Prism Switches for MCMC

James Cussens and Nicos Angelopoulos

{jc,nicos}@cs.york.ac.uk

Department of Computer Science,

University of York, England.

MCMC Overview

Class of sampling algorithms that estimate a posterior distribution.

Markov chain

construct a chain of visited values, $M_1, M_2, \ldots, M_n$, by proposing M_* from M_i , with probability $q(M_*, M_i)$. Use prior knowledge, $p(M_*)$ and relative likelihood of the two values, $p(D|M_*)/p(D|M_i)$ to decide chain construction.

Monte Carlo

Use the chain to approximate the posterior $p(M|D)$.

Bayesian learning with MCMC

Given some data D and a class of statistical models $\mathcal{M}$ ($M \in \mathcal{M}$) that can express relations in the data, use MCMC to approximate normalisation factor in Bayes' theorem

$$p(M|D) = \frac{p(D|M)p(M)}{\sum_M p(D|M)p(M)}$$

$p(M)$ is the prior probability of each model

$p(D|M)$ the likelihood (how well the model fits the data)

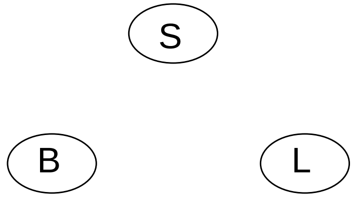
$p(M|D)$ the posterior

Example: Data

	smoker	bronchitis	l_cancer
person 1	y	y	n
person 2	y	n	n
person 3	y	y	y
person 4	n	y	n
person 5	n	n	n

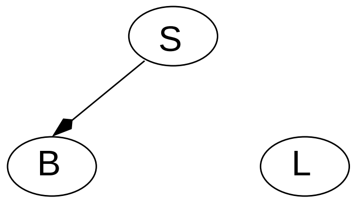
Example: Models

B_1



[b-[],l-[],s-[]]

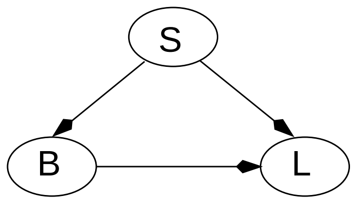
B_2



[b-[s],l-[],s-[]]

