Supplementary Material for the article
Quotient Normalized Maximum Likelihood Criterion for Learning Bayesian Network Structures

A qNML coincides with NML for many models

A.1 qNML EQUALS NML FOR MANY MODELS

The fNML criterion can be seen as a computationally feasible approximation of the more desirable NML criterion. However, the fNML criterion equals the NML criterion only for the Bayesian network structure with no arcs. We will next show that the qNML criterion equals the NML criterion for all the networks G whose connected components tournaments (i.e., complete directed acyclic subgraphs of G).

Theorem 1. *If G consists of C connected components $(G^1, \ldots, G^C)$ with variable sets $(V^1, \ldots, V^C)$, then $\log P_{NML}(D; G) = s^{qNML}(D; G)$ for all data sets D.*

Proof. We first show that the NML-criterion for a Bayesian network decomposes by the connected components.

Because the maximum likelihood for the data D decomposes, we can write

$$\begin{aligned}
&P_{NML}(D;G) \\
&= \frac{P(D;\hat{\theta}(D),G)}{\sum_{D'_{V_1}}\cdots\sum_{D'_{V_C}}\prod_{c=1}^{C}P(D'_{V^c};\hat{\theta}(D'_{V^c}),G)} \\
&= \frac{\prod_{c=1}^{C}P(D_{V^c};\hat{\theta}(D_{V^c}),G)}{\prod_{c=1}^{C}\sum_{D'_{V^c}}P(D'_{V^c};\hat{\theta}(D'_{V^c}),G)}. \\
&= \prod_{c=1}^{C}P_{NML}(D_{V^c};G). \qquad (1)
\end{aligned}$$

Clearly, the qNML score also decomposes by the connected components, so it remains to show that if the (sub)network G is a tournament, then for any data D, $s^{qNML}(D;G) = \log P_{NML}(D;G)$. Due to the score equivalence of the NML criterion and the qNML criterion, we may pick a tournament G such that the linear ordering defined by G matches the ordering of the data columns, i.e., $i < j$ implies $G_i \subset G_j$. Now from the definition (**??**) of the qNML criterion we see that for the tournament G, the sum telescopes leaving us with $s^{qNML}(D;G) = \log P^1_{NML}(D_G;G)$, thus it is enough to show that $P^1_{NML}(D;G) = P_{NML}(D;G)$. This follows, since for any data vector x in data D, we have $P^1(x;\hat{\theta}(D),G) = P(x;\hat{\theta}(D),G)$, where P^1 denotes the model that takes n-dimensional vectors to be values of the single (collapsed) categorical variable. Denoting prefixes of data vector x by $x^{:i}$, and the number of times such a prefix appears on N rows $[d_1,\ldots,d_N]$ of the $N\times n$ data matrix D by $N_D(x^{:i})$, (so that $N_D(x^{:0}) = N$), we have

$$\begin{aligned}
P(D;\hat{\theta}(D),G) &= \prod_{j=1}^{N}P(d_j;\hat{\theta}(D),G) \\
&= \prod_{j=1}^{N}\prod_{i=1}^{n}\frac{N_D(d_j^{:i})}{N_D(d_j^{:i-1})} \\
&= \prod_{j=1}^{N}\frac{N_D(d_j^{:n})}{N} \\
&= P^1(D;\hat{\theta}^1(D),G). \qquad (2)
\end{aligned}$$

Since both P^1_{NML} and P_{NML} are defined in terms of the maximum likelihood probabilities, the equality above implies the equality of these distributions. $\square$

Equality established, we would like to still state the number $a(n)$ of different n-node networks whose connected components are tournaments. We start by generating all the $p(n)$ integer partitions i.e. ways to partition n labelled into parts with different size-profiles. For example, for $n = 4$, we have $p(4) = 5$ partition profiles $[[4], [3, 1], [2, 2], [2, 1, 1], [1, 1, 1, 1]]$. Each of these partition size profiles corresponds to many different networks, apart from the last one that corresponds just to the empty network. We count number of networks for one such partition size-profile (and then later sum these counts up). For any such partition profile $(p_1, \ldots, p_k)$ we can count the ways we can assign the nodes to different parts and then order each part. This leads to the product $\binom{n}{p_1} p_1! \binom{n-p_1}{p_2} p_2! \binom{n-p_1-p2}{p_3} p_3! \ldots \binom{n-\sum_{j=1}^{k-1} p_j}{p_k} p_k!$. However, the order of different parts of the same size does not matter, so for all groups of parts having the same size, we have to divide the product above by the factorial of the size of such group. Notice also, that the product above telescopes, leaving us a formula for OEIS sequence A000262[1] as described by Thomas Wieder:

With $p(n)$ = the number of integer partitions of n, $d(i)$ = the number of different parts of the i^{th} partition of n, $m(i, j)$ = multiplicity of the j^{th} part of the i^{th} partition of n, one has:

$$a(n) = \sum_{i=1}^{p(n)} \frac{n!}{\prod_{j=1}^{d(i)} m(i, j)!}.$$

For example, $a(4) = \frac{4!}{1!} + \frac{4!}{1!1!} + \frac{4!}{2!} + \frac{4!}{1!2!} + \frac{4!}{4!} = 73$. In general this sequence grows rapidly; $1, 1, 3, 13, 73, 501, 4051, 37633, 394353, 4596553, \ldots$.

References

[1] https://oeis.org/A000108