1 Problem

The score function $Q_n(X \mid Y)$, where X refers to one random variable and Y is some proposed parent set for that variable, is *regular* if

$$H_n(X \mid U) \leq H_n(X \mid U \cup V) \quad \Rightarrow \quad Q_n(X \mid U) \geq Q_n(X \mid U \cup V),$$

where $H_n(\cdot \mid \cdot)$ refers to conditional empirical entropy and U and V are some sets of random variables.

Let X be a categorical random variable with r possible values. Let U denote a possible parent set with q different combinations of values for the variables and V a parent set with m different configurations. Assume that we have observed n samples of (X, U, V) and $H_n(X \mid U) \leq H_n(X \mid U \cup V)$ holds.

We want to show that

$$Q_n^{qnml}(X \mid U) \geq Q_n^{qnml}(X \mid U \cup V).$$

The assumption about entropy implies that the maximized likelihood terms of the qnml-score are equal. In order to prove the claim it suffices to study the penalty terms, and we want to show that

$$\begin{aligned} -[reg(n, rq) - reg(n, q)] &\geq -[reg(n, rqm) - reg(n, qm)] \\ reg(n, rq) - reg(n, q) &\leq reg(n, rqm) - reg(n, qm) \\ \frac{C(n, rq)}{C(n, q)} &\leq \frac{C(n, rqm)}{C(n, qm)}, \end{aligned}$$

where $C(n, k)$ is the stochastic complexity for multinomial random variable with k categories. Since, trivially, $q \leq qm$, we can prove the claim by showing that the function $k \mapsto C(n, rk)/C(n, k)$ is increasing for arbitrary $r \geq 2$ and $n \geq 1$.

2 Stochastic complexity

We can represent $C(n, k)$ via recursion

$$C(n, k) = C(n, k - 1) + C(n, k - 2)n/(k - 2)$$

or by using the following formula

$$
\begin{aligned}
C(n,k) &= \sum_{l=0}^{n} \frac{n^{\underline{l}}\ (k-1)^{\overline{l}}}{n^l\ l!} \\
&= \sum_{l=0}^{n-1} \frac{(n-1)^{\underline{l}} k^{\overline{l+1}}}{n^{l+1}\ l!}
\end{aligned}
$$

where $x^{\underline{l}}$ and $x^{\overline{l}}$ denote falling and rising factorials, respectively.

3 Case $n = 2$

$$
\frac{C(2,rk)}{C(2,k)} = \frac{1 + (rk-1) + (rk-1)rk/4}{1 + (k-1) + (k-1)k/4} = \frac{r^2k^2 + 3rk}{k^2 + 3k},
$$

and since

$$
\frac{d}{dk}\left(\frac{C(2,rk)}{C(2,k)}\right) = \frac{3r(r-1)k^2}{(k^2+3k)^2} = \frac{3r(r-1)}{(k+3)^2} > 0
$$

for every $r \geq 2$ and $k \geq 1$, the claim holds.

4 Case $n = 3$

$$
\begin{aligned}
\frac{C(3,rk)}{C(3,k)} &= \frac{r^3k^2 + 9r^2k + 17r}{k^2 + 9k + 17} \quad \text{and} \\
\frac{d}{dk}\left(\frac{C(3,rk)}{C(3,k)}\right) &= \frac{(r-1)r(153 + 34(r+1)k + 9rk^2)}{(17 + 9k + k^2)^2} > 0
\end{aligned}
$$

5 Case $n = 4$

$$
\frac{C(4,rk)}{C(4,k)} = \frac{142r + 95r^2k + 18r^3k^2 + r^4k^3}{142 + 95k + 18k^2 + k^3}
$$

It seems that the denominator of the derivative contains "always" a polynomial of k with positive coefficients, and since the nominator is some polynomial squared, the resulting quotient is always positive for arbitrary r.

6 Case general

We first derive a representation for $C(n,k)$ as a polynomial of k. We utilize the fact that the rising factorial can be represented as polynomial using unsigned Stirling numbers of the first kind

$$\begin{aligned}
C(n,k) &= \sum_{l=0}^{n-1} \frac{(n-1)^{\underline{l}} k^{\overline{l+1}}}{n^{l+1}\ l!} \\
&= \sum_{l=0}^{n-1} b_l\ k^{\overline{l+1}} \\
&= \sum_{l=0}^{n-1} b_l \left(\sum_{j=1}^{l+1} |s(l+1,j)|\ k^j \right) \\
&= \sum_{l=0}^{n-1} \left(\sum_{j=1}^{n} b_l\ |s(l+1,j)|\ k^j \right) \\
&= \sum_{j=1}^{n} \left(\sum_{l=0}^{n-1} b_l\ |s(l+1,j)|\ k^j \right) \\
&= \sum_{j=1}^{n} \left(\sum_{l=0}^{n-1} b_l\ |s(l+1,j)| \right) k^j \\
&= \sum_{j=1}^{n} a_j k^j,
\end{aligned}$$

where $s(x,y)$ denotes the Stirling number of the first kind and

$$a_j = \left(\sum_{l=0}^{n-1} \frac{(n-1)^{\underline{l}}}{n^{l+1} l!}\ |s(l+1,j)| \right),$$

