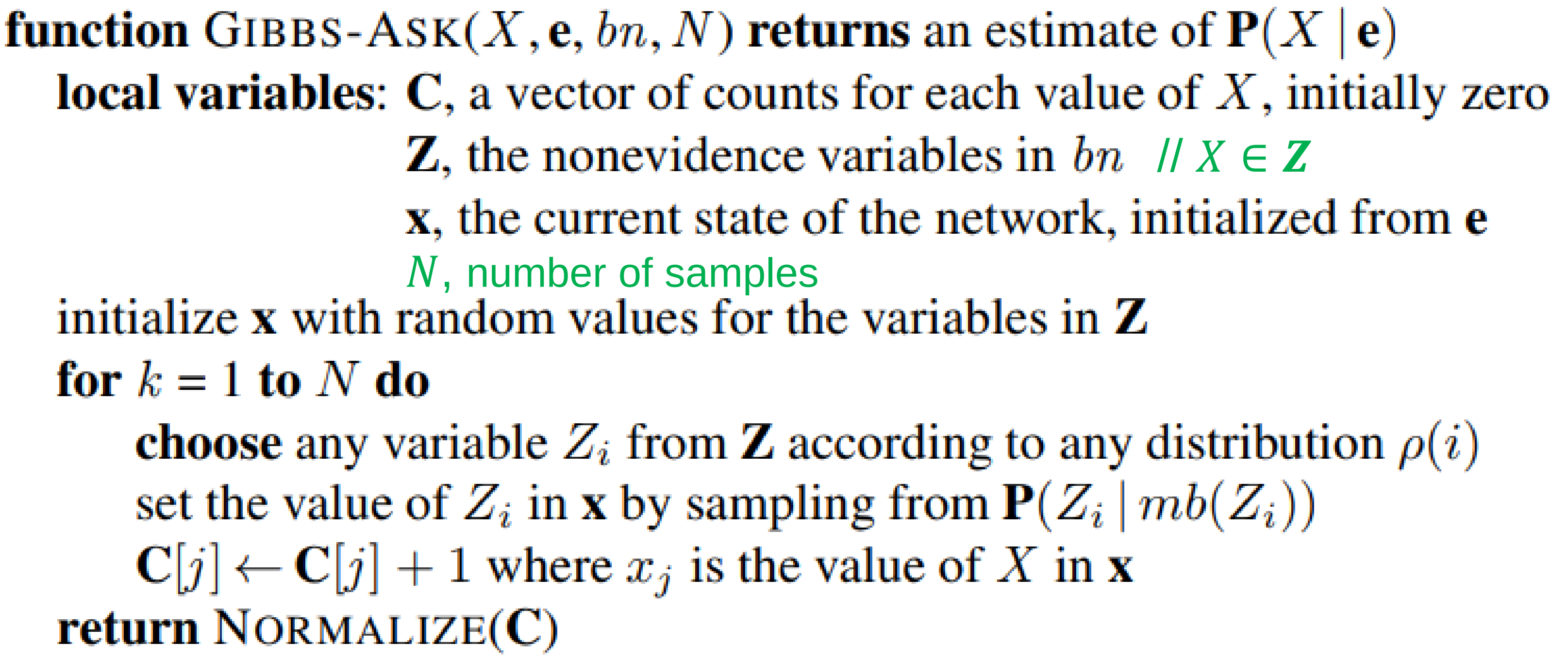


$\boldsymbol{P}(X_i \mid mb(X_i))$ // how to compute?
// (described later)

X_i is independent of all the variables outside of its *Markov blanket* $mb(X_i)$.

X_i's parents, children, and children's other parents

The Algorithm



```
function GIBBS-ASK($X$, $\mathbf{e}$, $bn$, $N$) returns an estimate of $\mathbf{P}(X \mid \mathbf{e})$
    local variables: $\mathbf{C}$, a vector of counts for each value of $X$, initially zero
                     $\mathbf{Z}$, the nonevidence variables in $bn$   // $X \in \mathbf{Z}$
                     $\mathbf{x}$, the current state of the network, initialized from $\mathbf{e}$
                     $N$, number of samples
    initialize $\mathbf{x}$ with random values for the variables in $\mathbf{Z}$
    for $k$ = 1 to $N$ do
        choose any variable $Z_i$ from $\mathbf{Z}$ according to any distribution $\rho(i)$
        set the value of $Z_i$ in $\mathbf{x}$ by sampling from $\mathbf{P}(Z_i \mid mb(Z_i))$
        $\mathbf{C}[j] \leftarrow \mathbf{C}[j] + 1$ where $x_j$ is the value of $X$ in $\mathbf{x}$
    return NORMALIZE($\mathbf{C}$)
```

The Algorithm

```
function GIBBS-ASK($X$, $\mathbf{e}$, $bn$, $N$) returns an estimate of $\mathbf{P}(X \mid \mathbf{e})$
  local variables: $\mathbf{C}$, a vector of counts for each value of $X$, initially zero
                   $\mathbf{Z}$, the nonevidence variables in $bn$   // $X \in \boldsymbol{Z}$
                   $\mathbf{x}$, the current state of the network, initialized from $\mathbf{e}$
                   $N$, number of samples
  initialize $\mathbf{x}$ with random values for the variables in $\mathbf{Z}$
  for $k$ = 1 to $N$ do
      choose any variable $Z_i$ from $\mathbf{Z}$ according to any distribution $\rho(i)$
      set the value of $Z_i$ in $\mathbf{x}$ by sampling from $\mathbf{P}(Z_i \mid mb(Z_i))$
      $\mathbf{C}[j] \leftarrow \mathbf{C}[j] + 1$ where $x_j$ is the value of $X$ in $\mathbf{x}$
  return NORMALIZE($\mathbf{C}$)
```

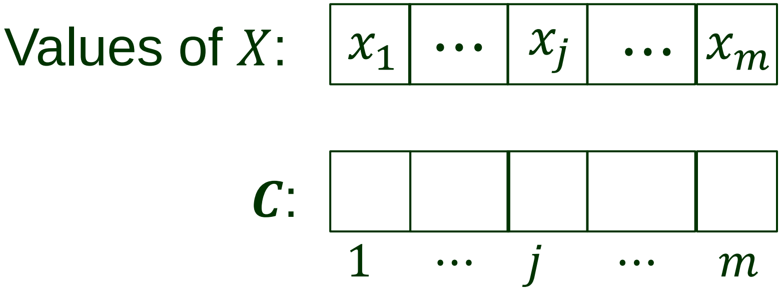
Values of X:	x_1	$\cdots$	x_j	$\cdots$	x_m
$\boldsymbol{C}$:					
	1	$\cdots$	j	$\cdots$	m

The Algorithm

```
function GIBBS-ASK($X$, $\mathbf{e}$, $bn$, $N$) returns an estimate of $\mathbf{P}(X \mid \mathbf{e})$
  local variables: $\mathbf{C}$, a vector of counts for each value of $X$, initially zero
                   $\mathbf{Z}$, the nonevidence variables in $bn$   // $X \in \boldsymbol{Z}$
                   $\mathbf{x}$, the current state of the network, initialized from $\mathbf{e}$
                   $N$, number of samples
  initialize $\mathbf{x}$ with random values for the variables in $\mathbf{Z}$
  for $k$ = 1 to $N$ do
      choose any variable $Z_i$ from $\mathbf{Z}$ according to any distribution $\rho(i)$
      set the value of $Z_i$ in $\mathbf{x}$ by sampling from $\mathbf{P}(Z_i \mid mb(Z_i))$
      $\mathbf{C}[j] \leftarrow \mathbf{C}[j] + 1$ where $x_j$ is the value of $X$ in $\mathbf{x}$
  return NORMALIZE($\mathbf{C}$)
```

Two cases in iteration k:

- $Z_i = X$. x_j is the newly sampled value of X. The index j changes from the previous iteration only if the value x_j does.
- $Z_i \neq X$. The value of X does not change, neither does j. The same counter $\boldsymbol{C}[j]$ as in the previous iteration is incremented again.



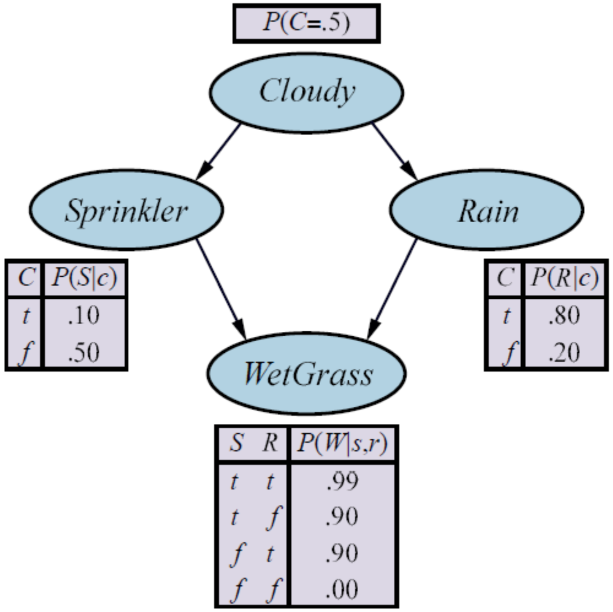
Values of X:	x_1	$\cdots$	x_j	$\cdots$	x_m
$\boldsymbol{C}$:					
	1	$\cdots$	j	$\cdots$	m

Example of Gibbs Sampling

Gibbs sampling for X_i is conditioned on the current values of the variables in its Markov blanket.

Example of Gibbs Sampling

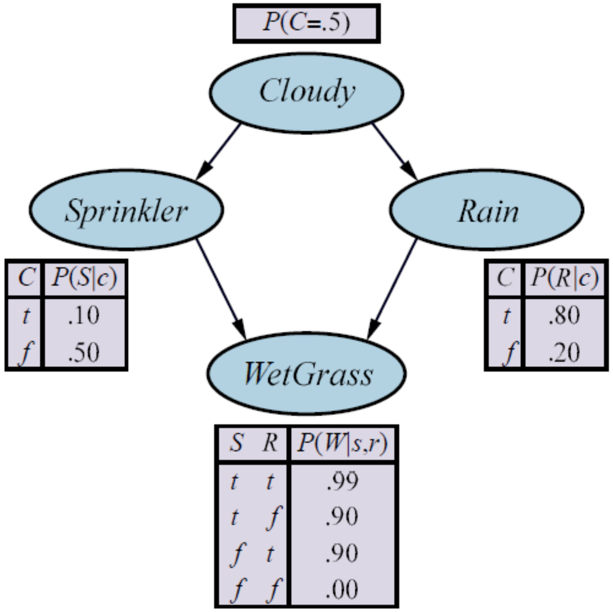
Gibbs sampling for X_i is conditioned on the current values of the variables in its Markov blanket.



Order: *Cloudy*, *Sprinkler*, *Rain*, *WetGrass*

Example of Gibbs Sampling

Gibbs sampling for X_i is conditioned on the current values of the variables in its Markov blanket.

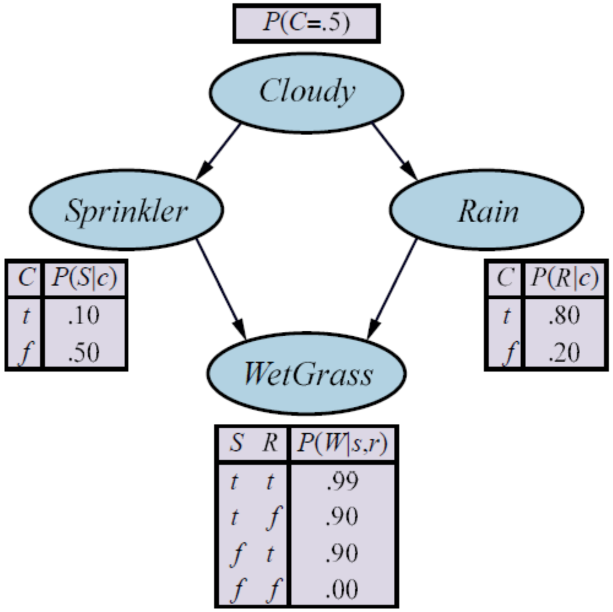


Query $\boldsymbol{P}(Rain \mid Sprinkler = true, WetGrass = true)$

Order: *Cloudy*, *Sprinkler*, *Rain*, *WetGrass*

Example of Gibbs Sampling

Gibbs sampling for X_i is conditioned on the current values of the variables in its Markov blanket.



C	P(S|c)
t	.10
f	.50

C	P(R|c)
t	.80
f	.20

S	R	P(W|s,r)
t	t	.99
t	f	.90
f	t	.90
f	f	.00

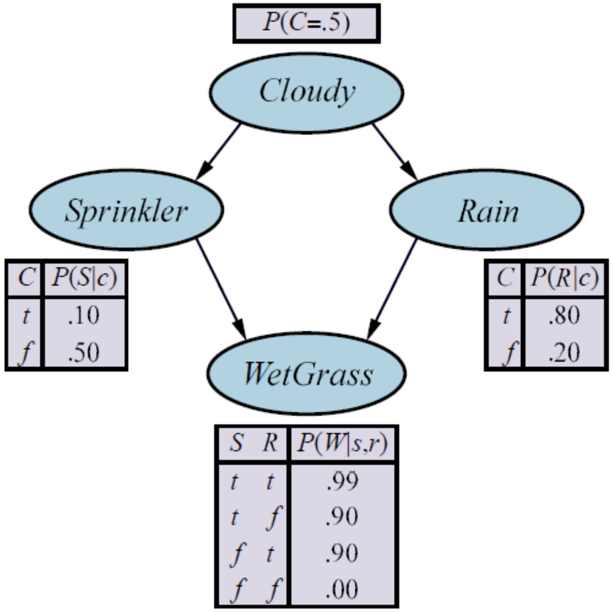
Query $\boldsymbol{P}(\mathit{Rain} \mid \mathit{Sprinkler} = \mathit{true}, \mathit{WetGrass} = \mathit{true})$

- Initial state [*true*,*true*,*false*,*true*]

Order: *Cloudy*, *Sprinkler*, *Rain*, *WetGrass*

Example of Gibbs Sampling

Gibbs sampling for X_i is conditioned on the current values of the variables in its Markov blanket.



$P(C{=}.5)$

C	$P(S\|c)$
t	.10
f	.50

C	$P(R\|c)$
t	.80
f	.20

S	R	$P(W\|s,r)$
t	t	.99
t	f	.90
f	t	.90
f	f	.00

Query $\mathbf{P}(Rain \mid Sprinkler = true, WetGrass = true)$

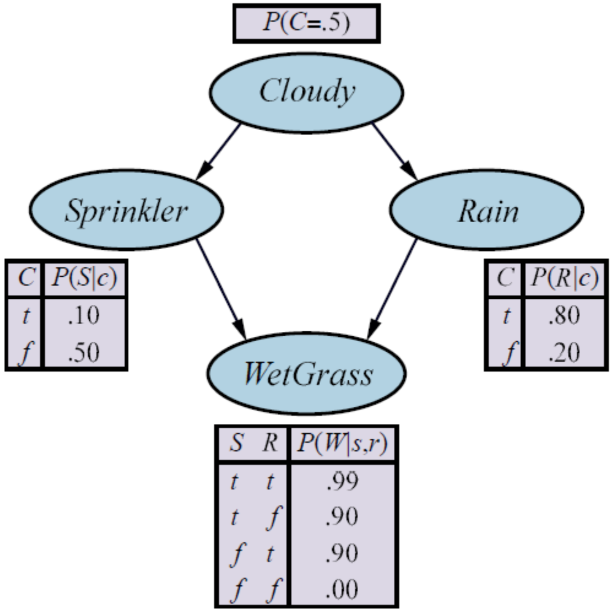
- Initial state [*true*,*true*,*false*,*true*]

evidence variables *Sprinkler* and
WetGrass fixed to their observed values

Order: *Cloudy*, *Sprinkler*, *Rain*, *WetGrass*

Example of Gibbs Sampling

Gibbs sampling for X_i is conditioned on the current values of the variables in its Markov blanket.



