Inferenza in Reti Bayesiane

Corso di Intelligenza Artificiale, Capitolo 14.4–5

Inferenza approssimata tramite MCMC

"Stato" della rete = assegnamento corrente a tutte le variabili

Genera lo stato successivo campionando una variabile dato il suo Markov blanket
Campiona ogni variabile a turno, mantenendo l'evidenza fissa

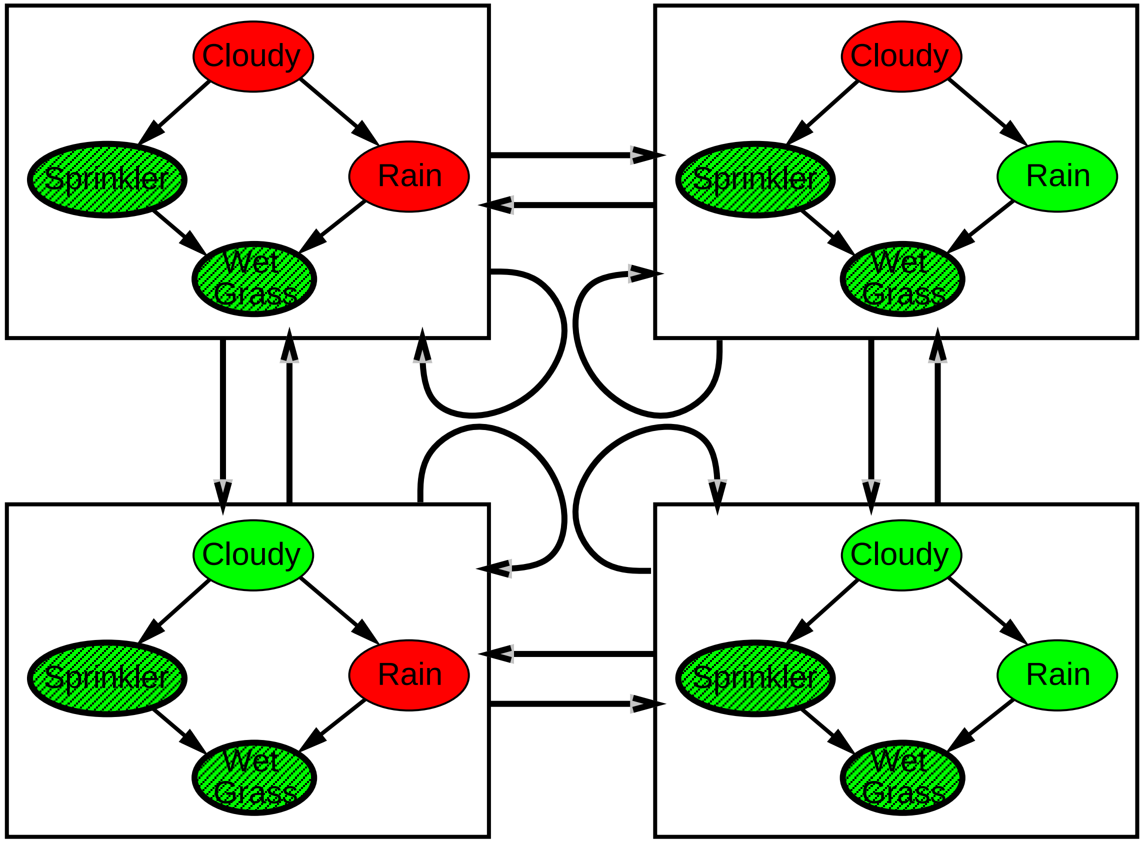
function MCMC-Ask(X, $\mathbf{e}$, bn, N) **returns** an estimate of $P(X|\mathbf{e})$
    local variables: $\mathbf{N}[X]$, a vector of counts over X, initially zero
                        $\mathbf{Z}$, the nonevidence variables in bn
                        $\mathbf{x}$, the current state of the network, initially copied from $\mathbf{e}$

    initialize $\mathbf{x}$ with random values for the variables in $\mathbf{Y}$
    for $j = 1$ to N **do**
        $\mathbf{N}[x] \leftarrow \mathbf{N}[x] + 1$ where x is the value of X in $\mathbf{x}$
        for each Z_i in $\mathbf{Z}$ **do**
                sample the value of Z_i in $\mathbf{x}$ from $\mathbf{P}(Z_i|MB(Z_i))$ given the values of $MB(Z_i)$ in $\mathbf{x}$
    return Normalize($\mathbf{N}[X]$)

