

# Markov chain Monte Carlo over Model Structures

Nicos Angelopoulos<sup>†</sup> and James Cussens<sup>‡</sup>

<sup>†</sup>Netherlands Cancer Institute

<sup>‡</sup>University of York, UK

**Introduction** MCMCMS implements a generic framework for constructing Markov chains. It can be used to perform statistical machine learning in a Bayesian framework. It presents a modular, high level approach to performing MCMC simulations over statistical models that can explain observed data [1]. The two main benefits of MCMCMS are its prior-centric construction of the chain, that does away with the need for an explicit proposal, and the fact that complex, crisp and probabilistic, information can be encoded in the prior by use of a high level language. MCMCMS priors can be encoded as SLPs (Stochastic logic programs, [4]) or as DLPs (Distributional). DLPs extend SLPs by allowing on-the-fly computations of the probabilistic labels.

**Implementation** The system is implemented in Prolog and it is open source. Through extensive use of term expansion the labelled clauses are transformed to Prolog ones with additional arguments. Inserted calls to the random number generator allow probabilistic clause selection, while stochastic backtracking is via an extra clausal argument holding a path structure. MCMCMS runs on the current Yap (6.2.0) and Swi (5.11.22) Prologs.

**Applications** It has been tested on two model spaces: C&ARTs and Bayes nets. In addition to a number of synthetic and literature test sets, it has been used to analyse chemical binding data [5] (C&RTs) and DNA assay data [3] (BNs). [2] presented an extensive comparison on effective priors for BNs.

**Availability:** <http://scibsfh.bch.ed.ac.uk/~nicos/sware/dlp/mcmcms>

## References

- [1] N. Angelopoulos and J. Cussens. MCMC using tree-based priors on model structure. In *UAI'01*, pages 16–23, 2001.
- [2] N. Angelopoulos and J. Cussens. Bayesian learning of bayesian networks with informative priors. *Annals of Math. and AI*, 54(1-3):53–98, 2008.
- [3] N. Angelopoulos and L. Wessels. Effective priors over model structures applied to DNA binding assay data. In *ProBioMed'11*, Bled, Slovenia, 2011.
- [4] J. Cussens. Stochastic logic programs: Sampling, inference and applications. In *UAI'2000*, pages 115–122, 2000.
- [5] N. Angelopoulos et al. Bayesian ligand discovery from high dimensional descriptor data. *ACS J. of Chemi Inf and Modeling*, 49(6):1547–1557, 2009.