C	P(S|c)
t	.10
f	.50

C	P(R|c)
t	.80
f	.20

S	R	P(W|s,r)
t	t	.99
t	f	.90
f	t	.90
f	f	.00

Query $\mathbf{P}(Rain \mid Sprinkler = true, WetGrass = true)$

randomly generated values for nonevidence variables *Cloudy* and *Rain*

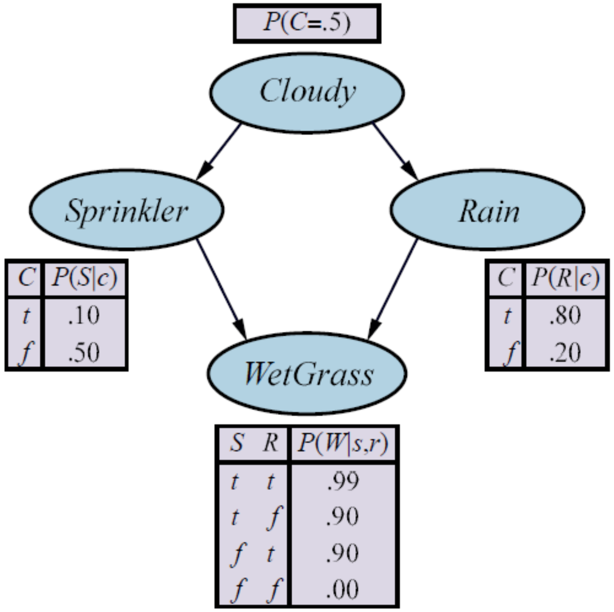
- Initial state [*true*,*true*,*false*,*true*]

evidence variables *Sprinkler* and *WetGrass* fixed to their observed values

Order: *Cloudy*, *Sprinkler*, *Rain*, *WetGrass*

Example (cont'd)

Query $\mathbf{P}(Rain \mid Sprinkler = true, WetGrass = true)$



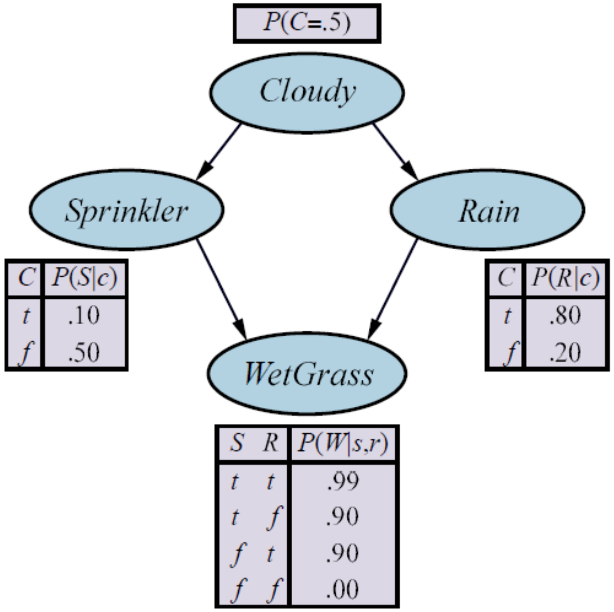
Order: *Cloudy*, *Sprinkler*, *Rain*, *WetGrass*

[*true*,*true*,*false*,*true*]

(initial state)

Example (cont'd)

Query $\boldsymbol{P}(Rain \mid Sprinkler = true, WetGrass = true)$



- Non-evidence variables are then sampled in random order following some probability distribution $\rho(i)$.

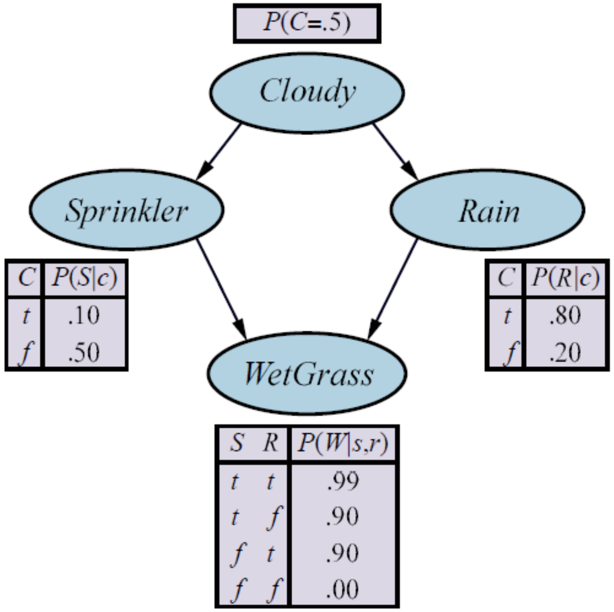
Order: *Cloudy*, *Sprinkler*, *Rain*, *WetGrass*

[*true*,*true*,*false*,*true*]

(initial state)

Example (cont'd)

Query $\boldsymbol{P}(Rain \mid Sprinkler = true, WetGrass = true)$



- Non-evidence variables are then sampled in random order following some probability distribution $\rho(i)$.
  - *Cloudy* is chosen and sampled given the current values of its Markov blanket {*Sprinkler*, *Rain*}.

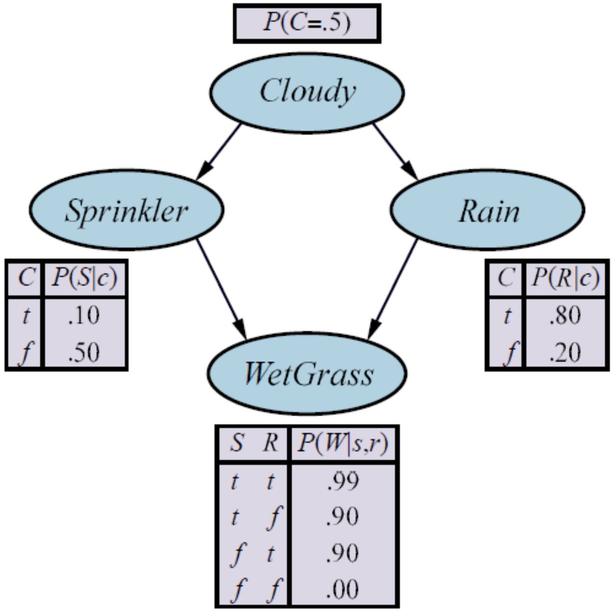
Order: *Cloudy*, *Sprinkler*, *Rain*, *WetGrass*

[*true*,*true*,*false*,*true*]

(initial state)

Example (cont'd)

Query $\boldsymbol{P}(Rain \mid Sprinkler = true, WetGrass = true)$



- Non-evidence variables are then sampled in random order following some probability distribution $\rho(i)$.
  - *Cloudy* is chosen and sampled given the current values of its Markov blanket {*Sprinkler*, *Rain*}.
    - Sampling distribution:

$$\boldsymbol{P}(Cloudy \mid Sprinkler = true, Rain = false\}$$

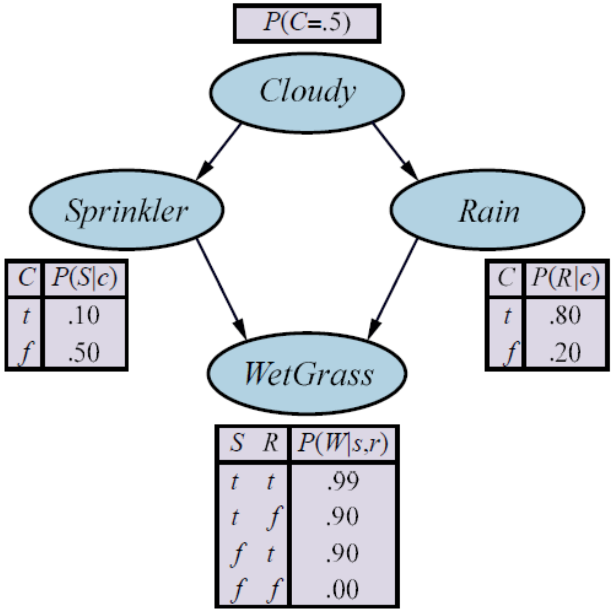
Order: *Cloudy*, *Sprinkler*, *Rain*, *WetGrass*

[*true*,*true*,*false*,*true*] ·

(initial state)

Example (cont’d)

Query $\boldsymbol{P}(Rain \mid Sprinkler = true, WetGrass = true)$



P(C=.5)

C	P(S|c)
t	.10
f	.50

C	P(R|c)
t	.80
f	.20

S	R	P(W|s,r)
t	t	.99
t	f	.90
f	t	.90
f	f	.00

Order: *Cloudy*, *Sprinkler*, *Rain*, *WetGrass*

[*true*,*true*,*false*,*true*]

(initial state)

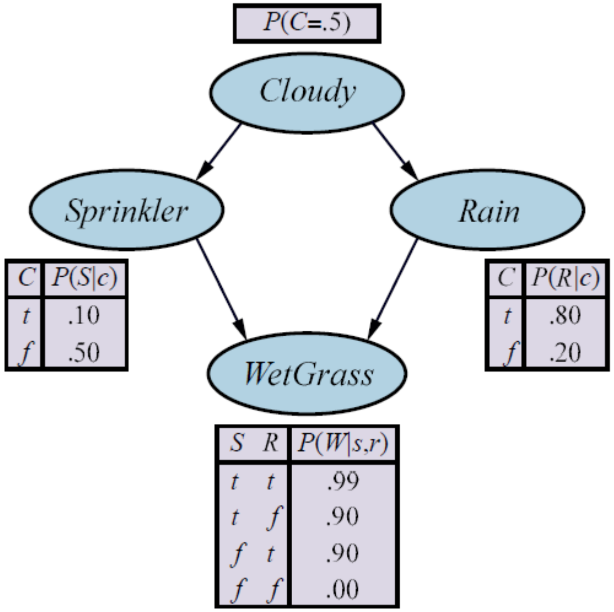
- Non-evidence variables are then sampled in random order following some probability distribution $\rho(i)$.
  - *Cloudy* is chosen and sampled given the current values of its Markov blanket {*Sprinkler*, *Rain*}.
    - Sampling distribution:

$$\boldsymbol{P}(Cloudy \mid Sprinkler = true, Rain = false\}$$

How to calculate Markov blanket distribution $\boldsymbol{P}(X_i \mid mb(X_i))$?

Example (cont'd)

Query $\boldsymbol{P}(Rain \mid Sprinkler = true, WetGrass = true)$



C	P(S|c)
t	.10
f	.50

C	P(R|c)
t	.80
f	.20

S	R	P(W|s,r)
t	t	.99
t	f	.90
f	t	.90
f	f	.00

Order: *Cloudy*, *Sprinkler*, *Rain*, *WetGrass*

[*true*,*true*,*false*,*true*] ·

(initial state)

- Non-evidence variables are then sampled in random order following some probability distribution $\rho(i)$.
  - *Cloudy* is chosen and sampled given the current values of its Markov blanket {*Sprinkler*, *Rain*}.
    - Sampling distribution:

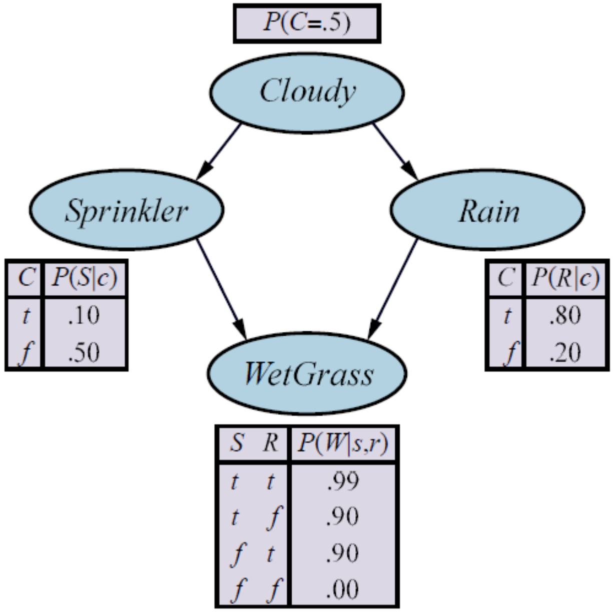
      $$\boldsymbol{P}(Cloudy \mid Sprinkler = true, Rain = false\}$$

    - Sampling result: $Cloudy = false$.

How to calculate Markov blanket distribution $\boldsymbol{P}(X_i \mid mb(X_i))$?

Example (cont'd)

Query $\boldsymbol{P}(\mathit{Rain} \mid \mathit{Sprinkler} = \mathit{true}, \mathit{WetGrass} = \mathit{true})$



C	P(S|c)
t	.10
f	.50

C	P(R|c)
t	.80
f	.20

S	R	P(W|s,r)
t	t	.99
t	f	.90
f	t	.90
f	f	.00

Order: *Cloudy*, *Sprinkler*, *Rain*, *WetGrass*

[*true*,*true*,*false*,*true*] → [*false*,*true*,*false*,*true*]

(initial state) (2nd state)

- Non-evidence variables are then sampled in random order following some probability distribution $\rho(i)$.
    - *Cloudy* is chosen and sampled given the current values of its Markov blanket $\{\mathit{Sprinkler}, \mathit{Rain}\}$.
        - Sampling distribution:

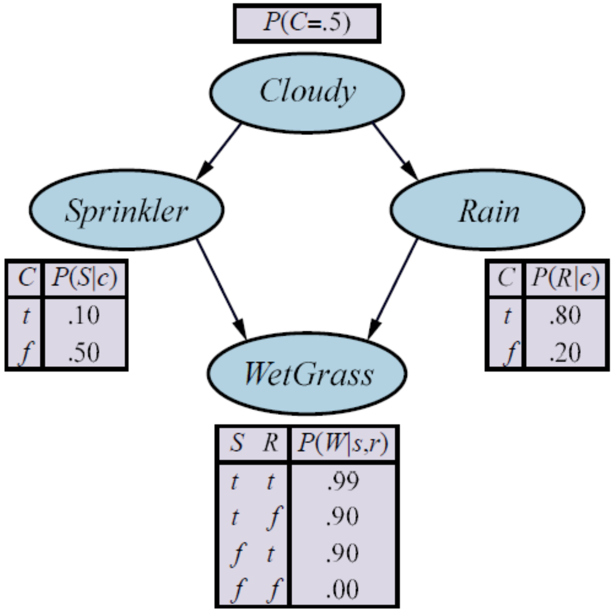
$$\boldsymbol{P}(\mathit{Cloudy} \mid \mathit{Sprinkler} = \mathit{true}, \mathit{Rain} = \mathit{false}\}$$

        - Sampling result: $\mathit{Cloudy} = \mathit{false}$.

