AMBN 2017

B&B for BNSL

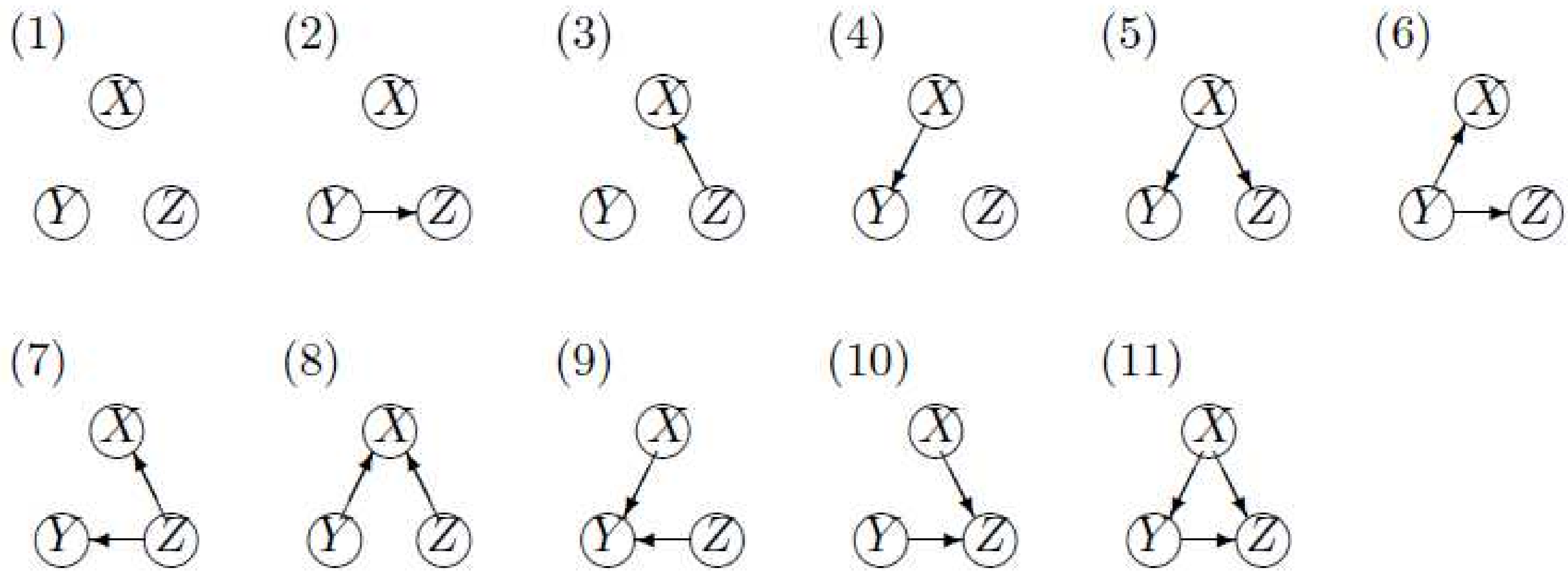
Joe Suzuki (Osaka U.)
joint work with Jun Kawahara (Nara Sen)

1. Define Regularity of BNSL
2. BDeu is not Regular
3. Markov equivalent structures share not just BDeu but also quotient scores
4. Regular BNSL offers efficient B & B
5. Propose a regular quotient score

BNSL

$p(\geq 3)$ variables such as X, Y, Z Discrete

X	Y	Z
x_1	y_1	z_1
$\vdots$	$\vdots$	$\vdots$
x_n	y_n	z_n



Scores

$$x^n = (x_1, \cdots, x_n) \in \{0,1\}^n \text{ contains } c \text{ ones and } n-c \text{ zeros}$$
$$\theta := P(X=1) \text{ is not known}$$

$$Q(x^n) := \int_0^1 \theta^c (1-\theta)^{n-c} w(\theta) d\theta = \prod_{i:x_i=1} \frac{c_{i-1}(x_i)+a}{i-1+a+b} \prod_{i:x_i=0} \frac{c_{i-1}(x_i)+b}{i-1+a+b}$$

$$w(\theta) \propto \theta^{a-1}(1-\theta)^{b-1} \text{ with } a, b > 0$$

$$c_{i-1}(x_i)\text{: the \# of } x_i \text{ in } (x_1, \cdots, x_{i-1})$$

$$a = 0.2, b = 0.1, x^5 = (0,1,0,1,1) \Longrightarrow Q(x^5) = \frac{0+0.2}{0+0.3} \cdot \frac{0+0.1}{1+0.3} \cdot \frac{1+0.2}{2+0.3} \cdot \frac{1+0.1}{3+0.3} \cdot \frac{2+0.1}{4+0.3}$$

Conditional/Quotient Scores

$$x^n = (x_1, \cdots, x_n) \in \{0, 1, \cdots, \alpha - 1\}^n,\ y^n = (y_1, \cdots, y_n) \in \{0, 1, \cdots, \beta - 1\}^n$$
$$h(\cdot),\ h(\cdot, \cdot):\ \text{constants}$$