$a_j \geq 0$ for all j (also it seems that $\sum_{j=1}^{n} a_j = 1$). On the row 4, we used the property of Stirling numbers: $s(i,j) = 0$ for all $j > i$. Similarly,

$$C(n,rk) = \sum_{j=1}^{n} a_j r^j k^j$$

Derivatives are obtained easily from this form

$$\begin{aligned}\frac{d}{dk}C(n,k) &= \sum_{j=1}^{n} ja_j k^{j-1} \\ &= \sum_{j=0}^{n-1}(j+1)a_{j+1}k^j\end{aligned}$$

and

$$\begin{aligned}\frac{d}{dk}C(n,rk) &= \sum_{j=1}^{n} ja_j r^j k^{j-1} \\ &= \sum_{j=0}^{n-1}(j+1)a_{j+1}r^{j+1}k^j.\end{aligned}$$

Consider next the products found in the derivative of the quotient. We obtain

$$\begin{aligned}\left(\frac{d}{dk}C(n,rk)\right)C(n,k) &= \left(\sum_{j=0}^{n-1}(j+1)a_{j+1}r^{j+1}k^j\right)\left(\sum_{l=1}^{n} a_l k^l\right) \\ &= \sum_{i=1}^{2n-1}\left(\sum_{j+l=i}(j+1)a_{j+1}r^{j+1}a_l\right)k^i\end{aligned}$$

and

$$\begin{aligned}\left(\frac{d}{dk}C(n,k)\right)C(n,rk) &= \left(\sum_{j=0}^{n-1}(j+1)a_{j+1}k^j\right)\left(\sum_{l=1}^{n} a_l r^l k^l\right) \\ &= \sum_{i=1}^{2n-1}\left(\sum_{j+l=i}(j+1)a_{j+1}a_l r^l\right)k^i.\end{aligned}$$

Subtracting these two expression yields

$$\begin{aligned}&\left(\frac{d}{dk}C(n,rk)\right)C(n,k) - \left(\frac{d}{dk}C(n,k)\right)C(n,rk) \\ &= \sum_{i=1}^{2n-1}\left(\sum_{j+l=i}(j+1)a_{j+1}r^{j+1}a_l\right)k^i - \sum_{i=1}^{2n-1}\left(\sum_{j+l=i}(j+1)a_{j+1}a_l r^l\right)k^i \\ &= \sum_{i=1}^{2n-1}\left(\sum_{j+l=i}(j+1)a_{j+1}a_l(r^{j+1}-r^l)\right)k^i\end{aligned}$$

which is the polynomial in the denominator of the derivative of $C(n, rk)/C(n, k)$. Next, we study the coefficient of k^i, if $i \leq n$

$$
\begin{aligned}
\sum_{j+l=i} (j+1)a_{j+1}a_l(r^{j+1} - r^l) &= \sum_{l=1}^{i} (i-l+1)a_{i-l+1}a_l(r^{i-l+1} - r^l) \\
&= \sum_{l=1}^{i} (i-l+1)c_l \\
&= \sum_{k=1}^{\lfloor i/2 \rfloor} (i-k+1)c_k + (i-(i-k+1)+1)c_{i-k+1} \\
&= \sum_{k=1}^{\lfloor i/2 \rfloor} (i-k+1)c_k + kc_{i-k+1} \\
&= \sum_{k=1}^{\lfloor i/2 \rfloor} (i-k+1)c_k - kc_k \\
&= \sum_{k=1}^{\lfloor i/2 \rfloor} (i-2k+1)c_k.
\end{aligned}
$$

On the first row, we re-wrote sum using only one running index. On the second row we denoted $c_l = a_{i-l+1}a_l(r^{i-l+1} - r^l)$. On the third row, we re-arranged the sum so that we are summing over pairs of terms of the original sum: the first and the last term, the second and the second to last, and so on. This resulting sum has $\lfloor i/2 \rfloor$ terms. We have to use the floor-function since if i is odd, there exists an index l' in the original sum such that $r^{i-l'+1} - r^{l'} = 0$. On the fifth row, we make use of the identity $c_k = -c_{i-k+1}$ which is straightforward to verify. From the last row, we can observe that every term of the sum is positive since $i - 2k + 1$ and $r^{i-k+1} - r^k$ are both positive if $k \leq (i+1)/2$ which holds since k ranges from 1 to $\lfloor i/2 \rfloor$.