How to calculate Markov blanket distribution $\boldsymbol{P}\big(X_i \mid mb(X_i)\big)$?

Example (cont'd)

Query $\boldsymbol{P}(Rain \mid Sprinkler = true, WetGrass = true)$



Order: *Cloudy*, *Sprinkler*, *Rain*, *WetGrass*

[*true*,*true*,*false*,*true*] → [*false*,*true*,*false*,*true*]

(initial state) (2nd state)

- Non-evidence variables are then sampled in random order following some probability distribution $\rho(i)$.
    - ♠ *Cloudy* is chosen and sampled given the current values of its Markov blanket {*Sprinkler*, *Rain*}.
        - ♣ Sampling distribution:

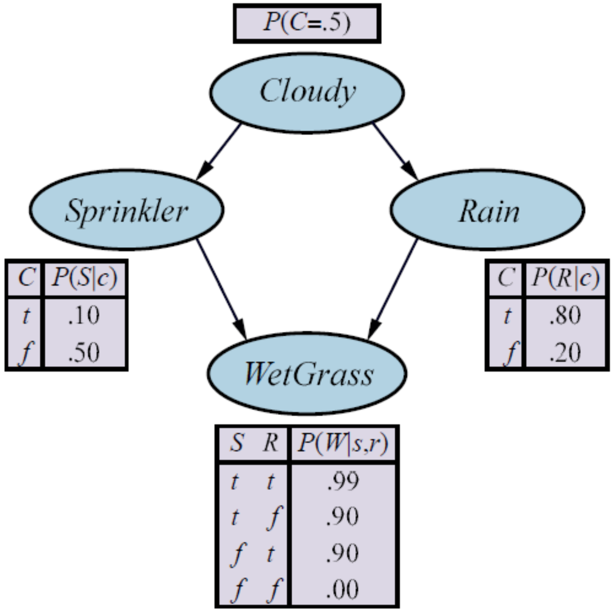
$$\boldsymbol{P}(Cloudy \mid Sprinkler = true, Rain = false\}$$

        - ♣ Sampling result: *Cloudy* = *false*.
    - ♠ *Rain* is chosen next and sampled given the current values of its Markov blanket {*Cloudy*, *Sprinkler*, *WetGrass*}.

How to calculate Markov blanket distribution $\boldsymbol{P}(X_i \mid mb(X_i))$?

Example (cont'd)

Query $\boldsymbol{P}(\mathit{Rain} \mid \mathit{Sprinkler} = \mathit{true}, \mathit{WetGrass} = \mathit{true})$



Order: *Cloudy*, *Sprinkler*, *Rain*, *WetGrass*

[*true*,*true*,*false*,*true*] → [*false*,*true*,*false*,*true*]

(initial state) (2nd state)

- Non-evidence variables are then sampled in random order following some probability distribution $\rho(i)$.
  - *Cloudy* is chosen and sampled given the current values of its Markov blanket {*Sprinkler*, *Rain*}.
    - Sampling distribution:

      $$\boldsymbol{P}(\mathit{Cloudy} \mid \mathit{Sprinkler} = \mathit{true}, \mathit{Rain} = \mathit{false}\}$$

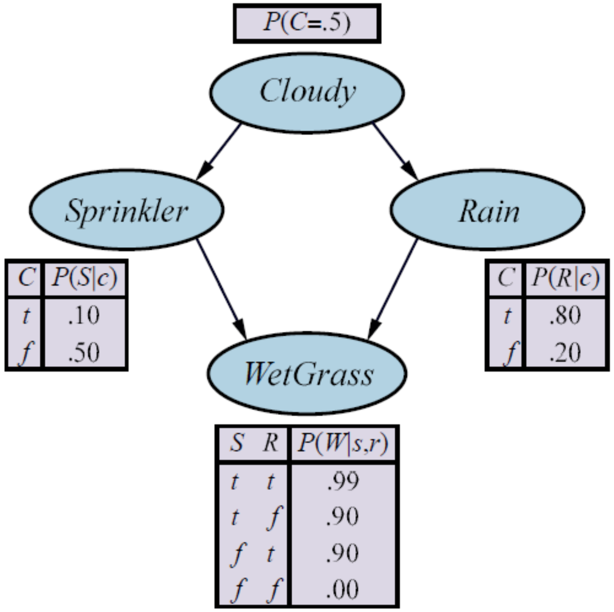
    - Sampling result: $\mathit{Cloudy} = \mathit{false}$.
  - *Rain* is chosen next and sampled given the current values of its Markov blanket {*Cloudy*, *Sprinkler*, *WetGrass*}.
    - Sampling distribution:

      $$\boldsymbol{P}(\mathit{Rain} \mid \mathit{Cloudy} = \mathit{false}, \mathit{Sprinkler} = \mathit{true}, \mathit{WetGrass} = \mathit{true})$$

How to calculate Markov blanket distribution $\boldsymbol{P}(X_i \mid mb(X_i))$?

Example (cont'd)

Query $\boldsymbol{P}(Rain \mid Sprinkler = true, WetGrass = true)$



Order: *Cloudy*, *Sprinkler*, *Rain*, *WetGrass*

[*true*,*true*,*false*,*true*] → [*false*,*true*,*false*,*true*]

(initial state) (2nd state)

- Non-evidence variables are then sampled in random order following some probability distribution $\rho(i)$.
  - *Cloudy* is chosen and sampled given the current values of its Markov blanket {*Sprinkler*, *Rain*}.
    - Sampling distribution:

      $$\boldsymbol{P}(Cloudy \mid Sprinkler = true, Rain = false\}$$

    - Sampling result: $Cloudy = false$.
  - *Rain* is chosen next and sampled given the current values of its Markov blanket {*Cloudy*, *Sprinkler*, *WetGrass*}.
    - Sampling distribution:

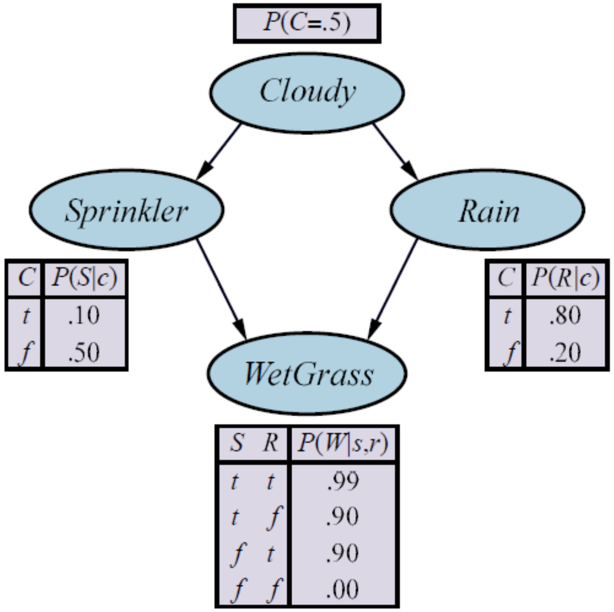
      $$\boldsymbol{P}(Rain \mid Cloudy = false,\ Sprinkler = true,\ WetGrass = true)$$

    - Sampling result: $Rain = true$.

How to calculate Markov blanket distribution $\boldsymbol{P}(X_i \mid mb(X_i))$?

Example (cont'd)

Query $\boldsymbol{P}(Rain \mid Sprinkler = true, WetGrass = true)$



Order: *Cloudy*, *Sprinkler*, *Rain*, *WetGrass*

[*true*,*true*,*false*,*true*] → [*false*,*true*,*false*,*true*]

(initial state) (2nd state)

→ [*false*,*true*,*true*,*true*]

(3rd state)

- Non-evidence variables are then sampled in random order following some probability distribution $\rho(i)$.
  - *Cloudy* is chosen and sampled given the current values of its Markov blanket {*Sprinkler*, *Rain*}.
    - Sampling distribution:

      $$\boldsymbol{P}(Cloudy \mid Sprinkler = true, Rain = false\}$$

    - Sampling result: *Cloudy* = *false*.
  - *Rain* is chosen next and sampled given the current values of its Markov blanket {*Cloudy*, *Sprinkler*, *WetGrass*}.
    - Sampling distribution:

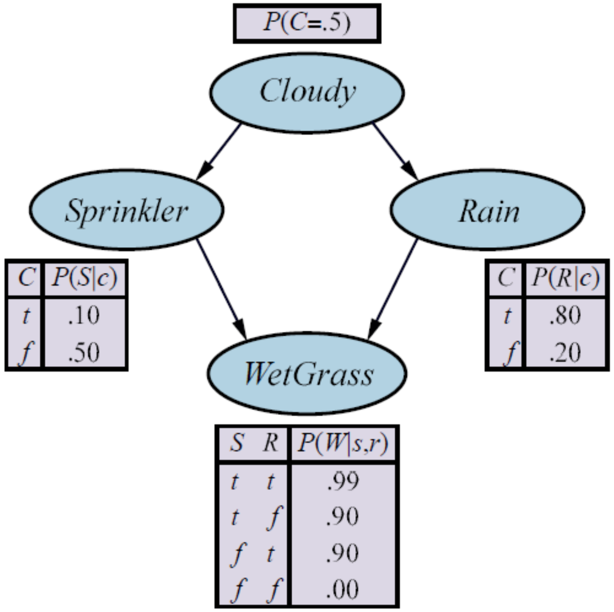
      $$\boldsymbol{P}(Rain \mid Cloudy = false, Sprinkler = true, WetGrass = true)$$

    - Sampling result: *Rain* = *true*.

How to calculate Markov blanket distribution $\boldsymbol{P}(X_i \mid mb(X_i))$?

Example (cont'd)

Query $\boldsymbol{P}(Rain \mid Sprinkler = true, WetGrass = true)$



C	P(S|c)
t	.10
f	.50

C	P(R|c)
t	.80
f	.20

S	R	P(W|s,r)
t	t	.99
t	f	.90
f	t	.90
f	f	.00

Order: *Cloudy*, *Sprinkler*, *Rain*, *WetGrass*

[*true*,*true*,*false*,*true*] → [*false*,*true*,*false*,*true*]

(initial state) (2nd state)

→ [*false*,*true*,*true*,*true*]

(3rd state)

- Non-evidence variables are then sampled in random order following some probability distribution $\rho(i)$.
  - *Cloudy* is chosen and sampled given the current values of its Markov blanket {*Sprinkler*, *Rain*}.
    - Sampling distribution:

      $$\boldsymbol{P}(Cloudy \mid Sprinkler = true, Rain = false\}$$

    - Sampling result: *Cloudy* = *false*.
  - *Rain* is chosen next and sampled given the current values of its Markov blanket {*Cloudy*, *Sprinkler*, *WetGrass*}.
    - Sampling distribution:

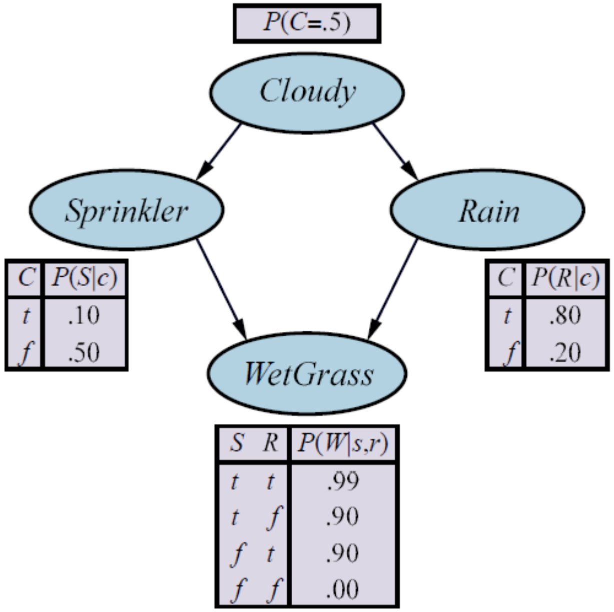
      $$\boldsymbol{P}(Rain \mid Cloudy = false,\ Sprinkler = true,\ WetGrass = true)$$

    - Sampling result: *Rain* = *true*.

How to calculate Markov blanket distribution $\boldsymbol{P}(X_i \mid mb(X_i))$?

Example (cont'd)

Query $\boldsymbol{P}(\mathit{Rain} \mid \mathit{Sprinkler} = \mathit{true}, \mathit{WetGrass} = \mathit{true})$



Order: *Cloudy*, *Sprinkler*, *Rain*, *WetGrass*

[*true*,*true*,*false*,*true*] → [*false*,*true*,*false*,*true*]

(initial state) (2nd state)

→ [*false*,*true*,*true*,*true*]

(3rd state)

→ ⋯

- Non-evidence variables are then sampled in random order following some probability distribution $\rho(i)$.
  - *Cloudy* is chosen and sampled given the current values of its Markov blanket {*Sprinkler*, *Rain*}.
    - Sampling distribution:

      $$\boldsymbol{P}(\mathit{Cloudy} \mid \mathit{Sprinkler} = \mathit{true}, \mathit{Rain} = \mathit{false}\}$$

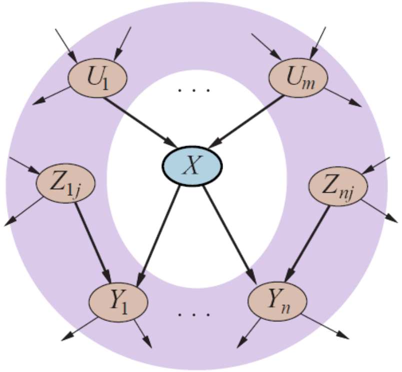
    - Sampling result: *Cloudy* = *false*.
  - *Rain* is chosen next and sampled given the current values of its Markov blanket {*Cloudy*, *Sprinkler*, *WetGrass*}.
    - Sampling distribution:

      $$\boldsymbol{P}(\mathit{Rain} \mid \mathit{Cloudy} = \mathit{false},\ \mathit{Sprinkler} = \mathit{true},\ \mathit{WetGrass} = \mathit{true})$$

    - Sampling result: *Rain* = *true*.

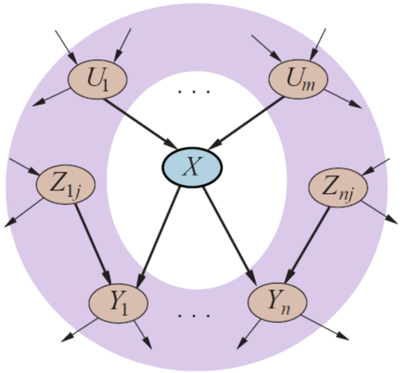
How to calculate Markov blanket distribution $\boldsymbol{P}(X_i \mid mb(X_i))$?