Conditional Scores:

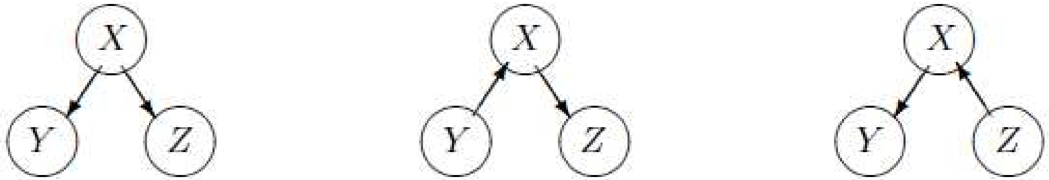
$$Q(x^n|y^n) := \prod_{y=0}^{\beta-1} \prod_{i:y_i=y} \frac{c_{i-1}(x_i, y) + h(x_i, y)}{i - 1 + \sum_{x=0}^{\alpha-1} h(x, y)}$$

Quotient Scores:

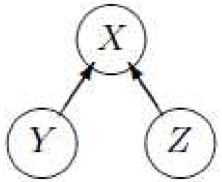
$$Q(x^n|y^n) := \frac{Q(x^n, y^n)}{Q(y^n)}$$

$$Q(x^n, y^n) := \prod_{i=1}^{n} \frac{c_{i-1}(x_i, y_i) + h(x_i, y_i)}{i - 1 + \sum_{x=0}^{\alpha-1} \sum_{y=0}^{\beta-1} h(x, y)} \qquad Q(y^n) := \prod_{i=1}^{n} \frac{c_{i-1}(y_i) + h(y_i)}{i - 1 + \sum_{y=0}^{\beta-1} h(y)}$$

Markov Equivalent Structures



$$\underbrace{P(X)P(Y|X)P(Z|X) = P(Y)P(X|Y)P(Z|X) = P(Z)P(X|Z)P(Y|X)}_{= \frac{P(XY)P(XZ)}{P(X)}} \qquad (5)$$



$$\underbrace{P(Y)P(Z)P(X|YZ)}_{= \frac{P(Y)P(Z)P(XYZ)}{P(YZ)}} \qquad (8)$$

$$\begin{cases} Q(x^n|y^n)Q(y^n) = Q(y^n|x^n)Q(x^n) \\ Q(y^n|z^n)Q(z^n) = Q(z^n|y^n)Q(y^n) \\ Q(z^n|x^n)Q(x^n) = Q(x^n|z^n)Q(z^n) \end{cases} \qquad \begin{cases} Q(x^n|y^n,z^n)Q(y^n,z^n) = Q(y^n,z^n|x^n)Q(x^n) \\ Q(y^n|z^n,x^n)Q(z^n,x^n) = Q(z^n,x^n|y^n)Q(y^n) \\ Q(z^n|x^n,y^n)Q(x^n) = Q(x^n,y^n|z^n)Q(z^n) \end{cases}$$

For Markov equivalent structures to share the Conditional Score, there should be $\delta > 0$ s.t.

$$\begin{cases} Q(x^n) = \prod_{i=1}^{n} \frac{c_{i-1}(x_i) + \frac{\delta}{\alpha}}{i-1+\delta} \\ Q(x^n|y^n) = \prod_{y=0}^{\beta-1} \prod_{i:y_i=y} \frac{c_{i-1}(x_i, y) + \frac{\delta}{\alpha\beta}}{i-1+\frac{\delta}{\beta}} \\ Q(x^n|y^n, z^n) = \prod_{y=0}^{\beta-1} \prod_{z=0}^{\gamma-1} \prod_{i:(y_i,z_i)=(y,z)} \frac{c_{i-1}(x_i, y, z) + \frac{\delta}{\alpha\beta}}{i-1+\frac{\delta}{\beta\gamma}} \end{cases}$$

$$\begin{cases} Q(y^n, z^n) = \prod_{i=1}^{n} \frac{c_{i-1}(y_i, z_i) + \frac{\delta}{\beta\gamma}}{i-1+\delta} \\ Q(y^n, z^n|x^n) = \prod_{x=0}^{\alpha-1} \prod_{i:x_i=x} \frac{c_{i-1}(x, y_i, z_i) + \frac{\delta}{\alpha\beta\gamma}}{i-1+\frac{\delta}{\alpha}} \end{cases}$$

$$Q(x^n, y^n, z^n) = \prod_{i=1}^{n} \frac{c_{i-1}(x_i, y_i, z_i) + \frac{\delta}{\alpha\beta\gamma}}{i-1+\frac{\delta}{\alpha\beta\gamma}}$$

α, β: # of values that X and Y take

BDeu (Buntine, 1991)