Let us now consider the situation where $n < i \leq 2n - 1$. We start with the special case where $i = 2n - 1$. Then, we have only one term in the sum

$$
\begin{aligned}
\sum_{j+l=i} (j+1)a_{j+1}a_l(r^{j+1} - r^l) &= \sum_{l=n}^{n} (2n-1-l+1)a_{2n-1-l+1}a_l(r^{2n-1-l+1} - r^l) \\
&= na_na_n(r^n - r^n) \\
&= 0.
\end{aligned}
$$

Now, let $n < i < 2n-1$, we follow a similar procedure as before to manipulate the sum

$$
\begin{aligned}
\sum_{j+l=i} (j+1)a_{j+1}a_l(r^{j+1} - r^l) &= \sum_{l=i-n+1}^{n} (i-l+1)a_{i-l+1}a_l(r^{i-l+1} - r^l) \\
&= \sum_{l=i-n+1}^{n} (i-l+1)c_l \\
&= \sum_{k=1}^{2n-i}(n-k+1)c_{i-n+k} \\
&= \sum_{k=1}^{\lfloor n-i/2 \rfloor} (n-k+1)c_{i-n+k} \\
&+ (n-(2n-i-k+1)+1)c_{i-n+(2n-i-k+1)} \\
&= \sum_{k=1}^{\lfloor n-i/2 \rfloor} (n-k+1)c_{i-n+k} + (i-n+k)c_{n-k+1} \\
&= \sum_{k=1}^{\lfloor n-i/2 \rfloor} (n-k+1)c_{i-n+k} - (i-n+k)c_{i-n+k} \\
&= \sum_{k=1}^{\lfloor n-i/2 \rfloor} (n-k+1-(i-n+k))c_{i-n+k} \\
&= \sum_{k=1}^{\lfloor n-i/2 \rfloor} (2n-i-2k+1)c_{i-n+k}.
\end{aligned}
$$

It is now easy to verify that $(2n-i-2k+1)$ and c_{i-n+k} are positive if $k \leq n-(i-1)/2$ which holds since k ranges from 1 to $\lfloor n-i/2 \rfloor$. The floor function is again used when we sum over pairs of terms since if i is odd there is zero-term. Since all the coefficients are non-negative and the $k \geq 2$, the derivative is positive. This implies that the original function is increasing.

7 Terms a_j sum to one

$$\begin{aligned}
\sum_{j=1}^{n} a_j &= \sum_{j=1}^{n} \left(\sum_{l=0}^{n-1} \frac{(n-1)^{\underline{l}}}{n^{l+1} l!} \; |s(l+1,j)| \right) \\
&= \sum_{l=0}^{n-1} \frac{(n-1)^{\underline{l}}}{n^{l+1} l!} \left(\sum_{j=1}^{n} |s(l+1,j)| \right) \\
&= \sum_{l=0}^{n-1} \frac{(n-1)^{\underline{l}}}{n^{l+1} l!} \left(\sum_{j=0}^{l-1} |s(l+1,j)| \right) \\
&= \sum_{l=0}^{n-1} \frac{(n-1)^{\underline{l}}}{n^{l+1} l!} \; (l+1)! \\
&= \sum_{l=0}^{n-1} \frac{(n-1)^{\underline{l}}}{n^{l+1}} (l+1)
\end{aligned}$$

8 ML terms

Assume we have n samples of X, Y and Z. We can write the logarithm of the maximized likelihood, $\hat{P}_n(X \mid Y)$, as follows

$$\begin{aligned}
\log \hat{P}_n(X \mid Y) &= -n \left(H_n(X) - I_n(X;Y) \right) \\
&= -n \; H_n(X \mid Y),
\end{aligned}$$

where $I_n(\cdot;\cdot)$ is the empirical mutual information. This implies that the assumption

$$H_n(X \mid Y) \leq H_n(X \mid Y \cup Z)$$

is equivalent to

$$\log \hat{P}_n(X \mid Y) \geq \log \hat{P}_n(X \mid Y \cup Z).$$

Actually we must have the equality holding in the above expression, since

$$H_n(X \mid Y) < H_n(X \mid Y \cup Z)$$

would imply that

$$I_n(X; Z \mid Y) < 0$$

which is impossible.