II. Distribution Given the Markov Blanket



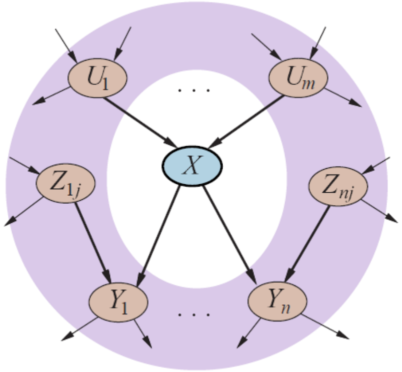
- $MB(X_i)$: variables in the Markov blanket of X_i.
  - ♣ *Parents*(X_i): parents of X_i (e.g., $U_1, \ldots, U_m$)
  - ♣ *Children*(X_i): children of X_i (e.g., $Y_1, \ldots, Y_n$)
  - ♣ *Others*(X_i): other parents of X_i's children (e.g, $Z_{1j}, \ldots, Z_{nj}, \ldots$)

II. Distribution Given the Markov Blanket



- $MB(X_i)$: variables in the Markov blanket of X_i.
  - ♣ *Parents*(X_i): parents of X_i (e.g., $U_1, \dots, U_m$)
  - ♣ *Children*(X_i): children of X_i (e.g., $Y_1, \dots, Y_n$)
  - ♣ *Others*(X_i): other parents of X_i's children (e.g, $Z_{1j}, \dots, Z_{nj}, \dots$)
- $mb(X_i)$: values of the variables in $MB(X_i)$.
  - ♣ *parents*(X_i): values of the parents of X_i
  - ♣ *children*(X_i): values of the children of X_i
  - ♣ *others*(X_i): values of the other parents of X_i's children

II. Distribution Given the Markov Blanket

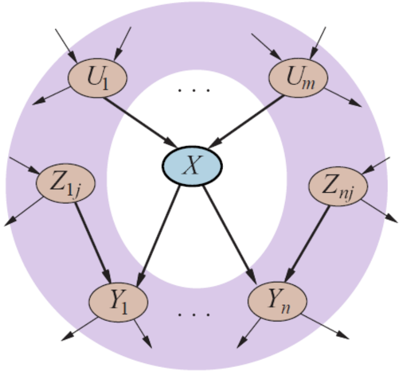


- $MB(X_i)$: variables in the Markov blanket of X_i.
  - ♣ *Parents*(X_i): parents of X_i (e.g., $U_1, \ldots, U_m$)
  - ♣ *Children*(X_i): children of X_i (e.g., $Y_1, \ldots, Y_n$)
  - ♣ *Others*(X_i): other parents of X_i's children (e.g, $Z_{1j}, \ldots, Z_{nj}, \ldots$)
- $mb(X_i)$: values of the variables in $MB(X_i)$.
  - ♣ *parents*(X_i): values of the parents of X_i
  - ♣ *children*(X_i): values of the children of X_i
  - ♣ *others*(X_i): values of the other parents of X_i's children

We want to compute the probability distribution $\boldsymbol{P}(X_i \mid mb(X_i)\}$.

$$\boldsymbol{P}\left(X_i \mid mb(X_i)\right) = \boldsymbol{P}(X_i \mid \textit{parents}(X_i), \textit{children}(X_i), \textit{others}(X_i))$$

II. Distribution Given the Markov Blanket



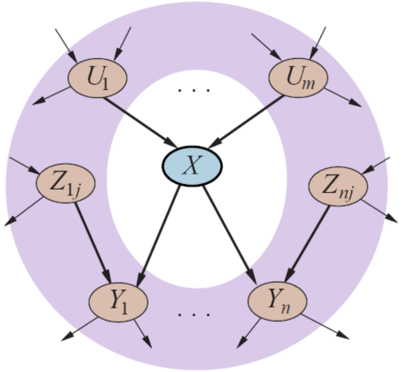
- $MB(X_i)$: variables in the Markov blanket of X_i.
    - ♣ *Parents*(X_i): parents of X_i (e.g., $U_1, \dots, U_m$)
    - ♣ *Children*(X_i): children of X_i (e.g., $Y_1, \dots, Y_n$)
    - ♣ *Others*(X_i): other parents of X_i's children (e.g, $Z_{1j}, \dots, Z_{nj}, \dots$)
- $mb(X_i)$: values of the variables in $MB(X_i)$.
    - ♣ *parents*(X_i): values of the parents of X_i
    - ♣ *children*(X_i): values of the children of X_i
    - ♣ *others*(X_i): values of the other parents of X_i's children

We want to compute the probability distribution $\boldsymbol{P}(X_i \mid mb(X_i)\}$.

$$\boldsymbol{P}\left(X_i \mid mb(X_i)\right) = \boldsymbol{P}(X_i \mid \textit{parents}(X_i), \textit{children}(X_i), \textit{others}(X_i))$$

$$= \alpha \boldsymbol{P}\ (X_i, \textit{parents}(X_i), \textit{children}(X_i), \textit{others}(X_i))$$

II. Distribution Given the Markov Blanket



- $MB(X_i)$: variables in the Markov blanket of X_i.
  - *Parents*(X_i): parents of X_i (e.g., $U_1, \ldots, U_m$)
  - *Children*(X_i): children of X_i (e.g., $Y_1, \ldots, Y_n$)
  - *Others*(X_i): other parents of X_i's children (e.g, $Z_{1j}, \ldots, Z_{nj}, \ldots$)
- $mb(X_i)$: values of the variables in $MB(X_i)$.
  - *parents*(X_i): values of the parents of X_i
  - *children*(X_i): values of the children of X_i
  - *others*(X_i): values of the other parents of X_i's children

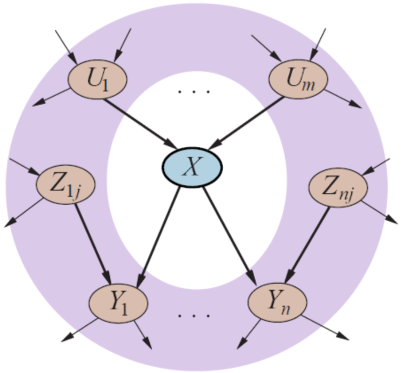
We want to compute the probability distribution $\boldsymbol{P}(X_i \mid mb(X_i)\}$.

$$\boldsymbol{P}\big(X_i \mid mb(X_i)\big) = \boldsymbol{P}(X_i \mid \mathit{parents}(X_i), \mathit{children}(X_i), \mathit{others}(X_i))$$

$$= \alpha \boldsymbol{P}\ (X_i, \mathit{parents}(X_i), \mathit{children}(X_i), \mathit{others}(X_i))$$

$$= \alpha \boldsymbol{P}(X_i \mid \mathit{parents}(X_i), \mathit{others}(X_i)) \cdot \boldsymbol{P}\big(\mathit{children}(X_i) \mid \mathit{parents}(X_i), X_i, \mathit{others}(X_i)\big)$$

II. Distribution Given the Markov Blanket

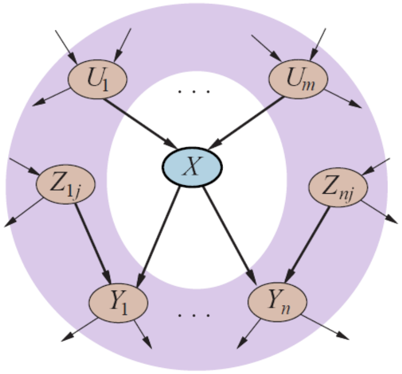


- $MB(X_i)$: variables in the Markov blanket of X_i.
    - *Parents*(X_i): parents of X_i (e.g., $U_1, \ldots, U_m$)
    - *Children*(X_i): children of X_i (e.g., $Y_1, \ldots, Y_n$)
    - *Others*(X_i): other parents of X_i's children (e.g, $Z_{1j}, \ldots, Z_{nj}, \ldots$)
- $mb(X_i)$: values of the variables in $MB(X_i)$.
    - *parents*(X_i): values of the parents of X_i
    - *children*(X_i): values of the children of X_i
    - *others*(X_i): values of the other parents of X_i's children

We want to compute the probability distribution $\boldsymbol{P}(X_i \mid mb(X_i)\}$.

$$
\begin{aligned}
\boldsymbol{P}\big(X_i \mid mb(X_i)\big) &= \boldsymbol{P}(X_i \mid \mathit{parents}(X_i), \mathit{children}(X_i), \mathit{others}(X_i)) \\
&= \alpha \boldsymbol{P}\ (X_i, \mathit{parents}(X_i), \mathit{children}(X_i), \mathit{others}(X_i)) \\
&= \alpha \boldsymbol{P}(X_i \mid \mathit{parents}(X_i), \mathit{others}(X_i)) \cdot \boldsymbol{P}\big(\mathit{children}(X_i) \mid \mathit{parents}(X_i), X_i, \mathit{others}(X_i)\big) \\
&= \alpha \boldsymbol{P}(X_i \mid \mathit{parents}(X_i)) \cdot \boldsymbol{P}\big(\mathit{children}(X_i) \mid X_i, \mathit{others}(X_i)\big)
\end{aligned}
$$

II. Distribution Given the Markov Blanket



- $MB(X_i)$: variables in the Markov blanket of X_i.
  - ♣ *Parents*(X_i): parents of X_i (e.g., $U_1, \ldots, U_m$)
  - ♣ *Children*(X_i): children of X_i (e.g., $Y_1, \ldots, Y_n$)
  - ♣ *Others*(X_i): other parents of X_i's children (e.g, $Z_{1j}, \ldots, Z_{nj}, \ldots$)
- $mb(X_i)$: values of the variables in $MB(X_i)$.
  - ♣ *parents*(X_i): values of the parents of X_i
  - ♣ *children*(X_i): values of the children of X_i
  - ♣ *others*(X_i): values of the other parents of X_i's children

We want to compute the probability distribution $\boldsymbol{P}(X_i \mid mb(X_i)\}$.

$$\boldsymbol{P}\big(X_i \mid mb(X_i)\big) = \boldsymbol{P}(X_i \mid parents(X_i), children(X_i), others(X_i))$$

$$= \alpha \boldsymbol{P}\ (X_i, parents(X_i), children(X_i), others(X_i))$$

$$= \alpha \boldsymbol{P}(X_i \mid parents(X_i), others(X_i)) \cdot \boldsymbol{P}\big(children(X_i) \mid parents(X_i), X_i, others(X_i)\big)$$

$$= \alpha \boldsymbol{P}(X_i \mid parents(X_i)) \cdot \boldsymbol{P}\big(children(X_i) \mid X_i, others(X_i)\big)$$

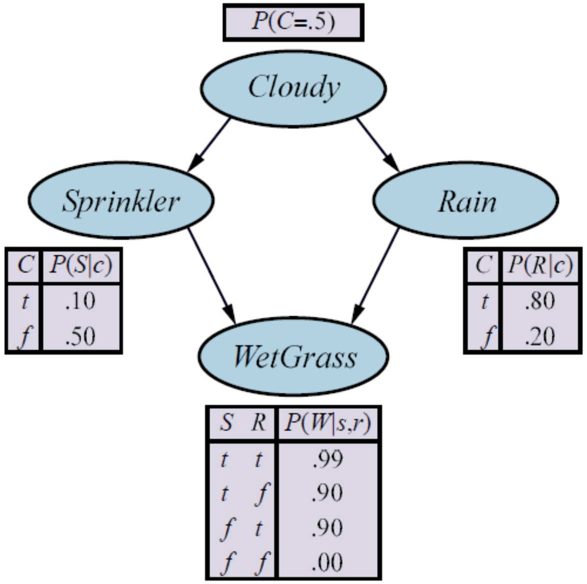
$$= \alpha \boldsymbol{P}(X_i \mid parents(X_i)) \prod_{Y_j \in Children(X_i)} P\big(y_j \mid parents(Y_j)\big)$$

Markov Blanket Distribution

$$\boldsymbol{P}(X_i \mid mb(X_i)) = \alpha \boldsymbol{P}(X_i \mid \mathit{parents}(X_i)) \prod_{Y_j \in \mathit{Children}(X_i)} P(y_j \mid \mathit{parents}(Y_j))$$

Markov Blanket Distribution

$$\boldsymbol{P}(X_i \mid mb(X_i)) = \alpha \boldsymbol{P}(X_i \mid parents(X_i)) \prod_{Y_j \in Children(X_i)} P(y_j \mid parents(Y_j))$$

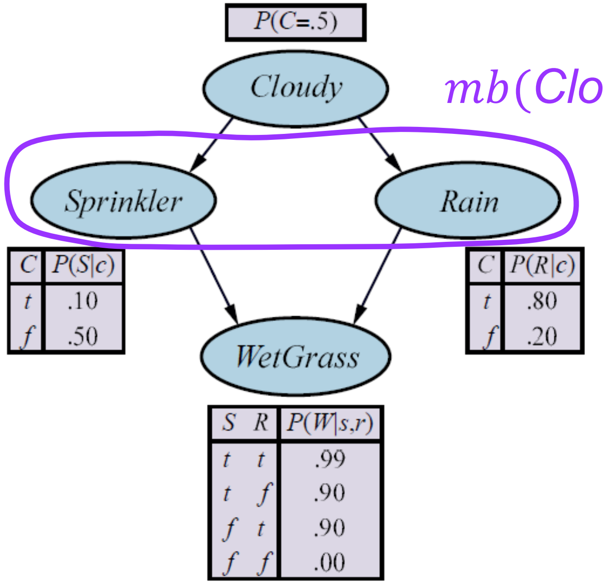


- Sampling distribution:

$$\boldsymbol{P}(Cloudy \mid Sprinkler = true, Rain = false\}$$

Markov Blanket Distribution

$$\boldsymbol{P}(X_i \mid mb(X_i)) = \alpha \boldsymbol{P}(X_i \mid parents(X_i)) \prod_{Y_j \in Children(X_i)} P(y_j \mid parents(Y_j))$$



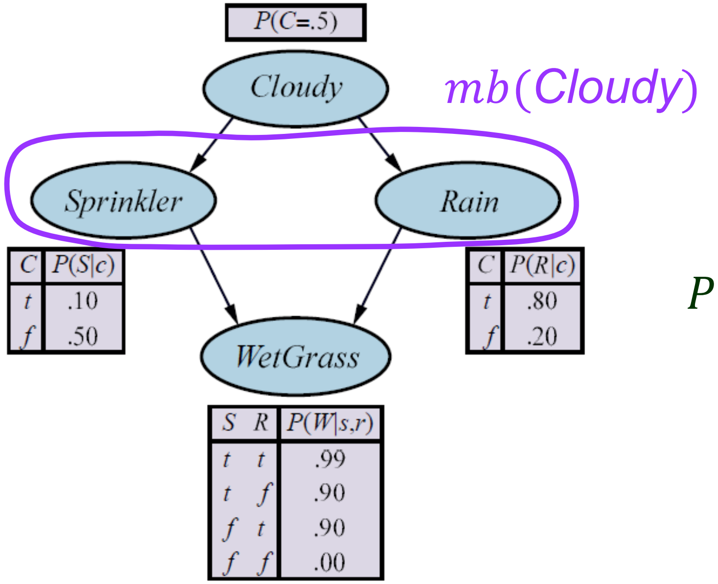
$mb(Cloudy)$

- Sampling distribution:

$$\boldsymbol{P}(Cloudy \mid Sprinkler = true, Rain = false\}$$

Markov Blanket Distribution

$$\mathbf{P}(X_i \mid mb(X_i)) = \alpha \mathbf{P}(X_i \mid parents(X_i)) \prod_{Y_j \in Children(X_i)} P(y_j \mid parents(Y_j))$$



