Efficient Sampling for Bayesian Inference of Conjunctive Bayesian Networks

for cancer progression modeling

Thomas Sakoparnig

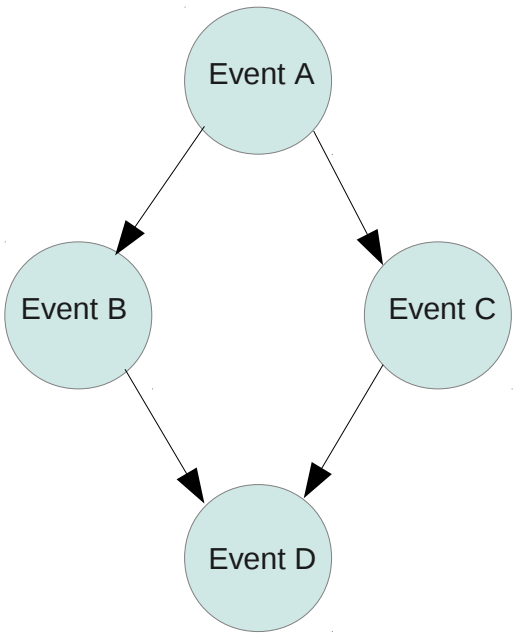
Computational Biology Group, ETH Zurich - Biosystems Science and Engineering

Sept 9th 2012

Characteristics of Cancer progression

- evolutionary process
- accumulation of advantageous mutations
- recurrent mutations
- mutations depend on presence of other mutations
- order in general unknown (dependency structure)

Temporal order: CBN



Mutation pattern: (cross-sectional data)

A	B	C	D
1	1	0	0
1	1	1	0
1	0	0	0
1	0	1	0
1	1	1	1
.	.	.	.

Problem at hand

- quantify uncertainty of structure estimation
- sampling in Bayesian network structure space
- local optima

Bayesian inference of CBNs I

dependency structure: $\prec$
mutation probabilities: θ_k

$$\Pr(Z \mid \prec, \theta) = \prod_{\{k\,:\,Z_k=1\}} \theta_k \prod_{k \in \mathrm{Exit}(Z)} (1-\theta_k) \tag{1}$$

if Z is compatible with $\prec$, and zero otherwise.

$$\Pr(X \mid Z, \varepsilon) = \varepsilon^{d(X,Z)}(1-\varepsilon)^{n-d(X,Z)} \tag{2}$$

where $d(X, Z)$ is the Hamming distance between X and Z and ε is the error probability.

Bayesian inference of CBNs II

The marginal likelihood of the m measured genotypes, denoted D, can then be written as

$$\Pr(D \mid \prec, \theta, \varepsilon) = \prod_{X \in D} \sum_{Z} \Pr(X \mid Z, \varepsilon) \Pr(Z \mid \prec, \theta) \tag{3}$$

$$\Pr(\prec, \theta, \varepsilon \mid D) \propto \prod_{X \in D} \sum_{Z} \left[\Pr(X \mid Z, \varepsilon) \times \right.$$
$$\left. \times \Pr(Z \mid \prec, \theta) \right] \prod_{k=1}^{n} \Pr(\theta_k) \Pr(\prec) \Pr(\varepsilon) \tag{4}$$

Priors:

$\Pr(\prec) = 1$

$\Pr(\theta_k) = 10^{-5}$

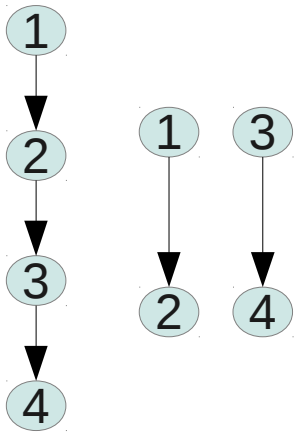
$\Pr(\varepsilon) = \mathrm{Beta}(5, 30)$