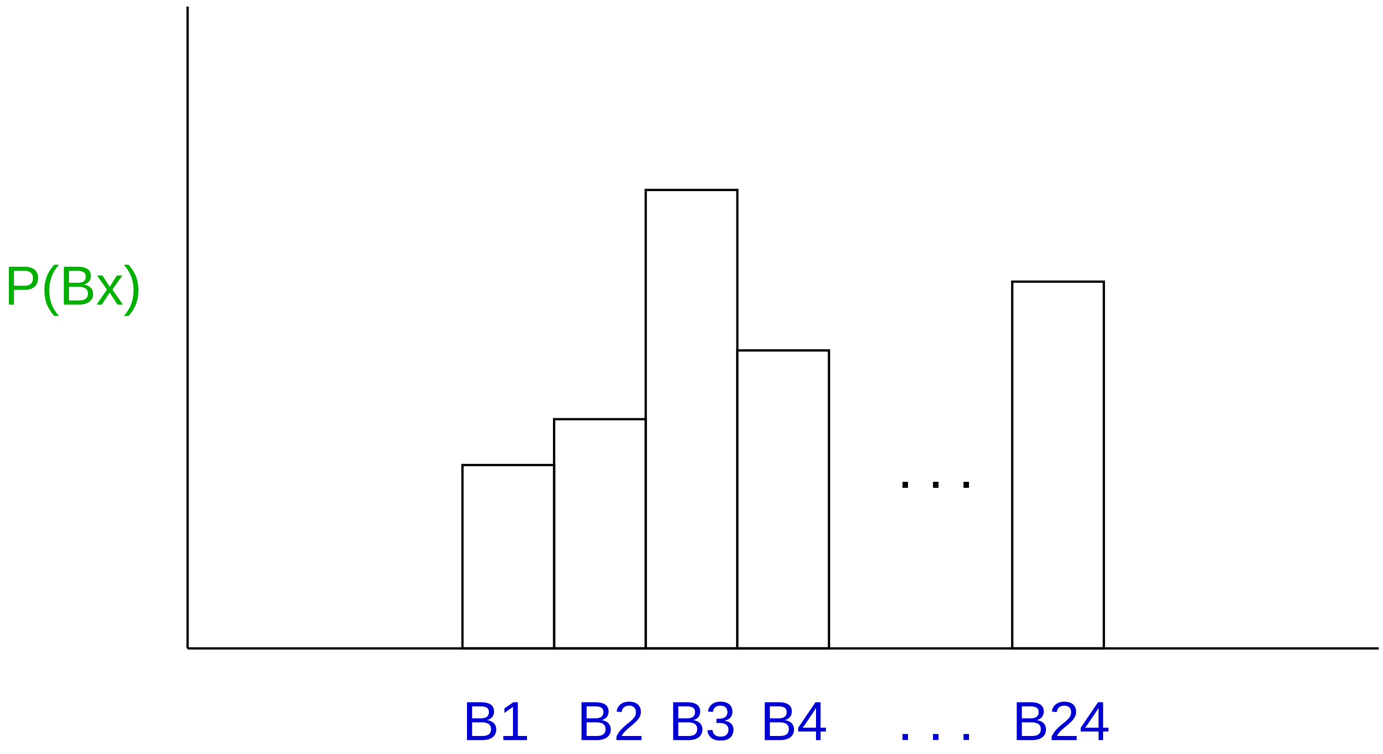
$\ldots$

B_{24}



[b-[s],l-[b,s],s-[]]

Example: Objective



$$\sum_{B_x} p(B_x) = 1$$

Metropolis-Hastings (M-H) MCMC

0. Set $i = 0$ and find M_0 using the prior.
1. From M_i produce a candidate model M_*. Let the probability of reaching M_* be $q(M_*, M_i)$.
2. Let

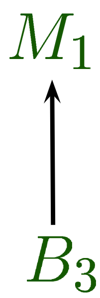
$$\alpha(M_i, M_*) = min\left\{\frac{q(M_*, M_i)P(D|M_*)P(M_*)}{q(M_i, M_*)P(D|M_i)P(M_i)}, 1\right\}$$

$$M_{i+1} = \begin{cases} M_* & \text{with probability } \alpha(M_i, M_*) \\ M_i & \text{with probability } 1 - \alpha(M_i, M_*) \end{cases}$$

3. If i reached limit then terminate, else set $i = i + 1$ and repeat from 1.

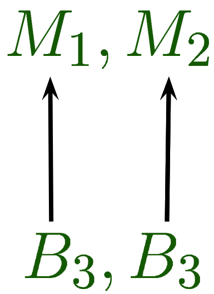
Example: MCMC

Markov Chain:



Example: MCMC

Markov Chain:



Example: MCMC

Markov Chain:

$$M_1, M_2, M_3, M_4, M_5, \ldots$$

$$\uparrow \quad \uparrow \quad \uparrow \quad \uparrow \quad \uparrow$$

$$B_3, B_3, B_{10}, B_3, B_{24}, \ldots$$

Example: MCMC

Markov Chain:

$$
\begin{array}{cccccc}
M_1, & M_2, & M_3, & M_4, & M_5, & \ldots \\
\uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \\
B_3, & B_3, & B_{10}, & B_3, & B_{24}, & \ldots
\end{array}
$$

Monte Carlo:

$$p(B_k) = \frac{\#(B_k)}{\sum_{B_x} \#(B_x)}$$

Independent sampler

Always sample from the prior: $q(M_*, M_i) = p(M_*)$. Thus,

$$\alpha(M_i, M_*) = min\left\{\frac{P(D|M_*)}{P(D|M_i)}, 1\right\}$$

Very simple to implement but only effective if prior is close to the posterior.