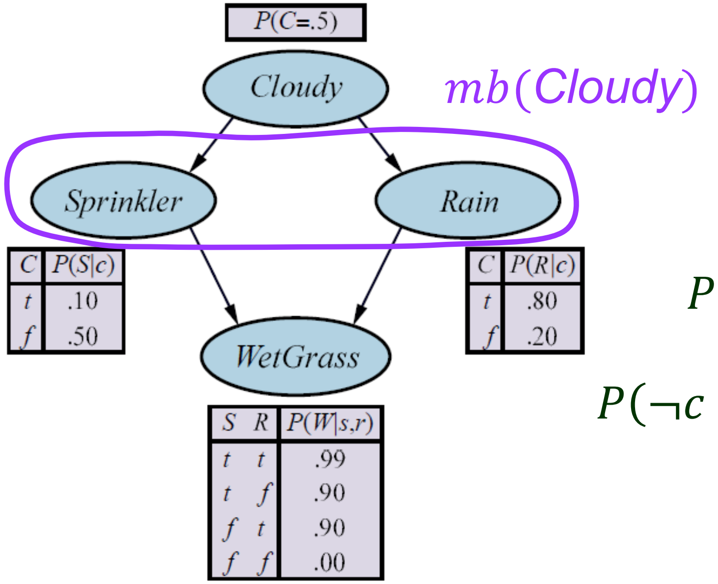
- Sampling distribution:

$$\mathbf{P}(Cloudy \mid Sprinkler = true, Rain = false\}$$

$$P(c \mid s, \neg r) = \alpha P(c) P(s \mid c) P(\neg r \mid c) = \alpha 0.5 \cdot 0.1 \cdot 0.2$$

Markov Blanket Distribution

$$\mathbf{P}(X_i \mid mb(X_i)) = \alpha \mathbf{P}(X_i \mid parents(X_i)) \prod_{Y_j \in Children(X_i)} P(y_j \mid parents(Y_j))$$



- Sampling distribution:

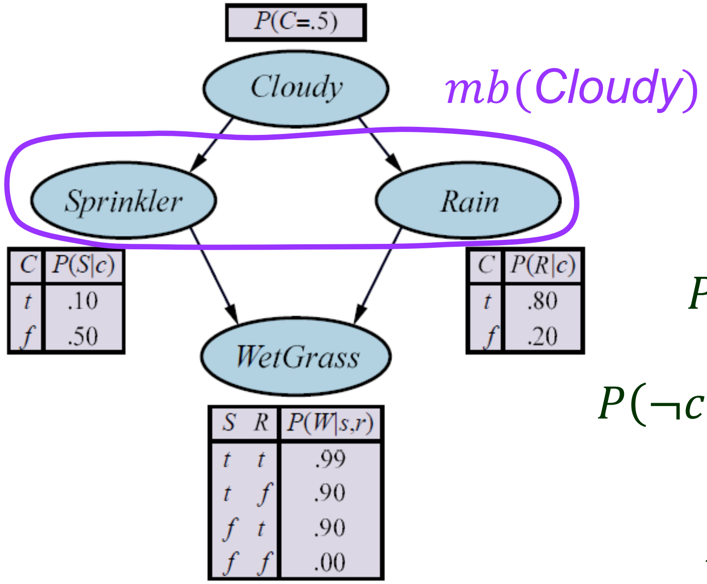
$$\mathbf{P}(Cloudy \mid Sprinkler = true, Rain = false\}$$

$$P(c \mid s, \neg r) = \alpha P(c)P(s \mid c)P(\neg r \mid c) = \alpha 0.5 \cdot 0.1 \cdot 0.2$$

$$P(\neg c \mid s, \neg r) = \alpha P(\neg c)P(s \mid \neg c)P(\neg r \mid \neg c) = \alpha 0.5 \cdot 0.5 \cdot 0.8$$

Markov Blanket Distribution

$$\mathbf{P}(X_i \mid mb(X_i)) = \alpha \mathbf{P}(X_i \mid parents(X_i)) \prod_{Y_j \in Children(X_i)} P(y_j \mid parents(Y_j))$$



C	P(S|c)
t	.10
f	.50

C	P(R|c)
t	.80
f	.20

S	R	P(W|s,r)
t	t	.99
t	f	.90
f	t	.90
f	f	.00

- Sampling distribution:

$$\mathbf{P}(Cloudy \mid Sprinkler = true, Rain = false\}$$

$$P(c \mid s, \neg r) = \alpha P(c)P(s \mid c)P(\neg r \mid c) = \alpha 0.5 \cdot 0.1 \cdot 0.2$$

$$P(\neg c \mid s, \neg r) = \alpha P(\neg c)P(s \mid \neg c)P(\neg r \mid \neg c) = \alpha 0.5 \cdot 0.5 \cdot 0.8$$

$$\mathbf{P}(C \mid s, \neg r\} = \alpha \langle 0.001, 0.020 \rangle \approx \langle 0.048, 0.952 \rangle$$

$mb(\mathit{Cloudy}) = \{\mathit{Sprinkler}, \mathit{Rain}\}$

$mb(\mathit{Rain}) = \{\mathit{Cloudy}, \mathit{Sprinkler}, \mathit{WetGrass}\}$

II. Markov Chain

Query $\boldsymbol{P}(\mathit{Rain} \mid \mathit{Sprinkler} = \mathit{true}, \mathit{WetGrass} = \mathit{true})$

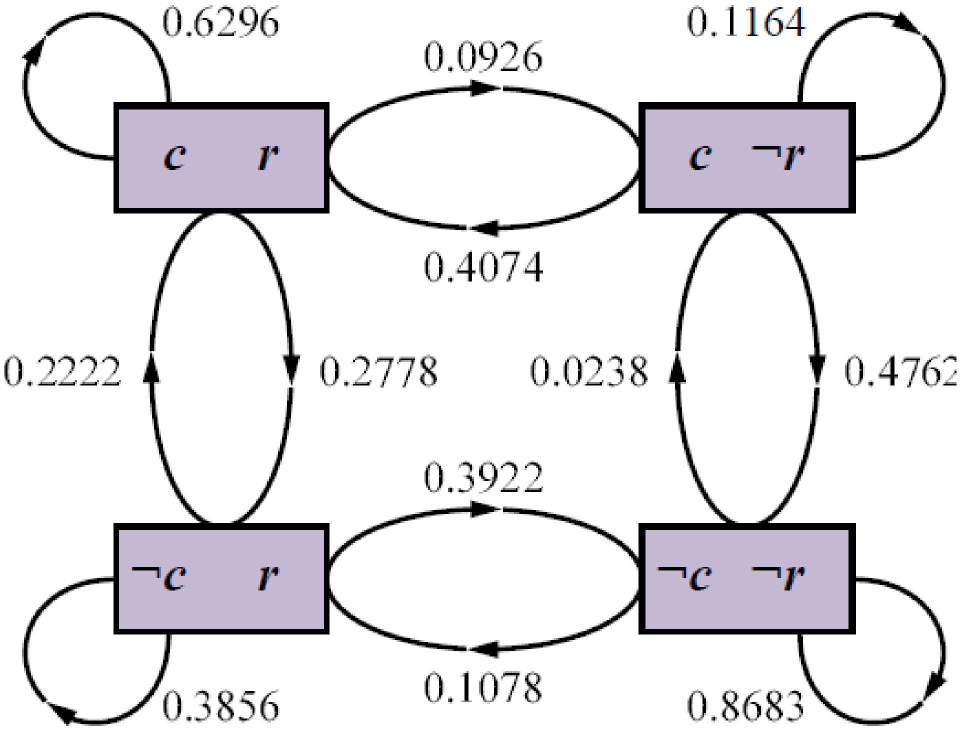
A state *need only include all nonevidence variables*, for example, $\{c, r\}$.

II. Markov Chain

$mb(\mathit{Cloudy}) = \{\mathit{Sprinkler}, \mathit{Rain}\}$
$mb(\mathit{Rain}) = \{\mathit{Cloudy}, \mathit{Sprinkler}, \mathit{WetGrass}\}$

Query $\boldsymbol{P}(\mathit{Rain} \mid \mathit{Sprinkler} = \mathit{true}, \mathit{WetGrass} = \mathit{true})$

A state *need only include all nonevidence variables*, for example, $\{c, r\}$.



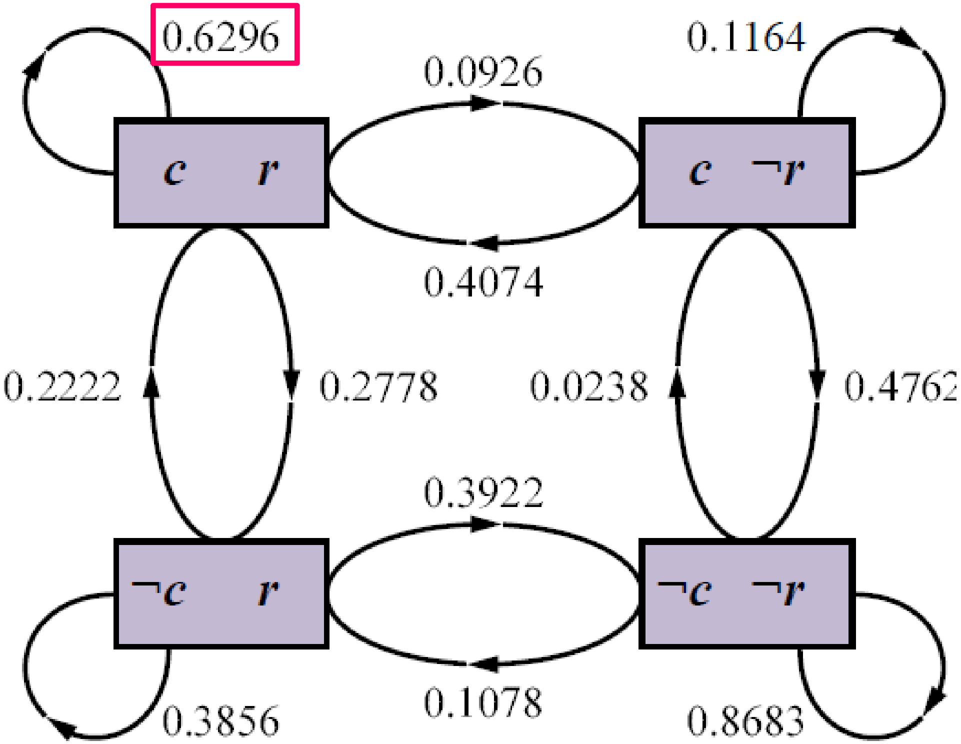
Markov chain from uniform choice of the two nonevidence variables ($\rho(\mathit{Cloudy}) = \rho(\mathit{Rain}) = 0.5$)

$mb(\textit{Cloudy}) = \{\textit{Sprinkler, Rain}\}$
$mb(\textit{Rain}) = \{\textit{Cloudy, Sprinkler, WetGrass}\}$

II. Markov Chain

Query $\mathbf{P}(\textit{Rain} \mid \textit{Sprinkler} = \textit{true}, \textit{WetGrass} = \textit{true})$

A state *need only include all nonevidence variables*, for example, $\{c, r\}$.



$$0.6296 = \rho(\textit{Cloudy}) \cdot P(c \mid r, s)$$
$$+\rho(\textit{Rain}) \cdot P(r \mid c, s, w)$$

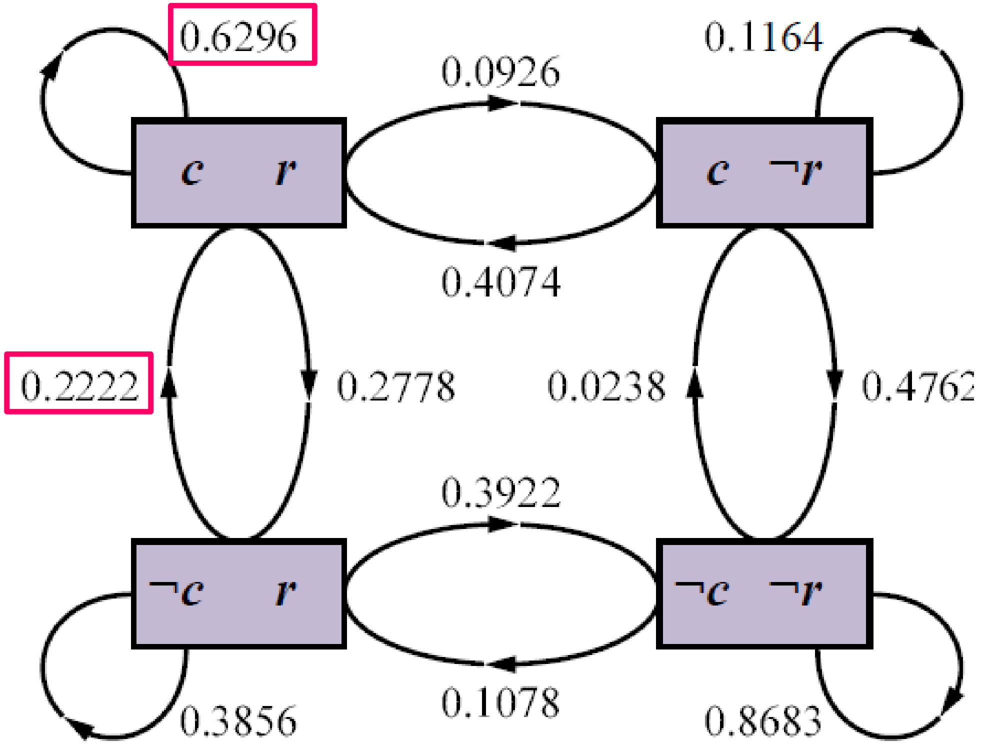
Markov chain from uniform choice of the two nonevidence variables ($\rho(\textit{Cloudy}) = \rho(\textit{Rain}) = 0.5$)

$mb(\mathit{Cloudy}) = \{\mathit{Sprinkler}, \mathit{Rain}\}$
$mb(\mathit{Rain}) = \{\mathit{Cloudy}, \mathit{Sprinkler}, \mathit{WetGrass}\}$

II. Markov Chain

Query $\mathbf{P}(\mathit{Rain} \mid \mathit{Sprinkler} = \mathit{true}, \mathit{WetGrass} = \mathit{true})$

A state *need only include all nonevidence variables*, for example, $\{c, r\}$.