Hybrid sampler

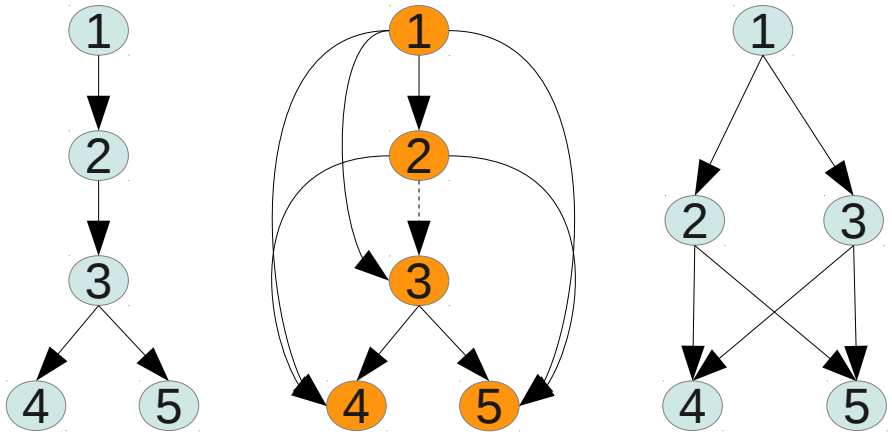
- random scan Metropolis-Hastings within Gibbs sampler
- eight move types (six structure and two continuous parameter moves)
- asymmetric move types have disjoint neighborhoods
- structure moves are complemented by theta moves

Structure moves

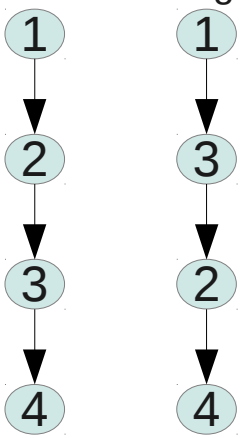
New/Delete cover relation



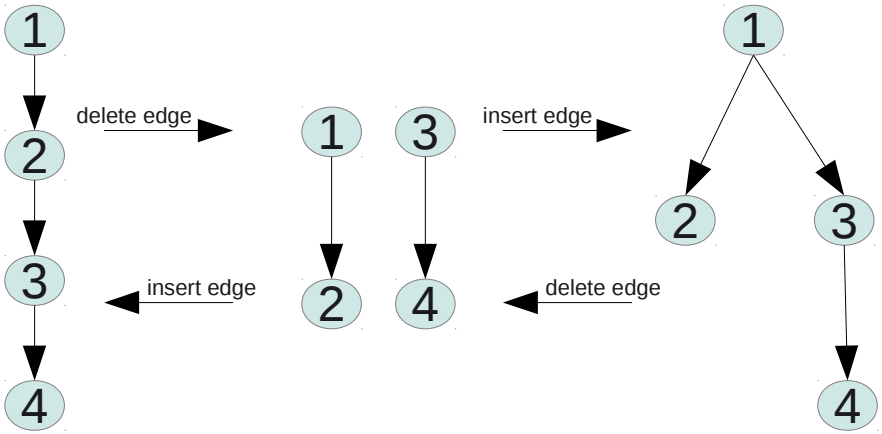
New/Delete transitive closure relation



Event exchange



Reincarnation



Remaining moves and convergence calling

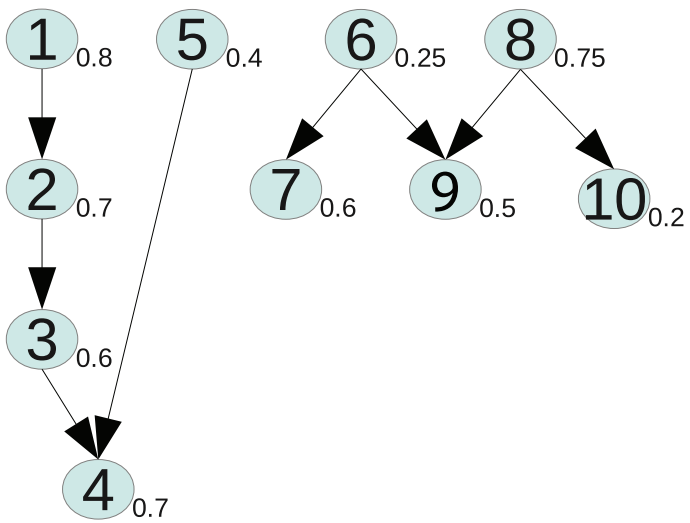
Other move types

- relocate theta (from Uniform(0,1) proposal)
- relocate epsilon (from Beta(2,20) proposal)

