An upper bound for BDeu scores

James Cussens, University of York

AIGM 2012

BDeu local scores and BN structure learning

$$\log P(G|D) = \sum_{i=1}^{p} z_i(G) = \sum_{i=1}^{p} z_i(\mathrm{Pa}_i(G))$$

Let W be a candidate parent set for i:

$$\begin{aligned} z_i(W) &= \sum_{j=1}^{q_i} \left(\log \frac{\Gamma(\frac{\alpha}{q_i})}{\Gamma(n_{ij} + \frac{\alpha}{q_i})} + \sum_{k=1}^{r_i} \log \frac{\Gamma(n_{ijk} + \frac{\alpha}{q_i r_i})}{\Gamma(\frac{\alpha}{q_i r_i})} \right) \\ &= \sum_{j=1}^{q^{(+)}} \left(\log \frac{\Gamma(\frac{\alpha}{q_i})}{\Gamma(n_{ij} + \frac{\alpha}{q_i})} + \sum_{k=1}^{r_i} \log \frac{\Gamma(n_{ijk} + \frac{\alpha}{q_i r_i})}{\Gamma(\frac{\alpha}{q_i r_i})} \right) \end{aligned}$$

where $q^{(+)} = |\{j : n_{ij} > 0\}|$

Bounding local scores: why do we care?

- If $W' \supset W$ and $z_i(W) > z_i(W')$ then W' cannot be a parent set for i in an optimal BN.
- If $\forall W' : W' \supset W$ we have $z_i(W) > z_i(W')$, then all these subsets can be pruned from a search for an optimal parent set.

If:

1. $W \subset W'$
2. $\alpha/q_i(W') \leq 0.8349$
3. $z_i(W) > -q^{(+)}(W') \log r_i$

then neither W' *nor any superset of* W' can be an optimal parent set for i. de Campos and Ji [1]

Column generation for integer linear programming approaches to BN learning

- In ILP approaches to BN structure learning we can represent choosing parent set W for child i by setting a binary variable $I(W \rightarrow i)$ to 1.
- Unless the number of possible parent sets is artificially restricted there will be too many variables to represent explicitly.
- This motivates a column generation (= variable generation) approach where variables with *negative reduced cost* are created.
- A bound on the local score (the 'unreduced' cost) provides a bound on negative reduced cost.

The beta function of r variables

Define the beta function of r variables:

$$B(x_1, \ldots, x_r) = \frac{\Gamma(x_1) \ldots \Gamma(x_r)}{\Gamma(x_1 + \cdots + x_r)}$$

$$\begin{aligned} z_i(G) \quad = \quad & q^{(+)} \log \Gamma\left(\frac{\alpha}{q_i}\right) - q^{(+)} r_i \log \Gamma\left(\frac{\alpha}{q_i r_i}\right) \\ & + \sum_{j=1}^{q^{(+)}} \log B\left(n_{ij1} + \frac{\alpha}{q_i r_i}, \ldots n_{ijr_i} + \frac{\alpha}{q_i r_i}\right) \end{aligned}$$

Alzer's (upper) bound

Let $c > 0$ be a real number and let $r \geq 2$ be an integer. Then we have for all real numbers $x_k \geq c$ $(k = 1, \ldots, r)$:

$$B(x_1, \ldots, x_r) \leq \beta_r(c) \frac{\prod_{k=1}^{r} x_k^{-1/2+x_k}}{\left(\sum_{k=1}^{r} x_k\right)^{-1/2+\sum_{k=1}^{r} x_k}}$$

with the best possible constant $\beta_r(c) = r^{rc-1/2} c^{(r-1)/2} \frac{(\Gamma(c))^r}{\Gamma(rc)}$.