$$0.6296 = \rho(\mathit{Cloudy}) \cdot P(c \mid r, s) \\ +\rho(\mathit{Rain}) \cdot P(r \mid c, s, w)$$

$$0.2222 = \rho(\mathit{Cloudy}) \cdot P(c \mid r, s)$$

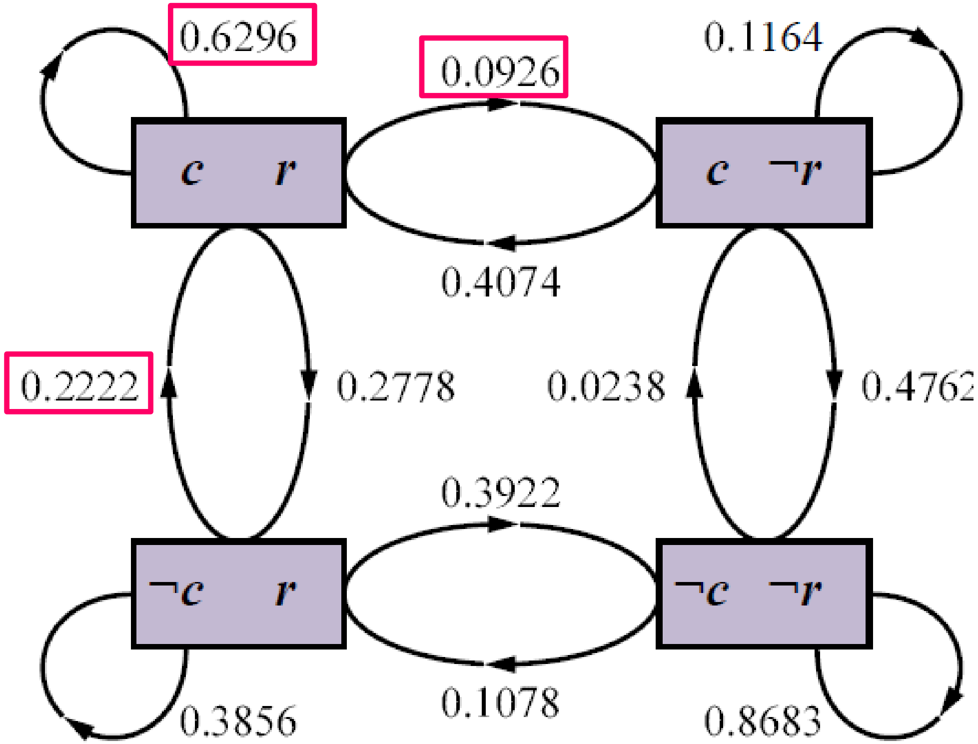
Markov chain from uniform choice of the two nonevidence variables ($\rho(\mathit{Cloudy}) = \rho(\mathit{Rain}) = 0.5$)

$mb(\text{Cloudy}) = \{\text{Sprinkler, Rain}\}$
$mb(\text{Rain}) = \{\text{Cloudy, Sprinkler, WetGrass}\}$

II. Markov Chain

Query $\boldsymbol{P}(\textit{Rain} \mid \textit{Sprinkler} = \textit{true}, \textit{WetGrass} = \textit{true})$

A state *need only include all nonevidence variables*, for example, $\{c, r\}$.



$$0.6296 = \rho(\textit{Cloudy}) \cdot P(c \mid r, s) + \rho(\textit{Rain}) \cdot P(r \mid c, s, w)$$

$$0.2222 = \rho(\textit{Cloudy}) \cdot P(c \mid r, s)$$

$$0.0926 = \rho(\textit{Rain}) \cdot P(\neg r \mid c, s, w)$$

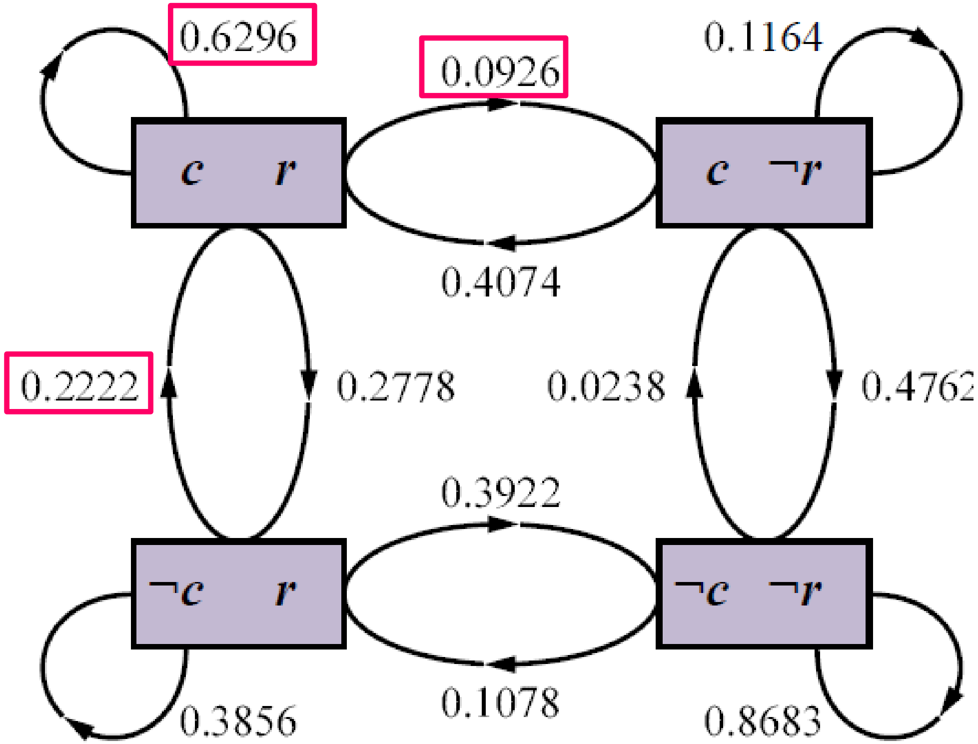
Markov chain from uniform choice of the two nonevidence variables ($\rho(\textit{Cloudy}) = \rho(\textit{Rain}) = 0.5$)

$mb(\textit{Cloudy}) = \{\textit{Sprinkler, Rain}\}$
$mb(\textit{Rain}) = \{\textit{Cloudy, Sprinkler, WetGrass}\}$

II. Markov Chain

Query $\mathbf{P}(\textit{Rain} \mid \textit{Sprinkler} = \textit{true}, \textit{WetGrass} = \textit{true})$

A state *need only include all nonevidence variables*, for example, $\{c, r\}$.



$$0.6296 = \rho(\textit{Cloudy}) \cdot P(c \mid r, s) + \rho(\textit{Rain}) \cdot P(r \mid c, s, w)$$

$$0.2222 = \rho(\textit{Cloudy}) \cdot P(c \mid r, s)$$

$$0.0926 = \rho(\textit{Rain}) \cdot P(\neg r \mid c, s, w)$$

Markov chain from uniform choice of the two nonevidence variables ($\rho(\textit{Cloudy}) = \rho(\textit{Rain}) = 0.5$)

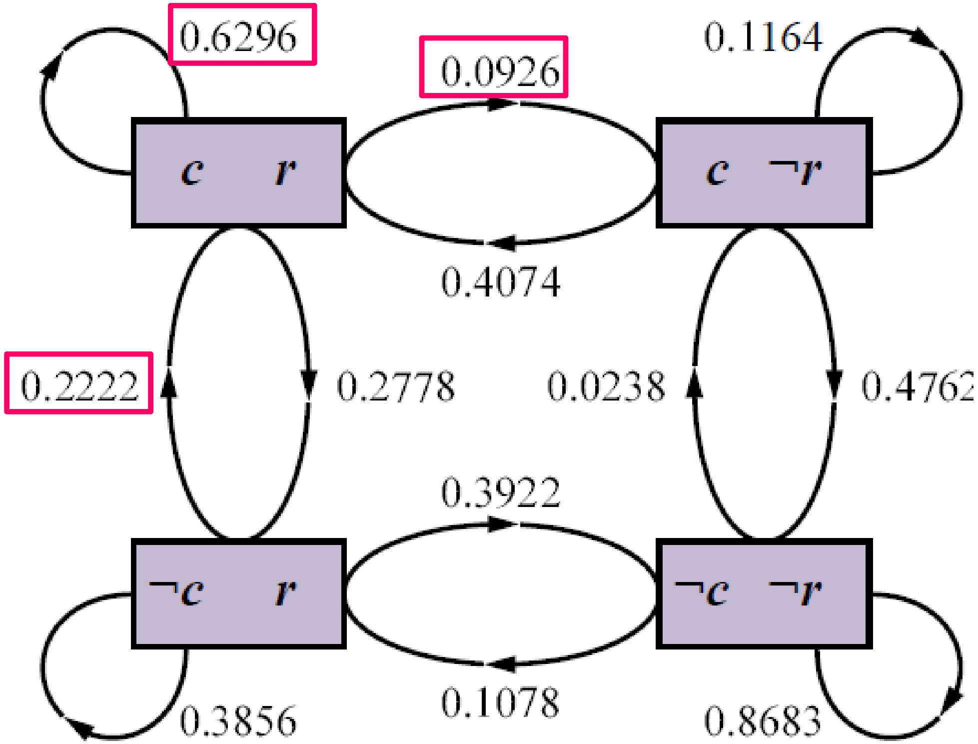
Probabilities with all the outgoing links of each node sum to 1, e.g., $0.6296 + 0.0926 + 0.2778 = 1$.

$mb(\text{Cloudy}) = \{\text{Sprinkler, Rain}\}$
$mb(\text{Rain}) = \{\text{Cloudy, Sprinkler, WetGrass}\}$

II. Markov Chain

Query $\mathbf{P}(\text{Rain} \mid \text{Sprinkler} = \text{true}, \text{WetGrass} = \text{true})$

A state *need only include all nonevidence variables*, for example, $\{c, r\}$.



$$0.6296 = \rho(\text{Cloudy}) \cdot P(c \mid r, s) + \rho(\text{Rain}) \cdot P(r \mid c, s, w)$$

$$0.2222 = \rho(\text{Cloudy}) \cdot P(c \mid r, s)$$

$$0.0926 = \rho(\text{Rain}) \cdot P(\neg r \mid c, s, w)$$

- Gibbs sampling simply wanders around in the graph, following links with probabilities.

Markov chain from uniform choice of the two nonevidence variables ($\rho(\text{Cloudy}) = \rho(\text{Rain}) = 0.5$)

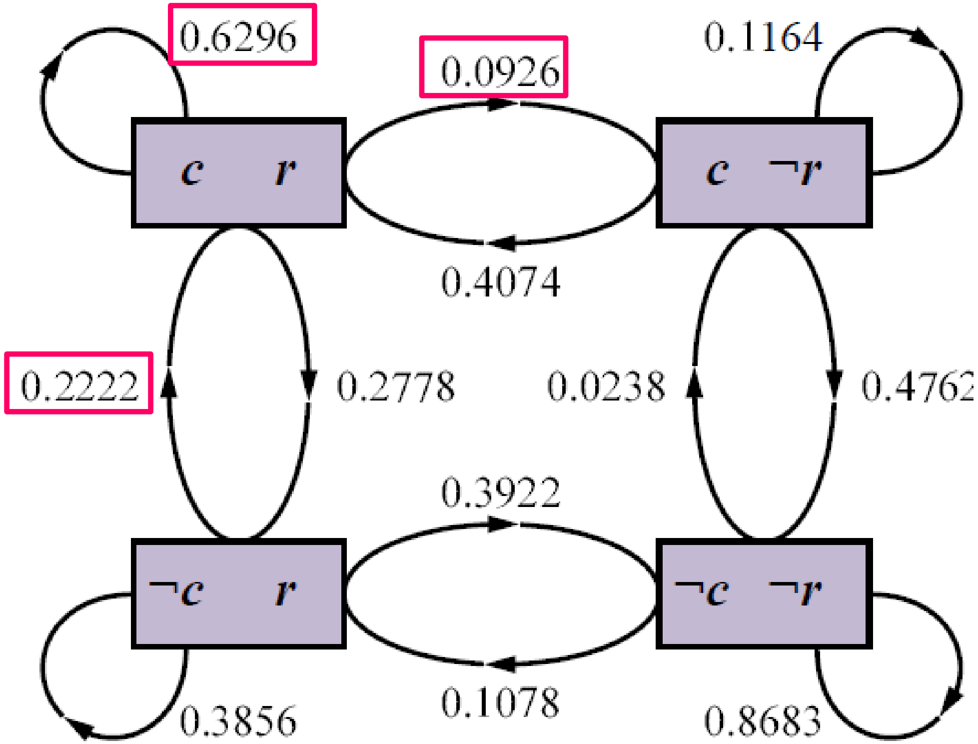
Probabilities with all the outgoing links of each node sum to 1, e.g., $0.6296 + 0.0926 + 0.2778 = 1$.

II. Markov Chain

$mb(\textit{Cloudy}) = \{\textit{Sprinkler, Rain}\}$
$mb(\textit{Rain}) = \{\textit{Cloudy, Sprinkler, WetGrass}\}$

Query $\mathbf{P}(\textit{Rain} \mid \textit{Sprinkler} = \textit{true}, \textit{WetGrass} = \textit{true})$

A state *need only include all nonevidence variables*, for example, $\{c, r\}$.



$$0.6296 = \rho(\textit{Cloudy}) \cdot P(c \mid r, s)$$
$$+\rho(\textit{Rain}) \cdot P(r \mid c, s, w)$$
$$0.2222 = \rho(\textit{Cloudy}) \cdot P(c \mid r, s)$$
$$0.0926 = \rho(\textit{Rain}) \cdot P(\neg r \mid c, s, w)$$

- Gibbs sampling simply wanders around in the graph, following links with probabilities.
- Every state visited is a sample that contributes to the estimate for the query variable *Rain*.

Markov chain from uniform choice of the two nonevidence variables ($\rho(\textit{Cloudy}) = \rho(\textit{Rain}) = 0.5$)

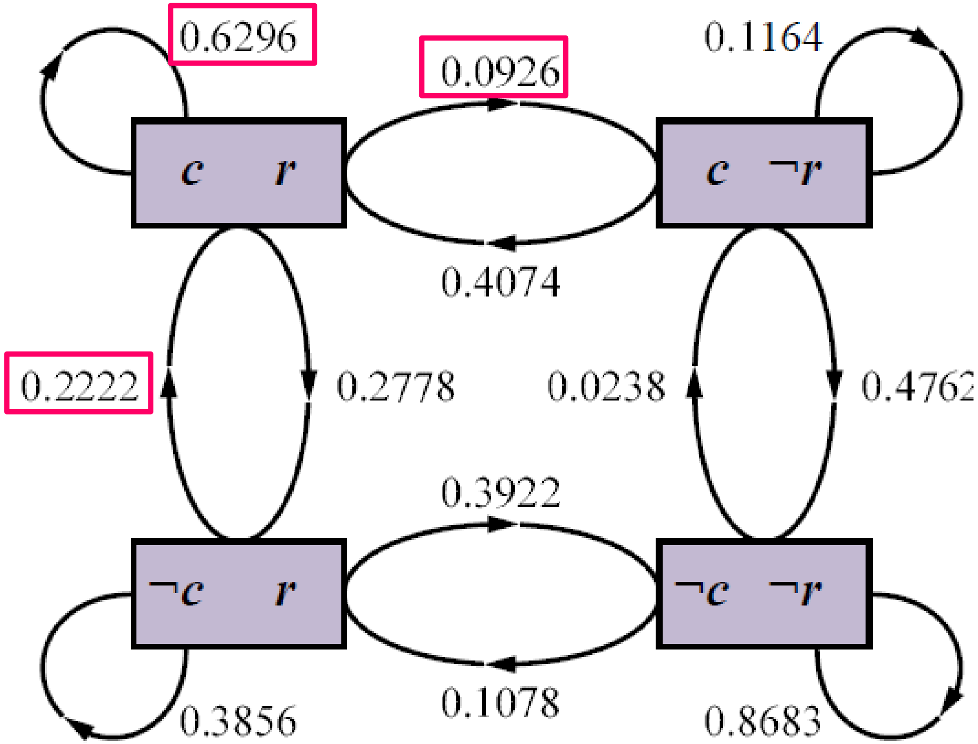
Probabilities with all the outgoing links of each node sum to 1, e.g., $0.6296 + 0.0926 + 0.2778 = 1$.