Convergence calling

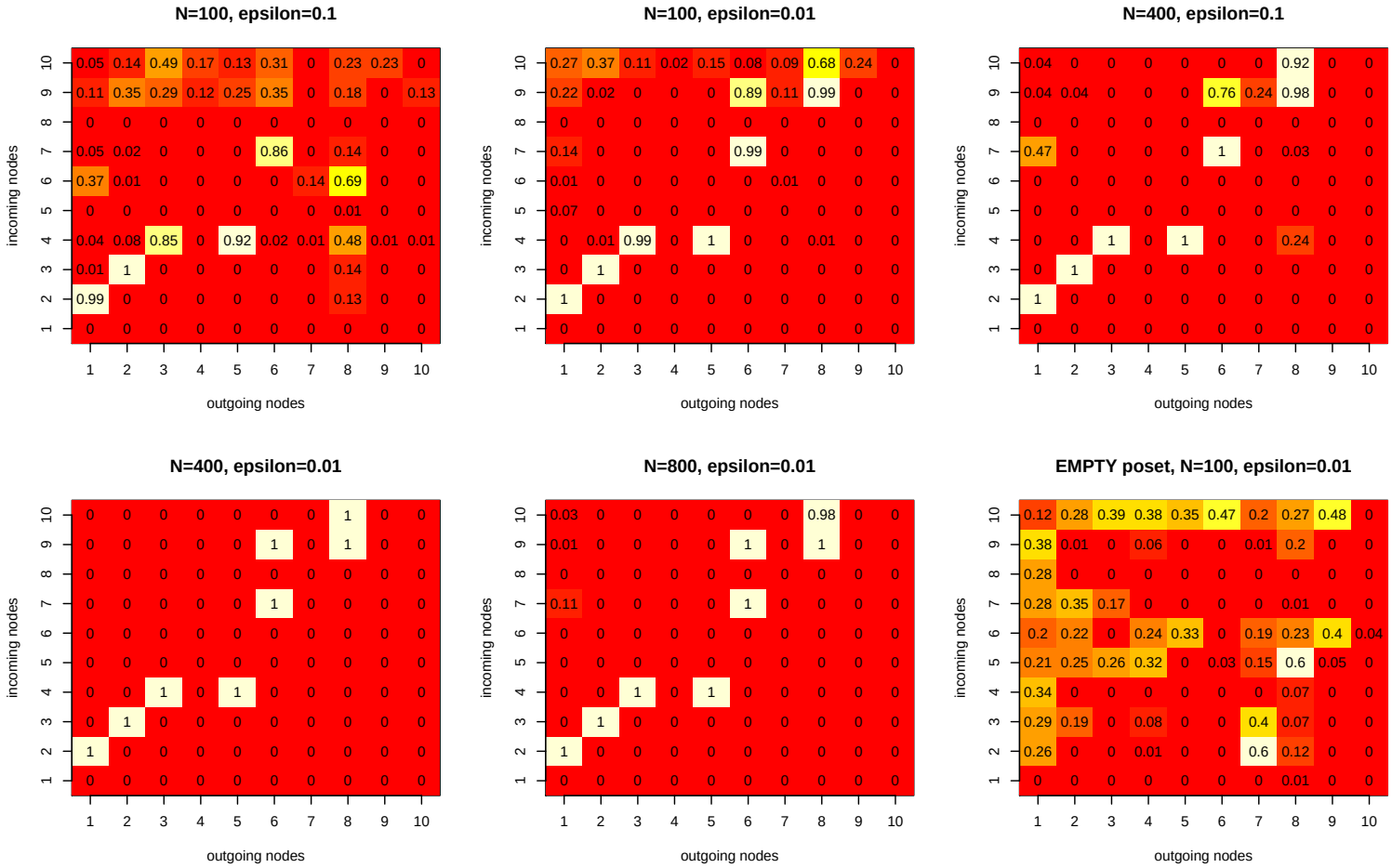
- multiple chains
- comparison of intra- and inter-chain variance