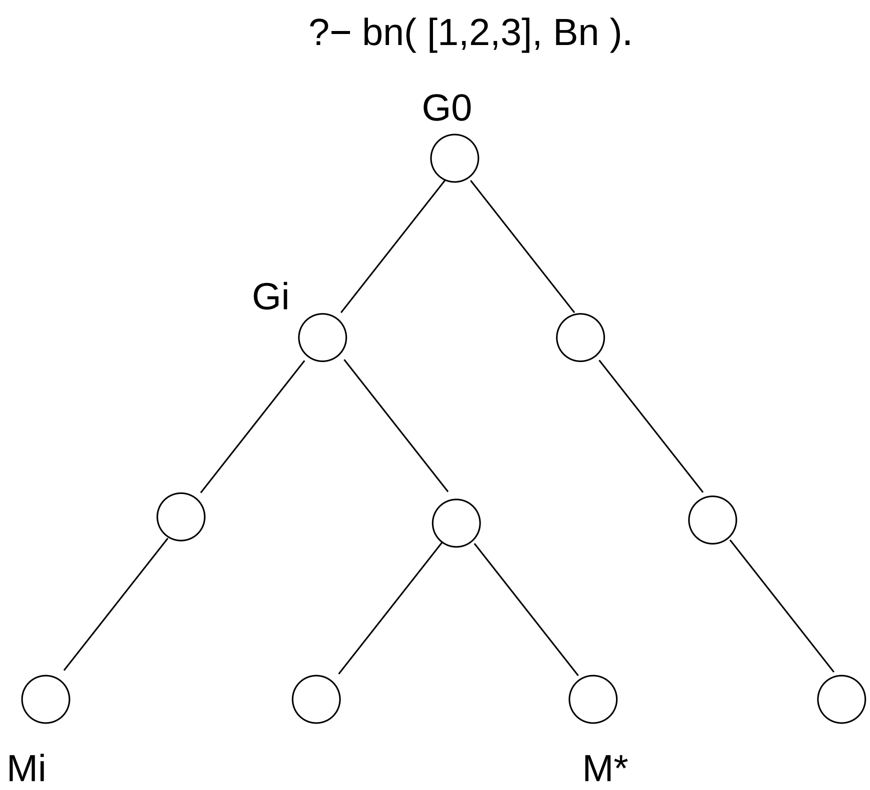
Single component M-H

If M_i can be decomposed to N components, use conditional sampling and a per component α.

M_i^{-k} model minus its kth component.

0. let $M_{c^1} = M_i$
1. for $k \in \{1, \ldots, N\}$
   - sample M_* with $p(M_* | M_{c^k}^{-k})$
   - $\alpha(M_{c^k}, M_*) = min\left\{\frac{P(D|M_*)}{P(D|M_{c^k})}, 1\right\}$
   - $M_{c^{k+1}} = \begin{cases} M_* & \text{with probability } \alpha(M_{c^k}, M_*) \\ M_{c^k} & \text{with probability } 1 - \alpha(M_{c^k}, M_*) \end{cases}$
2. $M_{i+1} = M_{c^{N+1}}$

Stochastic SLD trees



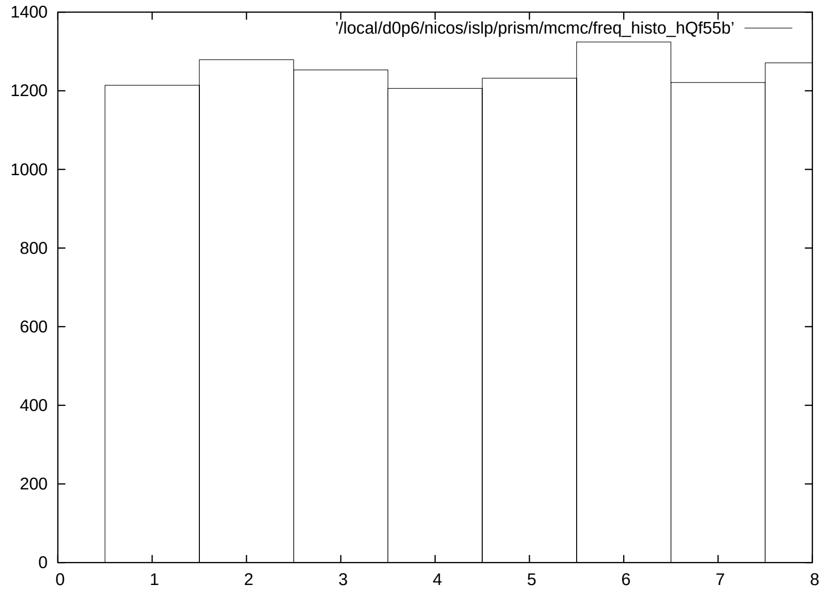
Statistical LP can provide rich language(s) for expressing $p(M_i), p(M_*)$ and disciplined ways for implementing alternative $q(M_*, M_i)$ kernels.