$mb(\mathit{Cloudy}) = \{\mathit{Sprinkler}, \mathit{Rain}\}$
$mb(\mathit{Rain}) = \{\mathit{Cloudy}, \mathit{Sprinkler}, \mathit{WetGrass}\}$

II. Markov Chain

Query $\mathbf{P}(\mathit{Rain} \mid \mathit{Sprinkler} = \mathit{true}, \mathit{WetGrass} = \mathit{true})$

A state *need only include all nonevidence variables*, for example, $\{c, r\}$.



Markov chain from uniform choice of the two nonevidence variables ($\rho(\mathit{Cloudy}) = \rho(\mathit{Rain}) = 0.5$)

Probabilities with all the outgoing links of each node sum to 1, e.g., $0.6296 + 0.0926 + 0.2778 = 1$.

$$0.6296 = \rho(\mathit{Cloudy}) \cdot P(c \mid r, s) + \rho(\mathit{Rain}) \cdot P(r \mid c, s, w)$$

$$0.2222 = \rho(\mathit{Cloudy}) \cdot P(c \mid r, s)$$

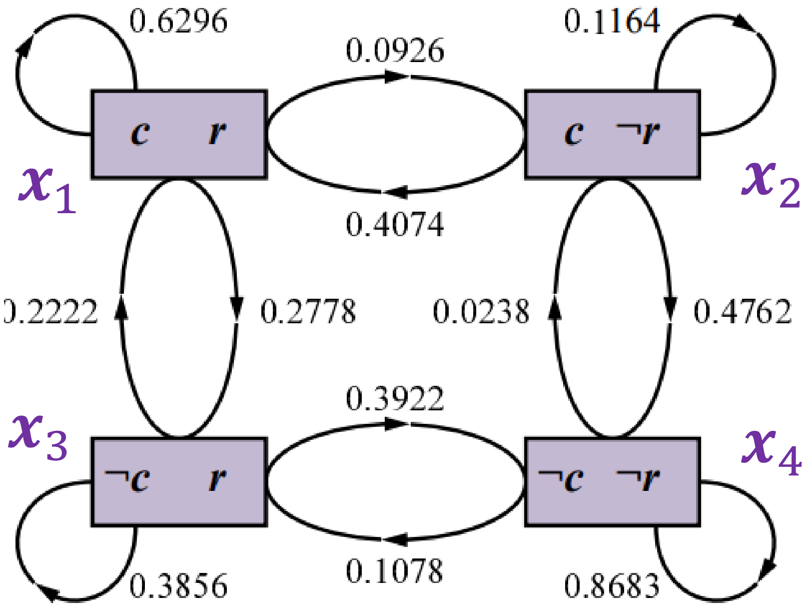
$$0.0926 = \rho(\mathit{Rain}) \cdot P(\neg r \mid c, s, w)$$

- Gibbs sampling simply wanders around in the graph, following links with probabilities.
- Every state visited is a sample that contributes to the estimate for the query variable *Rain*.

If the process visits 20 states with *Rain* = *true* and 60 states with *Rain* = *false*, then the answer to the query is $\alpha\langle 20,60\rangle = \langle 0.25,0.75\rangle$.

Analysis of Markov Chains

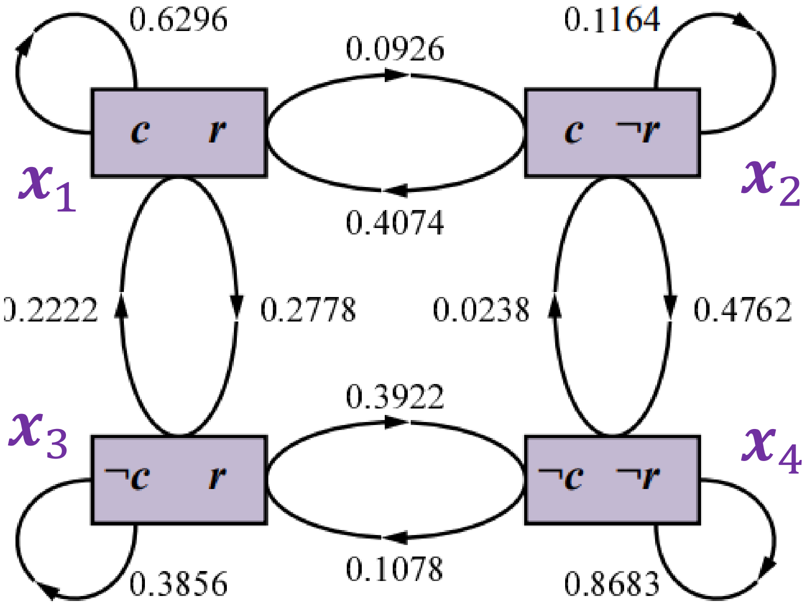
Why does Gibbs sampling work? Or, why does its estimates converge to correct values in the limit?



Analysis of Markov Chains

Why does Gibbs sampling work? Or, why does its estimates converge to correct values in the limit?

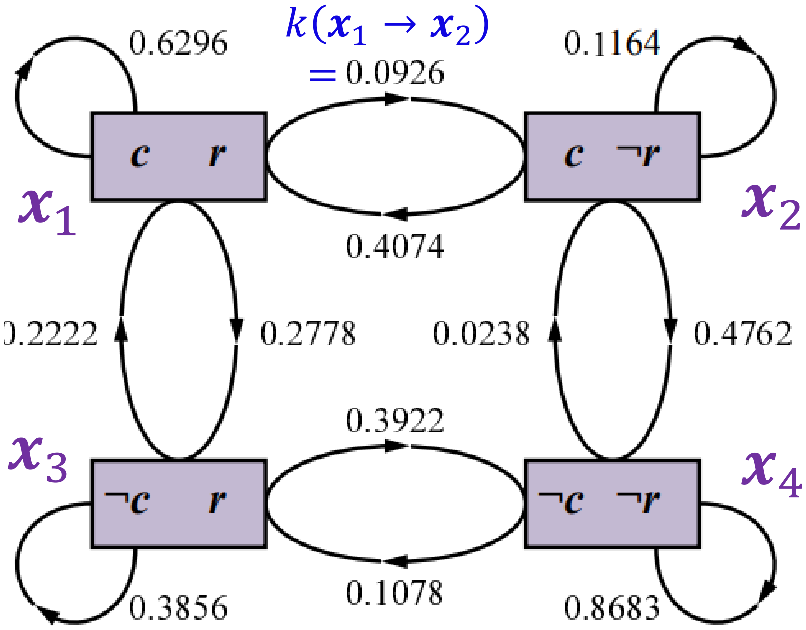
Transition kernel k assigns a probability $k(\boldsymbol{x} \rightarrow \boldsymbol{x}')$ to every transition from a state $\boldsymbol{x}$ to another state $\boldsymbol{x}'$ in the Markov chain.



Analysis of Markov Chains

Why does Gibbs sampling work? Or, why does its estimates converge to correct values in the limit?

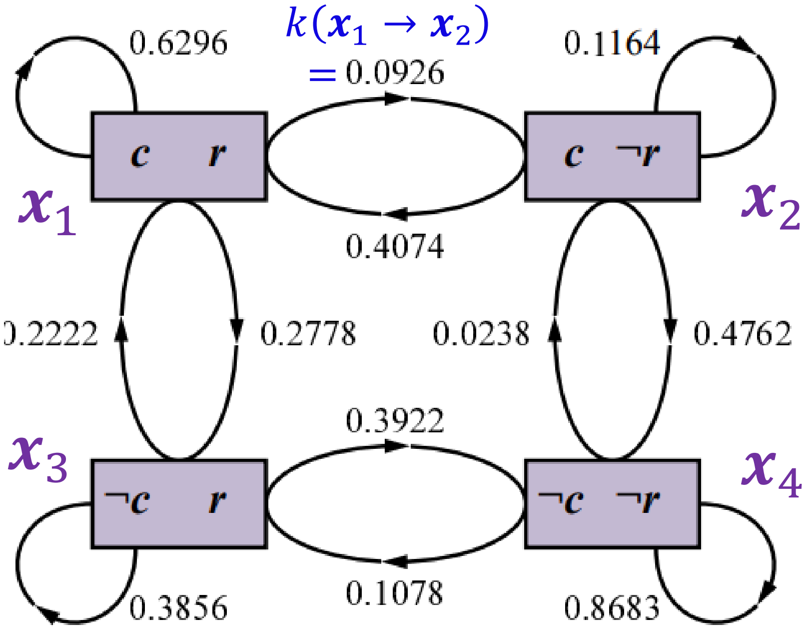
Transition kernel k assigns a probability $k(\boldsymbol{x} \to \boldsymbol{x}')$ to every transition from a state $\boldsymbol{x}$ to another state $\boldsymbol{x}'$ in the Markov chain.



Analysis of Markov Chains

Why does Gibbs sampling work? Or, why does its estimates converge to correct values in the limit?

Transition kernel k assigns a probability $k(\boldsymbol{x} \rightarrow \boldsymbol{x}')$ to every transition from a state $\boldsymbol{x}$ to another state $\boldsymbol{x}'$ in the Markov chain.

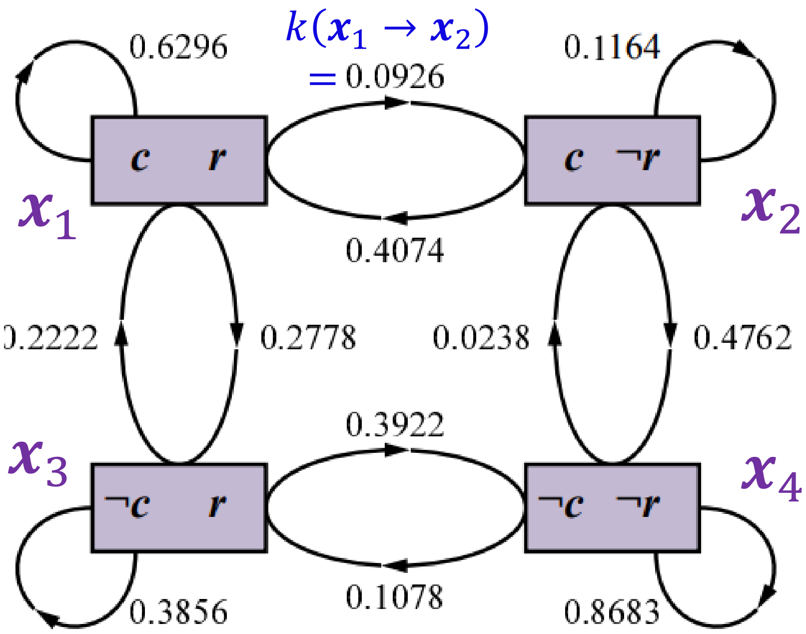


$\pi_t(\boldsymbol{x})$: probability that the system is in state $\boldsymbol{x}$ after t transitions

Analysis of Markov Chains

Why does Gibbs sampling work? Or, why does its estimates converge to correct values in the limit?

Transition kernel k assigns a probability $k(\boldsymbol{x} \to \boldsymbol{x}')$ to every transition from a state $\boldsymbol{x}$ to another state $\boldsymbol{x}'$ in the Markov chain.



$\pi_t(\boldsymbol{x})$: probability that the system is in state $\boldsymbol{x}$ after t transitions

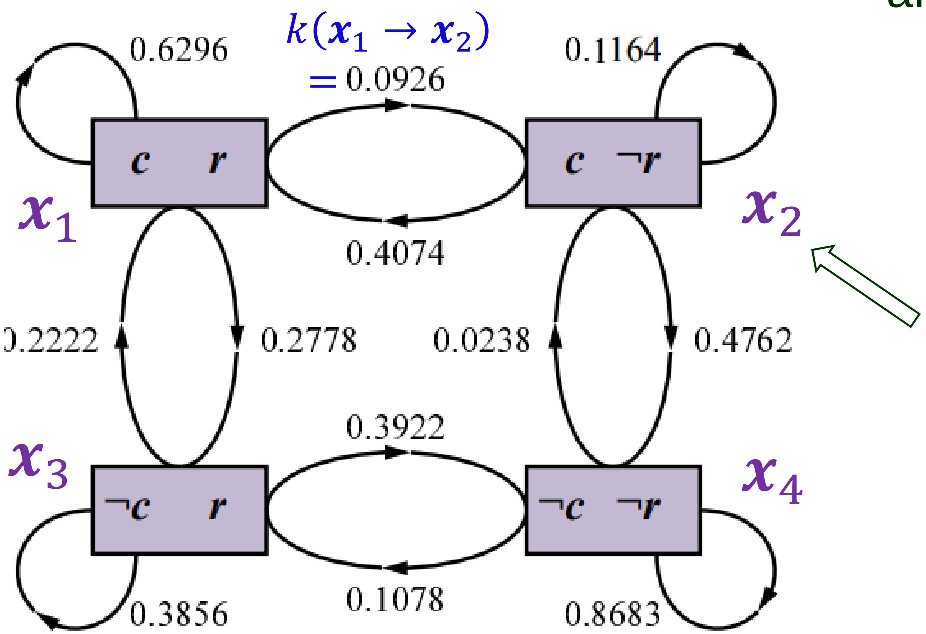
$$\pi_{t+1}(\boldsymbol{x}') = \sum_{\boldsymbol{x}} \pi_t(\boldsymbol{x}) k(\boldsymbol{x} \to \boldsymbol{x}')$$

Analysis of Markov Chains

Why does Gibbs sampling work? Or, why does its estimates converge to correct values in the limit?

Transition kernel k assigns a probability $k(\boldsymbol{x} \to \boldsymbol{x}')$ to every transition from a state $\boldsymbol{x}$ to another state $\boldsymbol{x}'$ in the Markov chain.

$\pi_t(\boldsymbol{x})$: probability that the system is in state $\boldsymbol{x}$ after t transitions

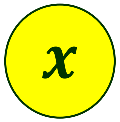


$$\pi_{t+1}(\boldsymbol{x}') = \sum_{\boldsymbol{x}} \pi_t(\boldsymbol{x}) k(\boldsymbol{x} \to \boldsymbol{x}')$$

$$\pi_{t+1}(\boldsymbol{x}_2) = 0.0926\,\pi_t(\boldsymbol{x}_1) + 0.1164\pi_t(\boldsymbol{x}_2) + 0.0238\,\pi_t(\boldsymbol{x}_4)$$

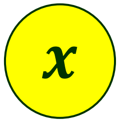
Stationary Distribution

The chain has reached its *stationary* distribution if $\pi_{t+1}(\boldsymbol{x}) = \pi_t(\boldsymbol{x})$ for all $\boldsymbol{x}$. We then call this stationary distribution π.



Stationary Distribution

The chain has reached its *stationary* distribution if $\pi_{t+1}(\boldsymbol{x}) = \pi_t(\boldsymbol{x})$ for all $\boldsymbol{x}$. We then call this stationary distribution π.



A probability distribution π over (the states of) the Markov chain is stationary if, for every state $\boldsymbol{x}'$,

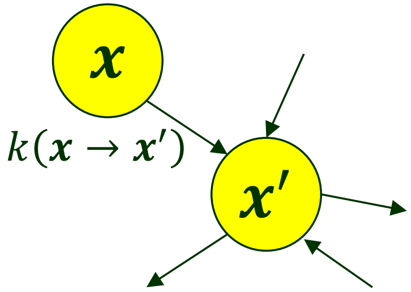
$$\pi(\boldsymbol{x}') = \sum_{\boldsymbol{x}} \pi(\boldsymbol{x}) k(\boldsymbol{x} \to \boldsymbol{x}')$$

Stationary Distribution

The chain has reached its *stationary* distribution if $\pi_{t+1}(\boldsymbol{x}) = \pi_t(\boldsymbol{x})$ for all $\boldsymbol{x}$. We then call this stationary distribution π.