- Apply to $B\left(n_{ij1} + \frac{\alpha}{q_i r_i}, \ldots n_{ijr_i} + \frac{\alpha}{q_i r_i}\right)$
- with $c = \frac{\alpha}{q_i r_i}$

Plugging in Alzer's bound

After much algebraic manipulation ...

$$
\begin{aligned}
& z_i(W) \\
& \quad \leq \quad \alpha \log r_i + q^{(+)}[(r_i - 1)\log(\alpha/q_i r_i) - \log(r_i)]/2 \\
& \qquad -(N+\alpha)H_{\tilde{p}}(i|W) \\
& \qquad +\frac{1}{2}\left[\sum_{j=1}^{q^{(+)}} \log\left(n_{ij} + \frac{\alpha}{q_i}\right) - \sum_{k=1}^{r_i} \log\left(n_{ijk} + \frac{\alpha}{q_i r_i}\right)\right]
\end{aligned}
$$

where

- $H_{\tilde{p}}(i|W)$ is conditional entropy ('fit to data')
- If $W' \supset W$ then $H_{\tilde{p}}(i|W) \geq H_{\tilde{p}}(i|W')$

Plugging in Alzer's bound

After much algebraic manipulation ...

$$
\begin{aligned}
& z_i(W) \\
& \leq \quad \alpha \log r_i + q^{(+)}[(r_i - 1)\log(\alpha/q_i r_i) - \log(r_i)]/2 \\
& \quad -(N+\alpha)H_{\tilde{p}}(i|W) \\
& \quad +\frac{1}{2}\left[\sum_{j=1}^{q^{(+)}} \log\left(n_{ij} + \frac{\alpha}{q_i}\right) - \sum_{k=1}^{r_i} \log\left(n_{ijk} + \frac{\alpha}{q_i r_i}\right)\right]
\end{aligned}
$$

where

- $H_{\tilde{p}}(i|W)$ is conditional entropy ('fit to data')
- If $W' \supset W$ then $H_{\tilde{p}}(i|W) \geq H_{\tilde{p}}(i|W')$
- $\tilde{p}$ is the posterior distribution in the saturated model when starting from a Dirichlet prior determined by α.

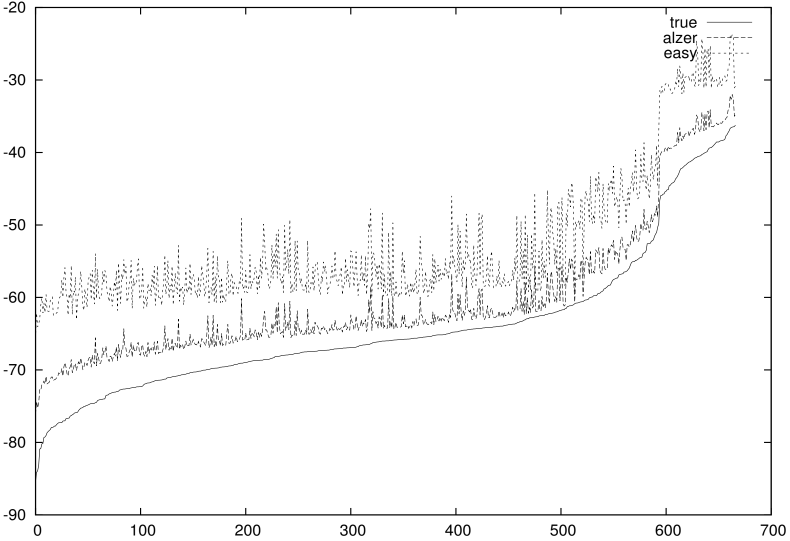
Rewarding determinism

The last term:

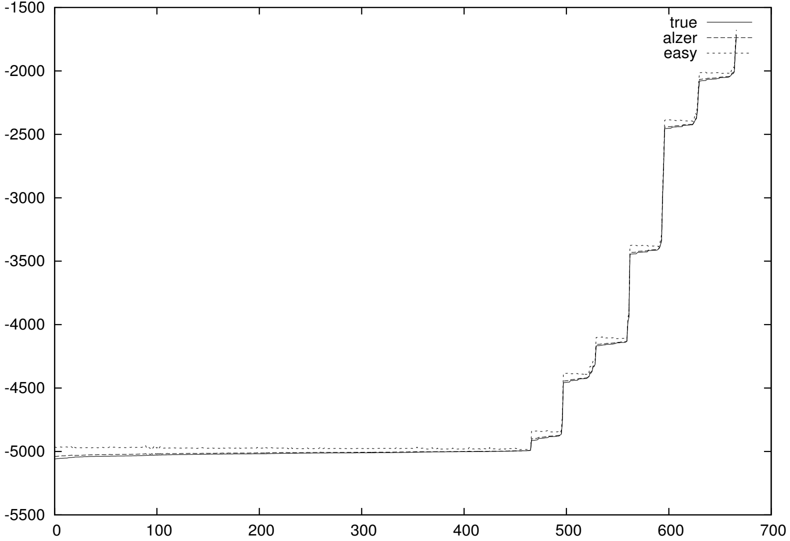
$$\sum_{j=1}^{q^{(+)}} \log\left(n_{ij} + \frac{\alpha}{q_i}\right) - \sum_{k=1}^{r_i} \log\left(n_{ijk} + \frac{\alpha}{q_i r_i}\right)$$

- Note that $n_{ij} > 0$
- Each $n_{ijk} > 0$ exacts a penalty
- Each $n_{ijk} = 0$ (typically) bestows a reward
- 'Best' situation is $n_{ij} = n_{ijk'}$ for some k' ...
- ...this is when (according to the data) parents determine the child's value

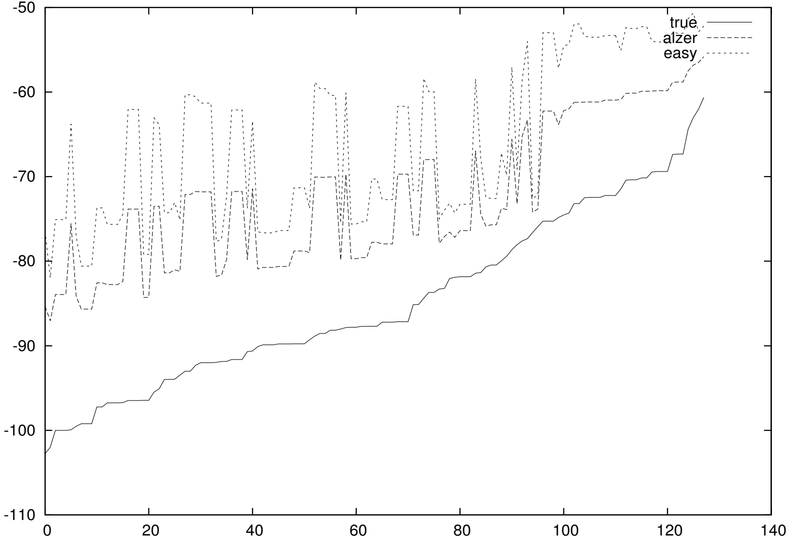
Scores and bounds. 100 datapoints. $\alpha = 1$



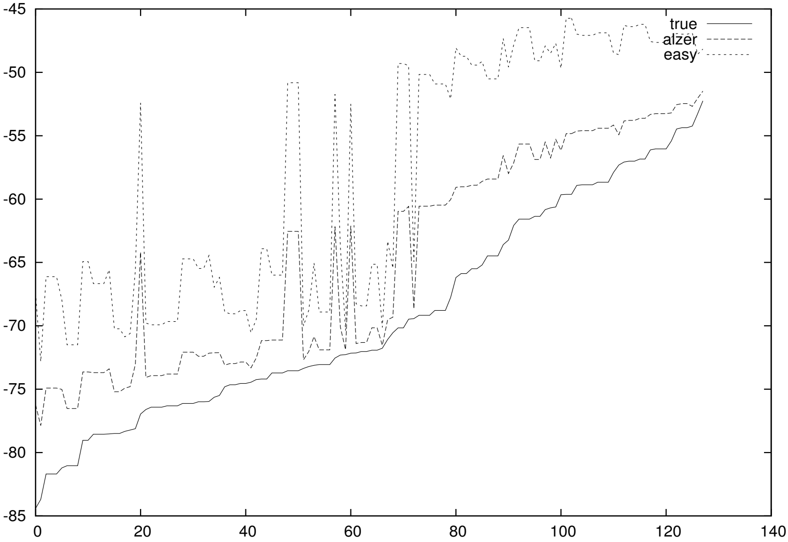
Scores and bounds. 10000 datapoints. $\alpha = 1$