BN Prior

```
values( coin, [yes,no] ).
:- set_sw( coin, [0.5,0.5] ).

bn( Nodes, Bn ) :- bn( Nodes, [], Bn ).

bn( [], _RecPar, [] ).
bn( [H|T], RecPar, [H-HPar|BnRec] ) :-
   append( RecPar, [H], NxPar ),
   bn( T, NxPar, BnRec ),
   select_parents( RecPar, H, 1, HPar ).

select_parents( [], _Ch, _N, [] ).
select_parents( [H|T], Ch, N, Pa ) :-
   msw( coin, Resp ),
   include_element( Resp, H, Pa, TPa ),
   NxN is N + 1,
   select_parents( T, Ch, NxN, TPa ).

include_element( yes, H, [H|TPa], TPa ).
include_element( no, _H, TPa, TPa ).
```

Sampling from the prior

10000 Samples (1/8 = 1250)



?- bn([1,2,3], X).

'[1-[],2-[],3-[]]-1214 '
'[1-[],2-[],3-[1]]-1279 '
'[1-[],2-[],3-[1,2]]-1253 '
'[1-[],2-[],3-[2]]-1206 '
'[1-[],2-[1],3-[]]-1232 '
'[1-[],2-[1],3-[1]]-1324 '
'[1-[],2-[1],3-[1,2]]-1221'
'[1-[],2-[1],3-[2]]-1271 '

Independent sampler experiments

Used code written by James Cussens to compute likelihood of BN structure given some data (BN parameters are integrated over).

Built loop that samples, computes likelihood, and chooses next model for chain.

Independent sampler example output

```
c([1-[],2-[],3-[1]]).
b([1-[],2-[1],3-[1,2]]).
rat(12.168987909287049)-rnd(0.847462734036231З).
c([1-[],2-[1],3-[1,2]]).
b([1-[],2-[],3-[2]]).
rat(7.22568572871225 4E-13)-rnd(0.2650961677396184).
c([1-[],2-[1],3-[1,2]]).
b([1-[],2-[],3-[2]]).
rat(7.225685728712254E-13)-rnd(0.031236445152826864).
c([1-[],2-[1],3-[1,2]]).
b([1-[],2-[1],3-[1]]).
rat(3.3704863109554304)-rnd(0.43330278240268494).
c([1-[],2-[1],3-[1]]).
b([1-[],2-[1],3-[2]]).
rat(8.792928226900739E-12)-rnd(0.4581041305393969).
c([1-[],2-[1],3-[1]]).
b([1-[],2-[],3-[]]).
rat(4.038590350518264E-14)-rnd(0.6152324293678713).
c([1-[],2-[1],3-[1]]).
b([1-[],2-[1],3-[1]]).
```

msw for conditional sampling

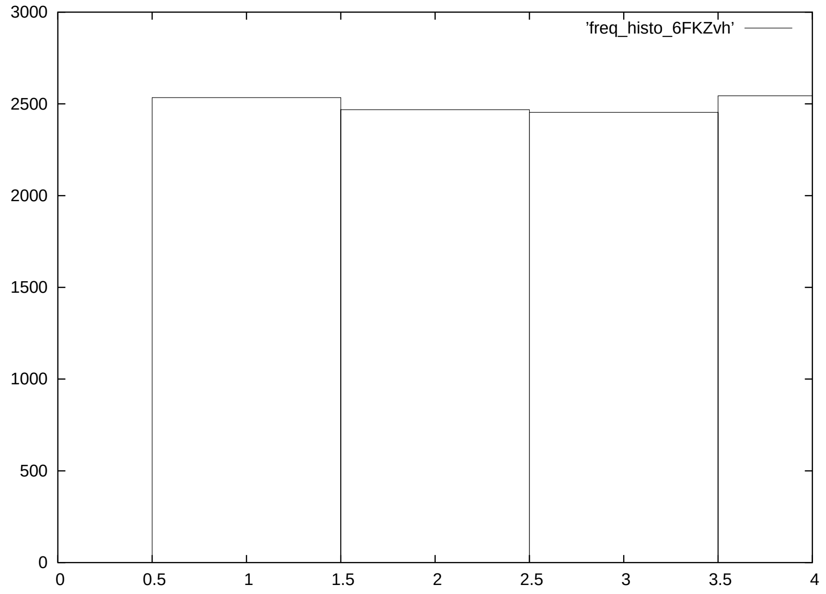
Need ability to sample with
e.g. $bn([1,2,3],[1-[],X,3-[1]])$.

Implemented a backtrackable version of msw.

For switch with N values, predicate succeeds N times.
On backtracking, the selected values so far are removed.
Probabilistically choose among remaining values.

Conditional sampling from prior

10000 Samples (1/4 = 2500)



?- bn([1,2,3], [1-[],X,3-[1]]).

