Round Table

Madigan, Raftery, 1994: Occam's window

- Reject models whose posterior probability given the data is very low. With uniform prior, they are rejecting models with low likelihood.
- Reject complex models whose posterior probability is not substantially better than a simpler model (the approximation step).

Even, with a simple s and a BDEu score, the likelihood approach makes sense: in general there are many networks that could explain the data. Selecting one seems to be *arbitrary*.

14 independent variables with two values (uniform).
A sample of 4 is obtained. Evolution of the score as a function of
the number of parents with $s = 2$.

Parents	Score
0	-2.9957322735539904
1	-1.8562979903657164
2	-1.67397643357172
3	-1.5488132906176393
4	-1.4733057381094796
5	-1.431414796400329
6	-1.4092838793445717
7	-2.087163587773726
8	-2.0833250417062317

Determining the prior distribution

- It is not simple how to determine the prior distribution.

- Even if we accept Dirichlet prior distributions in the multinomial case, it is not simple how to determine the weights.

- The problem in Bayesian networks is that increasing the number of parents, the weights should decrease (reduce the entropy of the distributions). How important is this reductions?

  - In K2 score there is no reduction (non-realistic).
  - In BDEu the weights are inversely proportional to the number of conditioning distributions. This is a strong assumption with important consequences and wrong behaviour if these assumptions are not satisfied.

Bayesian Solution and Average Case Complexity

- When Bayesians are not sure about an hypothesis, they average between different ones (averaging different sample sizes or averaging between the networks with score above a threshold).

- There is an analogy between Bayesian procedures and average case complexity versus likelihood based methods and worst case complexity.

- At the beginning average case complexity seems to be more appealing from an intuitive point of view, but in practice worst case complexity is more used in practice for two reasons: it is more robust and it is more simple to obtain results.