Può anche scegliere una variabile da campionare a caso ogni volta

La catena di Markov (Markov chain)

Con $Sprinkler = true, WetGrass = true$, ci sono quattro stati:



"gironzolare" per un pò, fare la media di quello che si osserva

MCMC: esempio

Stimare $\mathbf{P}(Rain|Sprinkler = true, WetGrass = true)$

Campionare $Cloudy$ o $Rain$ dato il suo Markov blanket, ripetere.
Contare il numero di volte in cui $Rain$ è true e false nei campioni.

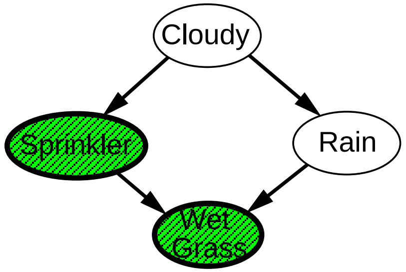
P.e., visita 100 stati
31 hanno $Rain = true$, 69 hanno $Rain = false$

$$\hat{\mathbf{P}}(Rain|Sprinkler = true, WetGrass = true)$$
$$= \textsc{Normalize}(\langle 31, 69 \rangle) = \langle 0.31, 0.69 \rangle$$

Teorema: la catena raggiunge la distribuzione stazionaria:
la frazione di tempo speso in ogni stato è esattamente proporzionale alla sua probabilità a posteriori

Markov blanket: campionamento

Il Markov blanket di $Cloudy$ è
$Sprinkler$ e $Rain$
Il Markov blanket di $Rain$ è
$Cloudy$, $Sprinkler$, e $WetGrass$



La probabilità dato il Markov blanket è calcolata come segue:
$$P(x_i'|MB(X_i)) = P(x_i'|Parents(X_i))\prod_{Z_j \in Children(X_i)} P(z_j|Parents(Z_j))$$

Facilmente implementabile in un sistema parallelo a scambio di messaggi

Principali problemi computazionali:
1) Difficoltà a riconoscere se la convergenza è avvenuta
2) Molto costoso se il Markov blanket è grande:
$P(X_i|MB(X_i))$ non cambierà molto (legge dei grandi numeri)

Analisi di MCMC: passi principali

Probabilità di transizione $q(\mathbf{x} \rightarrow \mathbf{x}')$

Probabilità di stazionamento $\pi_t(\mathbf{x})$ al tempo t

Condizione di equilibrio su π_t implica distribuzione stazionaria $\pi(\mathbf{x})$
Nota: la distribuzione stazionaria dipende dalla scelta di $q(\mathbf{x} \rightarrow \mathbf{x}')$

Equilibrio garantito dal detailed balance (condizione di equilibrio) su coppie di stati

Probabilità di transizione calcolata secondo Gibbs sampling:
campiona ogni variabile dati i valori correnti di tutte le altre
$\Rightarrow$ detailed balance con la vera probabilità a posteriori

Per reti Bayesiane, Gibbs sampling si riduce a
campionare condizionatamente alla Markov blanket di ogni variabile

Distribuzione stazionaria

$\pi_t(\mathbf{x})$ = probabilità di trovarsi nello stato $\mathbf{x}$ al tempo t
$\pi_{t+1}(\mathbf{x}')$ = probabilità di trovarsi nello stato $\mathbf{x}'$ al tempo $t+1$

π_{t+1} definibile in termini di π_t e $q(\mathbf{x} \rightarrow \mathbf{x}')$

$$\pi_{t+1}(\mathbf{x}') = \Sigma_{\mathbf{x}} \pi_t(\mathbf{x}) q(\mathbf{x} \rightarrow \mathbf{x}')$$