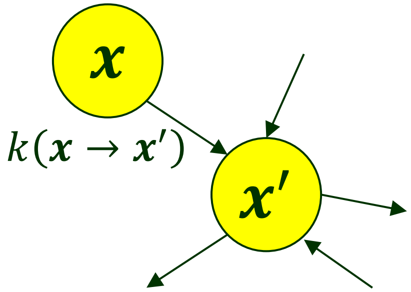
A probability distribution π over (the states of) the Markov chain is stationary if, for every state $\boldsymbol{x}'$,

$$\boxed{\pi(\boldsymbol{x}') = \sum_{\boldsymbol{x}} \pi(\boldsymbol{x}) k(\boldsymbol{x} \to \boldsymbol{x}')}$$



Stationary Distribution

The chain has reached its *stationary* distribution if $\pi_{t+1}(\boldsymbol{x}) = \pi_t(\boldsymbol{x})$ for all $\boldsymbol{x}$. We then call this stationary distribution π.



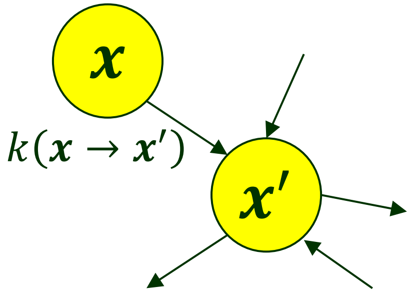
A probability distribution π over (the states of) the Markov chain is stationary if, for every state $\boldsymbol{x}'$,

$$\pi(\boldsymbol{x}') = \sum_{\boldsymbol{x}} \pi(\boldsymbol{x}) k(\boldsymbol{x} \rightarrow \boldsymbol{x}')$$

A stationary distribution remains unchanged in the Markov chain as time progresses.

Stationary Distribution

The chain has reached its *stationary* distribution if $\pi_{t+1}(\boldsymbol{x}) = \pi_t(\boldsymbol{x})$ for all $\boldsymbol{x}$. We then call this stationary distribution π.



A probability distribution π over (the states of) the Markov chain is stationary if, for every state $\boldsymbol{x}'$,

$$\pi(\boldsymbol{x}') = \sum_{\boldsymbol{x}} \pi(\boldsymbol{x}) k(\boldsymbol{x} \to \boldsymbol{x}')$$

A stationary distribution remains unchanged in the Markov chain as time progresses.

A transition kernel k is *ergodic* if

a) every state is reachable from every other state, and

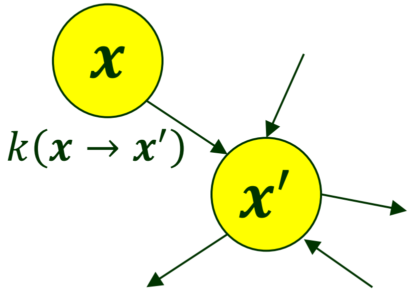
b) there exists *exactly one* stationary distribution π.

Achieving a Stationary Distribution

In a stationary distribution π, the expected “outflow” from each state is equal to the expected “inflow” from all the other states.

“population”

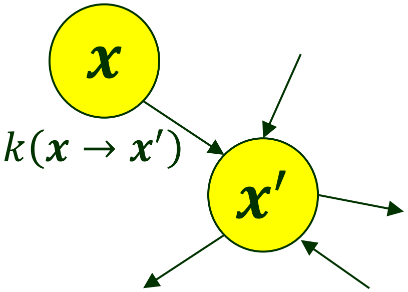
$$\underbrace{\pi(\boldsymbol{x}')}_{\text{expected “outflow”}} = \underbrace{\sum_{\boldsymbol{x}} \pi(\boldsymbol{x})k(\boldsymbol{x} \to \boldsymbol{x}')}_{\text{expected “inflow”}}$$



Achieving a Stationary Distribution

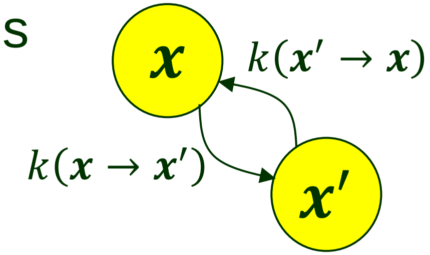
In a stationary distribution π, the expected “outflow” from each state is equal to the expected “inflow” from all the other states.

$$\overset{\text{“population”}}{\underbrace{\pi(\boldsymbol{x}')}_{\text{expected “outflow”}}} = \underbrace{\sum_{\boldsymbol{x}} \pi(\boldsymbol{x})k(\boldsymbol{x} \to \boldsymbol{x}')}_{\text{expected “inflow”}}$$



A *detailed balance* k with π is a distribution that satisfies

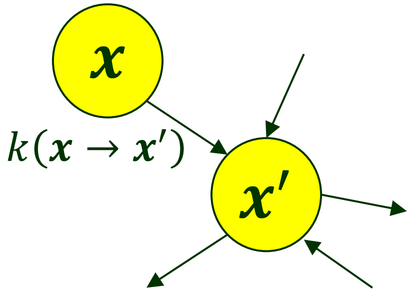
$\pi(\boldsymbol{x})k(\boldsymbol{x} \to \boldsymbol{x}') = \pi(\boldsymbol{x}')k(\boldsymbol{x}' \to \boldsymbol{x})$ for any states $\boldsymbol{x}, \boldsymbol{x}'$.



Achieving a Stationary Distribution

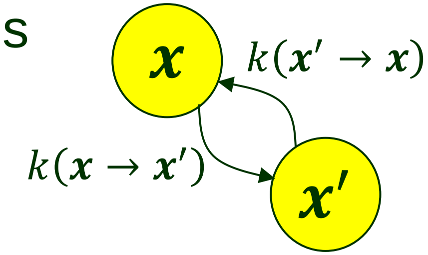
In a stationary distribution π, the expected “outflow” from each state is equal to the expected “inflow” from all the other states.

$$\underbrace{\overset{\text{“population”}}{\pi}(\boldsymbol{x}')}_{\text{expected “outflow”}} = \underbrace{\sum_{\boldsymbol{x}} \pi(\boldsymbol{x})k(\boldsymbol{x} \rightarrow \boldsymbol{x}')}_{\text{expected “inflow”}}$$



A *detailed balance* k with π is a distribution that satisfies

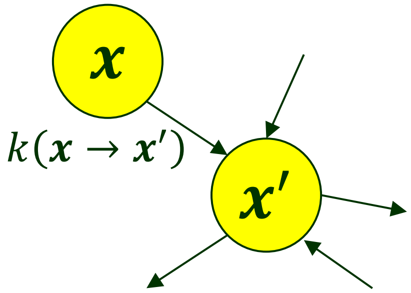
$$\pi(\boldsymbol{x})k(\boldsymbol{x} \rightarrow \boldsymbol{x}') = \pi(\boldsymbol{x}')k(\boldsymbol{x}' \rightarrow \boldsymbol{x}) \quad \text{for any states } \boldsymbol{x}, \boldsymbol{x}'.$$



The detailed balance k makes $\pi(\boldsymbol{x})$ stationary because

Achieving a Stationary Distribution

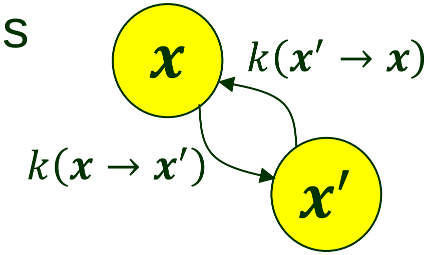
In a stationary distribution π, the expected “outflow” from each state is equal to the expected “inflow” from all the other states.



$$\overset{\text{“population”}}{\underbrace{\pi(\boldsymbol{x}')}_{\text{expected “outflow”}}} = \underbrace{\sum_{\boldsymbol{x}} \pi(\boldsymbol{x})k(\boldsymbol{x} \to \boldsymbol{x}')}_{\text{expected “inflow”}}$$

A *detailed balance* k with π is a distribution that satisfies

$$\pi(\boldsymbol{x})k(\boldsymbol{x} \to \boldsymbol{x}') = \pi(\boldsymbol{x}')k(\boldsymbol{x}' \to \boldsymbol{x}) \quad \text{for any states } \boldsymbol{x}, \boldsymbol{x}'.$$



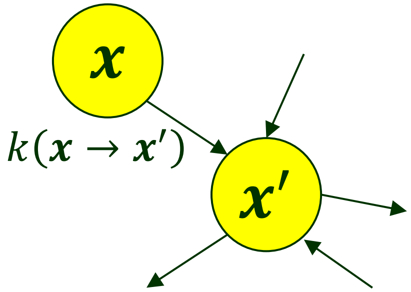
The detailed balance k makes $\pi(\boldsymbol{x})$ stationary because

$$\sum_{\boldsymbol{x}} \pi(\boldsymbol{x})k(\boldsymbol{x} \to \boldsymbol{x}') = \sum_{\boldsymbol{x}} \pi(\boldsymbol{x}')k(\boldsymbol{x}' \to \boldsymbol{x})$$

Achieving a Stationary Distribution

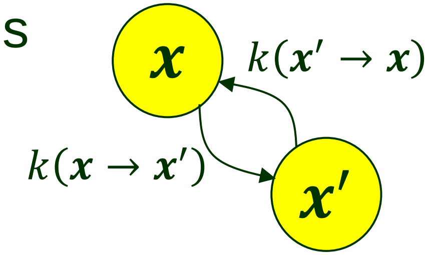
In a stationary distribution π, the expected "outflow" from each state is equal to the expected "inflow" from all the other states.

$$\underbrace{\overset{\text{“population”}}{\pi(\boldsymbol{x}')}}_{\text{expected “outflow”}} = \underbrace{\sum_{\boldsymbol{x}} \pi(\boldsymbol{x}) k(\boldsymbol{x} \to \boldsymbol{x}')}_{\text{expected “inflow”}}$$



A *detailed balance* k with π is a distribution that satisfies

$$\pi(\boldsymbol{x}) k(\boldsymbol{x} \to \boldsymbol{x}') = \pi(\boldsymbol{x}') k(\boldsymbol{x}' \to \boldsymbol{x}) \quad \text{for any states } \boldsymbol{x}, \boldsymbol{x}'.$$



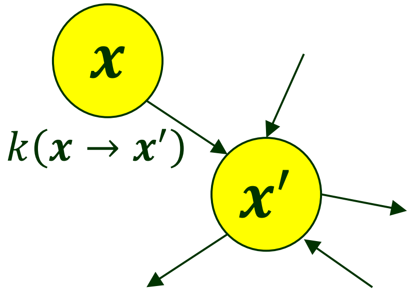
The detailed balance k makes $\pi(\boldsymbol{x})$ stationary because

$$\sum_{\boldsymbol{x}} \pi(\boldsymbol{x}) k(\boldsymbol{x} \to \boldsymbol{x}') = \sum_{\boldsymbol{x}} \pi(\boldsymbol{x}') k(\boldsymbol{x}' \to \boldsymbol{x})$$

$$= \pi(\boldsymbol{x}') \sum_{\boldsymbol{x}} k(\boldsymbol{x}' \to \boldsymbol{x})$$

Achieving a Stationary Distribution

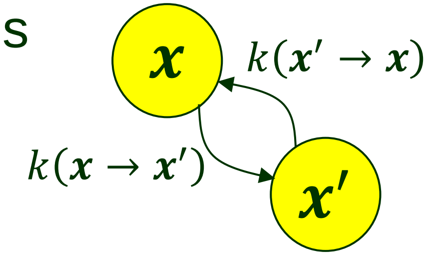
In a stationary distribution π, the expected “outflow” from each state is equal to the expected “inflow” from all the other states.



$$\underbrace{\overset{\text{“population”}}{\pi(\boldsymbol{x}')}}_{\text{expected “outflow”}} = \underbrace{\sum_{\boldsymbol{x}} \pi(\boldsymbol{x}) k(\boldsymbol{x} \to \boldsymbol{x}')}_{\text{expected “inflow”}}$$

A *detailed balance* k with π is a distribution that satisfies

$$\pi(\boldsymbol{x}) k(\boldsymbol{x} \to \boldsymbol{x}') = \pi(\boldsymbol{x}') k(\boldsymbol{x}' \to \boldsymbol{x}) \quad \text{for any states } \boldsymbol{x}, \boldsymbol{x}'.$$



The detailed balance k makes $\pi(\boldsymbol{x})$ stationary because

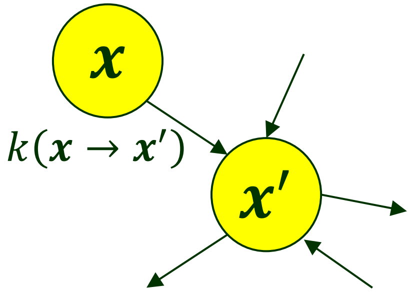
$$\sum_{\boldsymbol{x}} \pi(\boldsymbol{x}) k(\boldsymbol{x} \to \boldsymbol{x}') = \sum_{\boldsymbol{x}} \pi(\boldsymbol{x}') k(\boldsymbol{x}' \to \boldsymbol{x})$$

$$= \pi(\boldsymbol{x}') \boxed{\sum_{\boldsymbol{x}} k(\boldsymbol{x}' \to \boldsymbol{x})} = 1$$

Achieving a Stationary Distribution

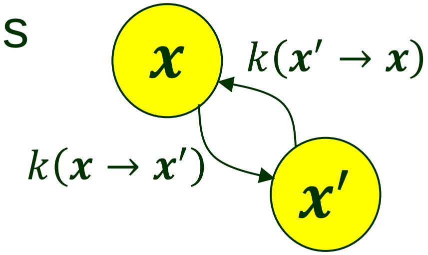
In a stationary distribution π, the expected "outflow" from each state is equal to the expected "inflow" from all the other states.

$$\underbrace{\overset{\text{"population"}}{\pi(\boldsymbol{x}')}}_{\text{expected "outflow"}} = \underbrace{\sum_{\boldsymbol{x}} \pi(\boldsymbol{x}) k(\boldsymbol{x} \to \boldsymbol{x}')}_{\text{expected "inflow"}}$$



A *detailed balance* k with π is a distribution that satisfies

$\pi(\boldsymbol{x}) k(\boldsymbol{x} \to \boldsymbol{x}') = \pi(\boldsymbol{x}') k(\boldsymbol{x}' \to \boldsymbol{x})$ for any states $\boldsymbol{x}, \boldsymbol{x}'$.



The detailed balance k makes $\pi(\boldsymbol{x})$ stationary because

$$\sum_{\boldsymbol{x}} \pi(\boldsymbol{x}) k(\boldsymbol{x} \to \boldsymbol{x}') = \sum_{\boldsymbol{x}} \pi(\boldsymbol{x}') k(\boldsymbol{x}' \to \boldsymbol{x})$$

$$= \pi(\boldsymbol{x}') \overset{= 1}{\boxed{\sum_{\boldsymbol{x}} k(\boldsymbol{x}' \to \boldsymbol{x})}} = \pi(\boldsymbol{x}')$$