Simulation study I



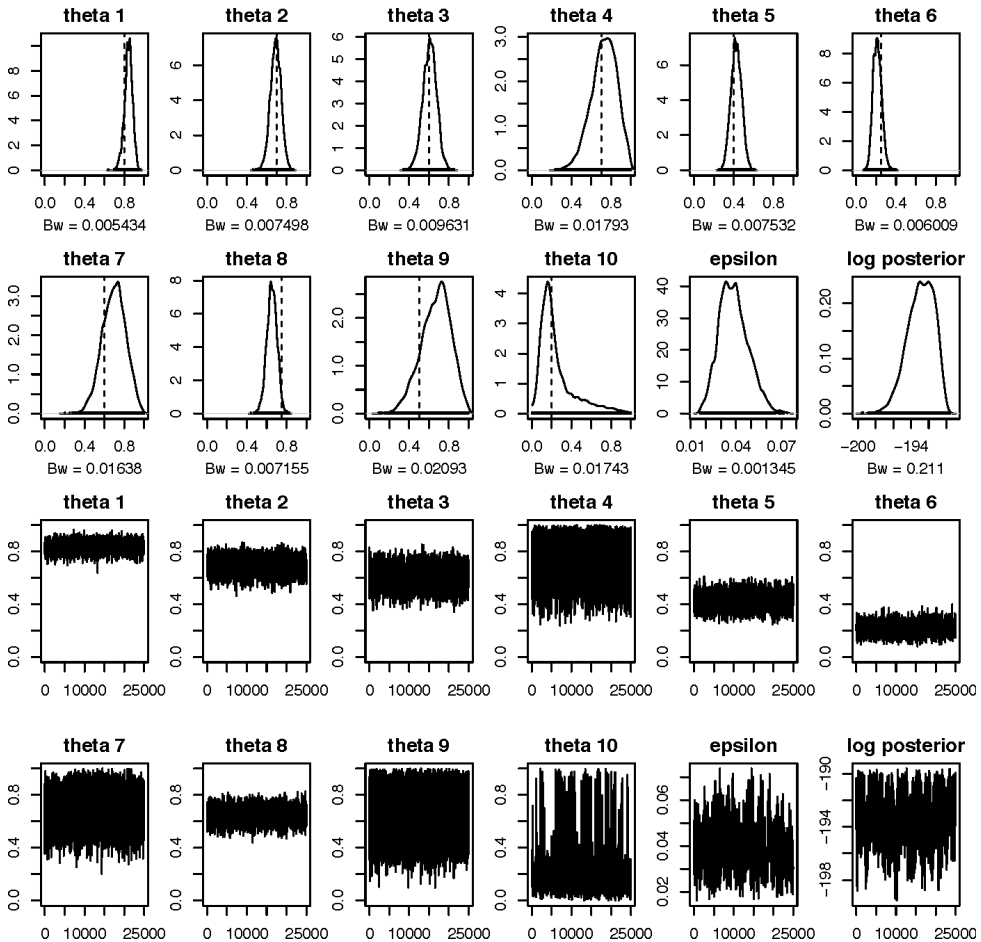
- four chains
- 25.000 samples per chain
- keeping every 20th samples
- convergence in one to five rounds

