An Experimental Study of Prior Dependence in Bayesian Network Structure Learning

Alvaro H. C. Correia
Cassio P. de Campos
Linda C. van der Gaag
Department of Information and Computing Sciences - Utrecht University

Prior Dependence in BDeu Structure Learning

In the Bayesian Dirichlet equivalent uniform (BDeu) the prior is expressed by α, the Equivalent Sample Size (ESS). What is its influence?

$$\mathrm{BDeu}(G,\alpha) = \prod_{i=1}^{n}\prod_{j=1}^{q_i}\frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij}+N_{ij})}\prod_{k=1}^{r_i}\frac{\Gamma(\alpha_{ijk}+N_{ijk})}{\Gamma(\alpha_{ijk})},$$

In a Bayesian framework, the prior should become less relevant in determining the final structure as we gather more data. Does that hold for the BDeu?

Experimental Analysis

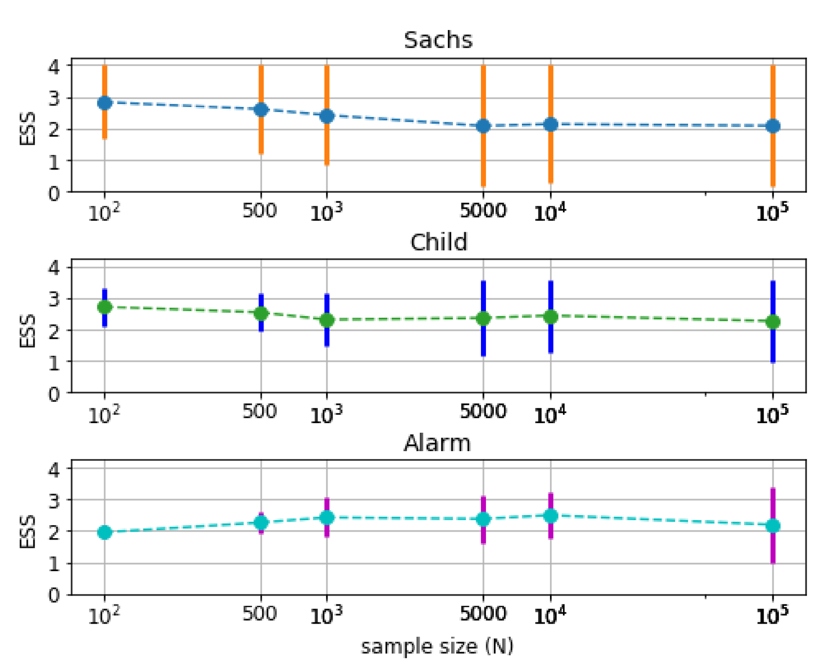
We analyse the influence of the ESS from two different angles:

Graph Complexity

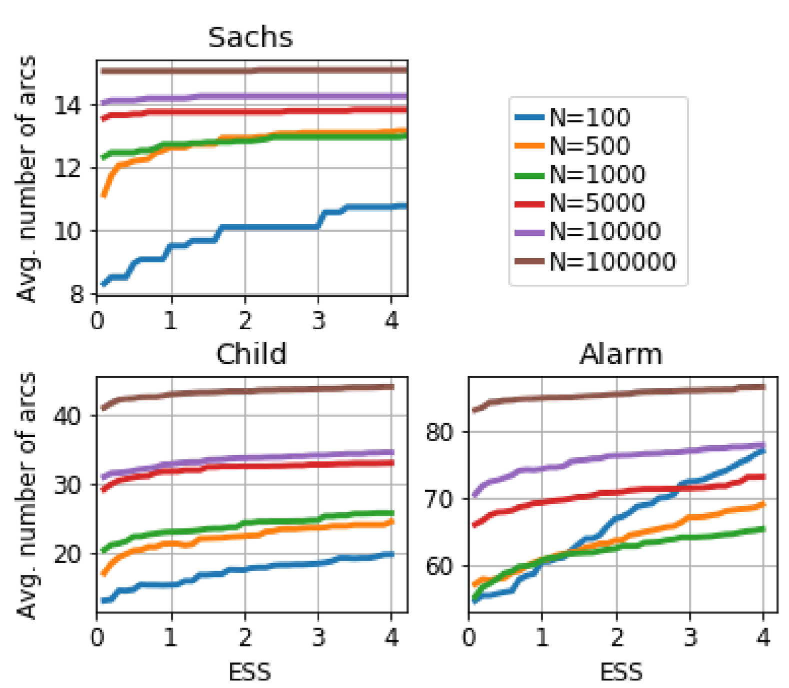
How variations in the ESS affect the number of arcs in the structure.

Robustness

What conditions are required for prior-independence.



Robust Interval for three well known Bayesian networks—Sachs, Child and Alarm—in function of the sample size (N).



Average number of arcs as a function of the ESS for different N. Scale of the graphs varies.

Utrecht University

Available data is likely insufficient to avoid prior dependence in BDeu-based Structure Learning.

Take a picture to download the full paper

Structure Learning

To learn the structure of a Bayesian network from data by searching for the graph G that maximises the posterior $P(G|D)$.

$$P(G)\int P(D|G,\Theta_G)\boxed{P(\Theta_G|G)}\,d\Theta_G$$

What is the influence of this prior?

Robustness Interval

The largest range of ESS values for which all obtained optimal structures (for each ESS) are Markov equivalent.

$$\mathrm{RI} := \underset{[\alpha_1,\alpha_2]}{\arg\max}\{|\alpha_2-\alpha_1| : G^*(\alpha') \equiv G^*(\alpha''), \forall \alpha',\alpha'' \in [\alpha_1,\alpha_2]\},$$

where $G^*(\alpha) = \arg\max_G \mathrm{BDeu}(G,\alpha)$ is the optimal graph for a given ESS, and $\equiv$ denotes Markov equivalence.

RI for UCI datasets

N and n are the number of samples and variables, RIo the average RI of 10 orderings, and RIf the RI with no order constraints.

Dataset	n	N	RIo	RIf
car	7	1728	(0.1, 4.0)	(0.4, 4.0)
glass	8	214	(1.3, 2.3)	(0.3, 4.0)
spambase	8	4601	(1.2, 4.0)	(1.7, 4.0)
diabetes	9	768	(0.2, 1.7)	(1.6, 4.0)
nursery	9	12960	(1.4, 2.9)	(1.4, 4.0)
breast-cancer	10	286	(1.9, 4.0)	(2.2, 4.0)
tic-tac-toe	10	958	(1.8, 2.1)	(1.7, 2.2)
cmc	10	1473	(1.7, 2.9)	(0.8, 2.8)
heart-h	12	294	(0.8, 1.6)	(2.2, 2.9)
vowel	14	990	(0.6, 1.8)	(1.9, 4.0)
zoo	17	101	(0.6, 1.3)	(0.9, 2.1)
vote	17	435	(0.8, 1.8)	(2.3, 3.1)
segment	17	2310	(1.5, 2.9)	(2.3, 4.0)
primary-tumor	18	339	(1.1, 1.5)	(3.1, 3.5)
vehicle	19	846	(0.9, 1.7)	(3.3, 4.0)

ISIPTA 2019