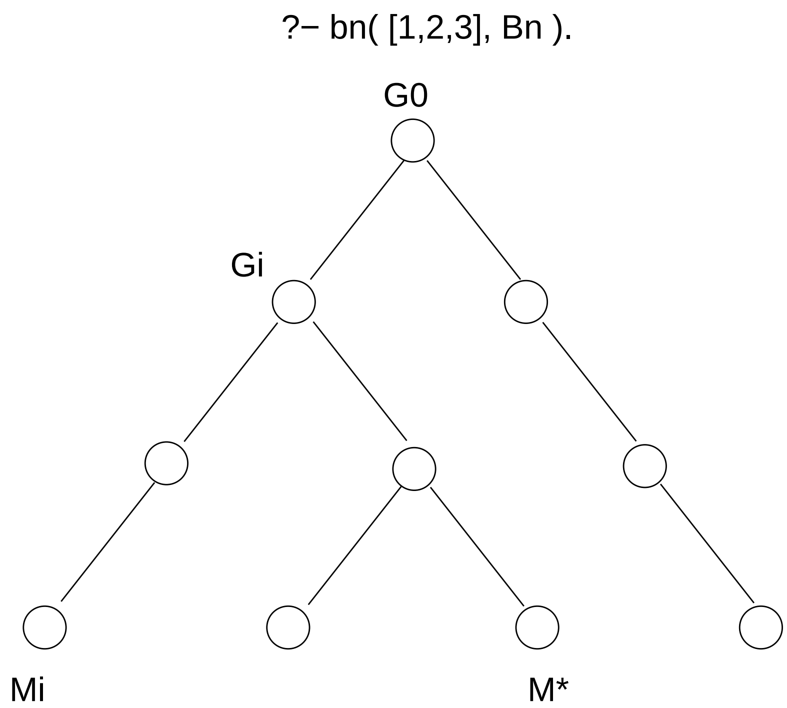
'[1-[],2-[],3-[]]-2534 ' '[1-[],2-[],3-[1]]-2468 '

'[1-[],2-[],3-[1,2]]-2454 ' '[1-[],2-[],3-[2]]-2544 '

Single component M-H example output

```
...
c([1-[],2-[],3-[1]]).
     r(1.0)-s(1-[]).
     r(41.01540716634144)-s(2-[1]).
     r(1.0)-s(3-[1]).
c([1-[],2-[1],3-[1]]).
     r(1.0)-s(1-[]).
     r(1.0)-s(2-[1]).
     r(8.79292822690739E-12)-s(3-[1]).
c([1-[],2-[1],3-[1]]).
     r(1.0)-s(1-[]).
     r(1.0)-s(2-[1]).
     r(8.79292822690739E-12)-s(3-[1]).
c([1-[],2-[1],3-[1]]).
     r(1.0)-s(1-[]).
     r(0.02438108186862940З)-s(2-[1]).
     r(8.79292822690739E-12)-s(3-[1]).
c([1-[],2-[1],3-[1]]).
     r(1.0)-s(1-[]).
     r(1.0)-s(2-[1]).
     r(1.656444276049385 8E-12)-s(3-[1]).
```

Proposals revisited



From M_i identify G_i then sample forward to $M_\star$.
$q(M_i, M_\star)$ is the probability of proposing $M_\star$ when M_i is the current model.

MCMC Scheme

10. sample initial goal deriving M_1.
20. backtrack to arbitrary G_i and sample M_*
    - do not destroy choice points $G_{i+1} \dots G_{l_i}$
    - add $G^*_{i+1} \dots G^*_{l_*}$ as $G_{l_i+1} \dots G_{l_i+l_*-i}$
30. set M_{i+1} to either M_i or M_* ; reclaim memory for
    - $G_{i+1} \dots G_{l_i}$, if M_{i+1} = M_i
    - or $G_{l_i+1} \dots G_{l_i+l_*-i}$, if M_{i+1} = M_*
40. unless termination conditions reached, go to 20

The challenge

The use of efficient techniques for implementing generic and user-specific proposals over stochastic SLD trees.

First impressions

For MCMC simulations Prism's switch can be used. Three possible extensions:

(a) allow shorthand msw(+Vals,+Prbs,-Val)

(b) P(msw(+Vals,+Prbs,+Val)) = 1, when $Val \in Vals$

(c) backtrackable version(s), bk_msw/2,3