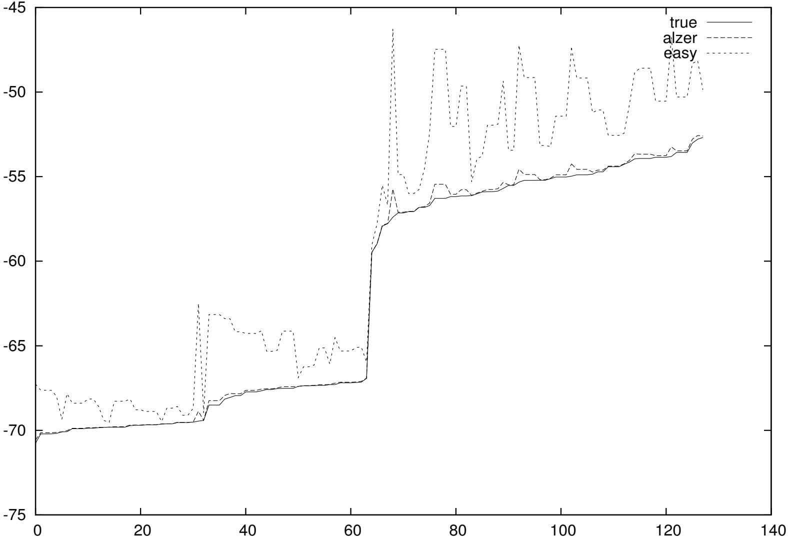
Scores and bounds. 100 datapoints. $\alpha = 0.01$



Scores and bounds. 100 datapoints. $\alpha = 1$



Scores and bounds. 100 datapoints. $\alpha = 100$



A looser bound for pruning

This is not in the paper ...

$$
\begin{aligned}
z_i(W) \quad \leq \quad & \alpha \log r_i \\
& -(N+\alpha) H_{\tilde{p}}(i|W) \\
& + \frac{1}{2} q^{(+)} \log \left(\frac{\frac{1}{r_i} + \frac{\alpha}{q_i r_i}}{1 + \frac{\alpha}{q_i r_i}} \right)
\end{aligned}
$$

Suppose $\underline{W} \subset W \subset \overline{W}$
If

$$
\alpha \log r_i - (N+\alpha) H_{\tilde{p}}(i|\overline{W}) + \frac{1}{2} q^{(+)} \log \left(\frac{\frac{1}{r_i} + \frac{\alpha}{q_i r_i}}{1 + \frac{\alpha}{q_i r_i}} \right) < z_i(\underline{W})
$$

then W can not be an optimal parent set for i.

Comparison with de Campos and Ji

$$z_i(\underline{W}) > \alpha \log r_i - (N + \alpha) H_{\tilde{p}}(i|\overline{W}) + \frac{1}{2} q^{(+)} \log \left(\frac{\frac{1}{r_i} + \frac{\alpha}{q_i r_i}}{1 + \frac{\alpha}{q_i r_i}} \right) \quad (1)$$

$$z_i(\underline{W}) > -q^{(+)} \log r_i \quad (2)$$

- Prune if either (1) or (2) applies.
- (1) motivates a bidirectional search.

C de Campos and Q Ji.
Properties of Bayesian Dirichlet scores to learn Bayesian network structures.
In *AAAI-10*, pages 431–436, 2010.