Distribuzione stazionaria: $\pi_t = \pi_{t+1} = \pi$

$$\pi(\mathbf{x}') = \Sigma_{\mathbf{x}} \pi(\mathbf{x}) q(\mathbf{x} \rightarrow \mathbf{x}') \qquad \text{per tutti gli stati } \mathbf{x}'$$

Se π esiste, essa è unica (e specifica per $q(\mathbf{x} \rightarrow \mathbf{x}')$)

All'equilibrio, "flusso in uscita" aspettato = "flusso in entrata" aspettato

Detailed balance

"flusso in uscita" $=$ "flusso in entrata" per ogni coppia di stati

$$\pi(\mathbf{x})q(\mathbf{x} \to \mathbf{x}') = \pi(\mathbf{x}')q(\mathbf{x}' \to \mathbf{x}) \qquad \text{per tutte le coppie } \mathbf{x},\ \mathbf{x}'$$

Detailed balance $\Rightarrow$ stazionarietà:

$$\begin{aligned}
\Sigma_{\mathbf{x}}\pi(\mathbf{x})q(\mathbf{x} \to \mathbf{x}') &= \Sigma_{\mathbf{x}}\pi(\mathbf{x}')q(\mathbf{x}' \to \mathbf{x}) \\
&= \pi(\mathbf{x}')\Sigma_{\mathbf{x}}q(\mathbf{x}' \to \mathbf{x}) \\
&= \pi(\mathbf{x}')
\end{aligned}$$

L'algoritmo MCMC è costruito in modo che la probabilità di transizione q sia in detailed balance con la probabilità desiderata π

Gibbs sampling

Campiona ogni variabile a turno, date **tutte le altre variabili**

Campionando X_i, denotiamo con $\bar{\mathbf{X}}_i$ tutte le altre variabili non di evidenza
Siano x_i e $\bar{\mathbf{x}}_i$ i valori correnti, e $\mathbf{e}$ l'evidenza fissata
La probabilità di transizione è data da

$$q(\mathbf{x} \to \mathbf{x}') = q(x_i, \bar{\mathbf{x}}_i \to x_i', \bar{\mathbf{x}}_i) = P(x_i' | \bar{\mathbf{x}}_i, \mathbf{e})$$

Questo conduce al detailed balance con la vera probabilità a posteriori $P(\mathbf{x}|\mathbf{e})$:

$$\begin{aligned}
\pi(\mathbf{x})q(\mathbf{x} \to \mathbf{x}') &= P(\mathbf{x}|\mathbf{e})P(x_i'|\bar{\mathbf{x}}_i, \mathbf{e}) = P(x_i, \bar{\mathbf{x}}_i|\mathbf{e})P(x_i'|\bar{\mathbf{x}}_i, \mathbf{e}) \\
&= P(x_i|\bar{\mathbf{x}}_i, \mathbf{e})P(\bar{\mathbf{x}}_i|\mathbf{e})P(x_i'|\bar{\mathbf{x}}_i, \mathbf{e}) \quad \text{(chain rule)} \\
&= P(x_i|\bar{\mathbf{x}}_i, \mathbf{e})P(x_i', \bar{\mathbf{x}}_i|\mathbf{e}) \qquad \text{(chain rule invertita)} \\
&= q(\mathbf{x}' \to \mathbf{x})\pi(\mathbf{x}') = \pi(\mathbf{x}')q(\mathbf{x}' \to \mathbf{x})
\end{aligned}$$

Riassunto

Inferenza esatta tramite l'eliminazione di variabile:
– polinomiale su polialberi, NP-hard in generale
– spazio $=$ tempo, dipendente dalla topologia

Inferenza approssimata tramite LW e MCMC:
– LW si comporta male quando c'è molta evidenza (soprattutto a "valle")
– LW, MCMC in genere indipendenti dalla topologia
– La convergenza può essere molto lenta per probabilità vicine a 1 o 0
– Possono trattare combinazioni arbitrarie di variabili discrete e continue