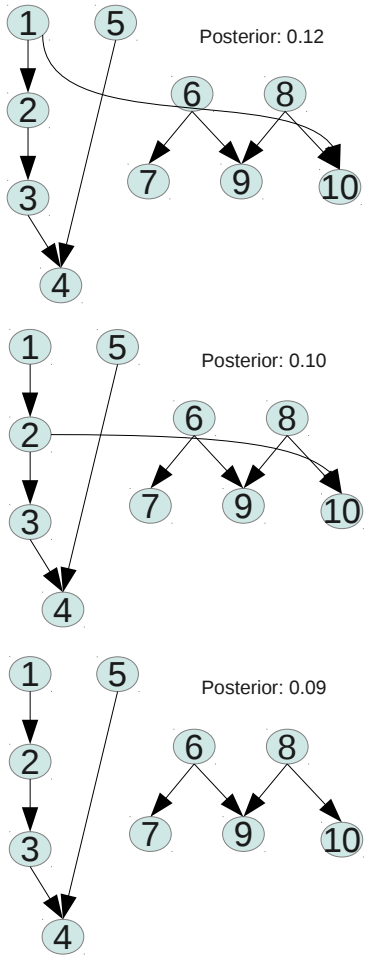
Simulation study II



Simulation study III

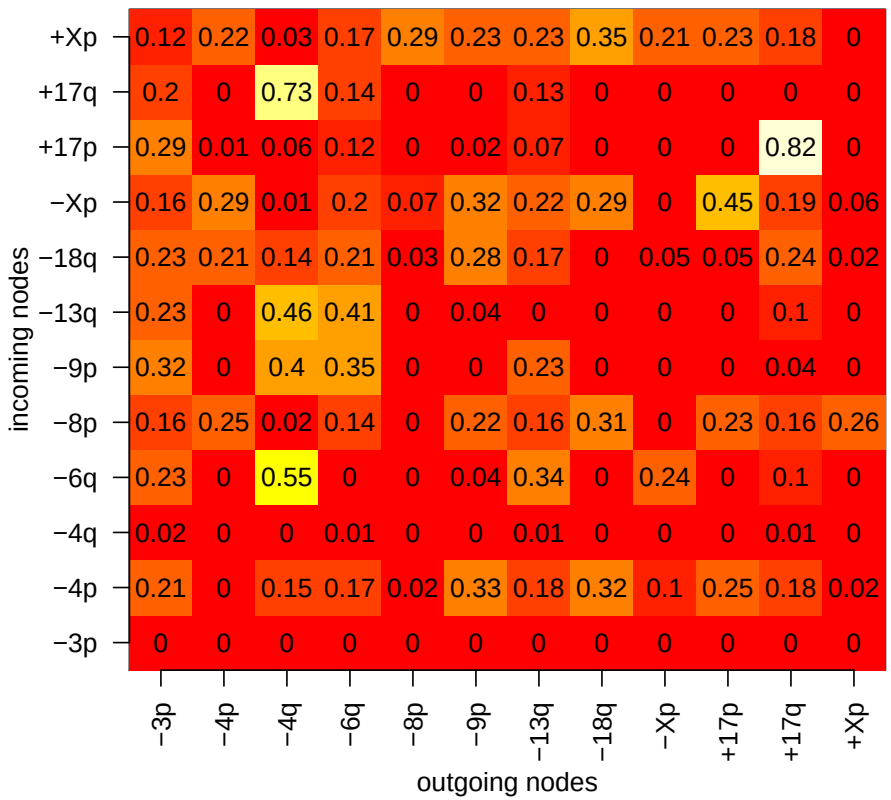
Details of $N = 100$ and $\varepsilon = 0.01$:





RCC CGH data analysis

- 251 renal cell carcinomas (RCC) (Jiang et al. 2000, Cancer Res.)
- comparative genome hybridization (CGH)
- only three dependencies have posterior probability > 0.5



Discussion

- only up to 15 loci
- convergence calling for the structure
- move types may be useful for other Bayesian networks (with unambiguous edge directions)
- usage in predictive modeling

Acknowledgement

Thanks to
Niko Beerenwinkel and the whole Beerenwinkel group