Markov equivalent structures share any Quotient Score

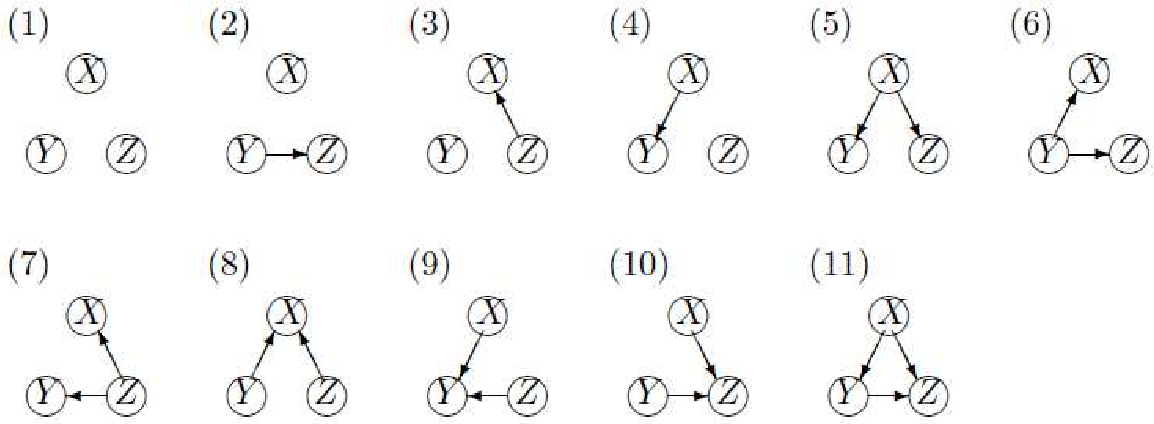
$$Q(x^n)Q^n(y^n)Q^n(z^n)$$

$$Q(x^n)Q(y^n, z^n), Q(y^n)Q(z^n, x^n), Q(z^n)Q(x^n, y^n)$$

$$\frac{Q(z^n, x^n)Q^n(x^n, y^n)}{Q(x^n)}, \frac{Q(x^n, y^n)Q(y^n, z^n)}{Q(y^n)}, \frac{Q(z^n, x^n)Q(x^n, y^n)}{Q(z^n)}$$

$$\frac{Q(y^n)Q(z^n)Q(x^n, y^n, z^n)}{Q(y^n, z^n)}, \frac{Q(z^n)Q(x^n)Q(x^n, y^n, z^n)}{Q(z^n, x^n)}, \frac{Q(x^n)Q(y^n)Q(x^n, y^n, z^n)}{Q(x^n, y^n)},$$

$$Q(x^n, y^n, z^n)$$



B&B for BNSL

Silander & Myllimaki 2008

$Q(x^n|S)$: the Conditional/Quotient Score w.r.t. parent set S

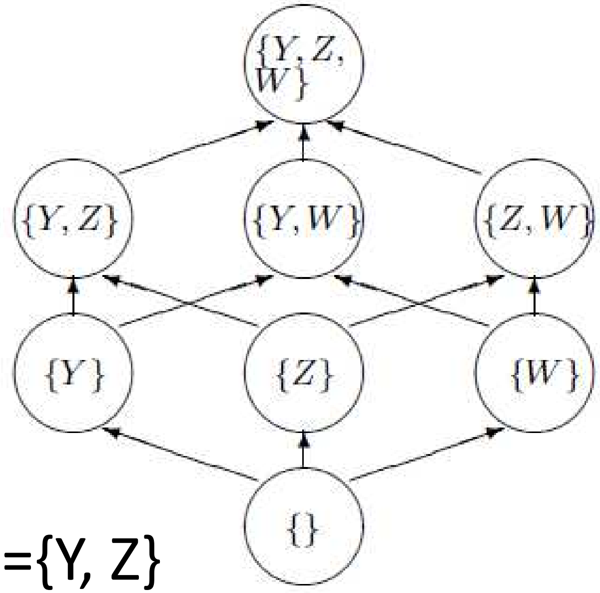
$$R(x^n|S) := \max_{U \subseteq S} Q(x^n|U) = \max\{Q(x^n|S), \max_{Y \in S} R(x^n|S \setminus \{Y\})\}$$

$$R(x^n|\{\}) = Q(x^n)$$

$$R(x^n|\{Y\}) = \max\{Q(x^n|y^n), Q(x^n)\}$$

$$R(x^n|\{Z\}) = \max\{Q(x^n|z^n), Q(x^n)\}$$

$$R(x^n|\{Y,Z\}) = \max\{Q(x^n|y^n, z^n), Q(x^n|y^n), Q(x^n|z^n), Q(x^n)\}$$



S={Y, Z}

T={Y, Z}, {Y, Z, W}

$$R(x^n|S \setminus \{Z\}) \geq R_*(x^n|S) \geq \sup_{T \supseteq S} Q(x^n|T) \text{ for any } Z \in S$$

BOUND

$$\Longrightarrow \text{no } Q(x^n|T) \text{ need to be computed for any } T \supseteq S$$

For BDeu, only a loose pruning rule has been found (Campos & Ji, 2011)

BDeu violates regularity

The empirical conditional entropy $H(x^n|y^n) = 0$
BDeu chooses $\{Y, Z\}$

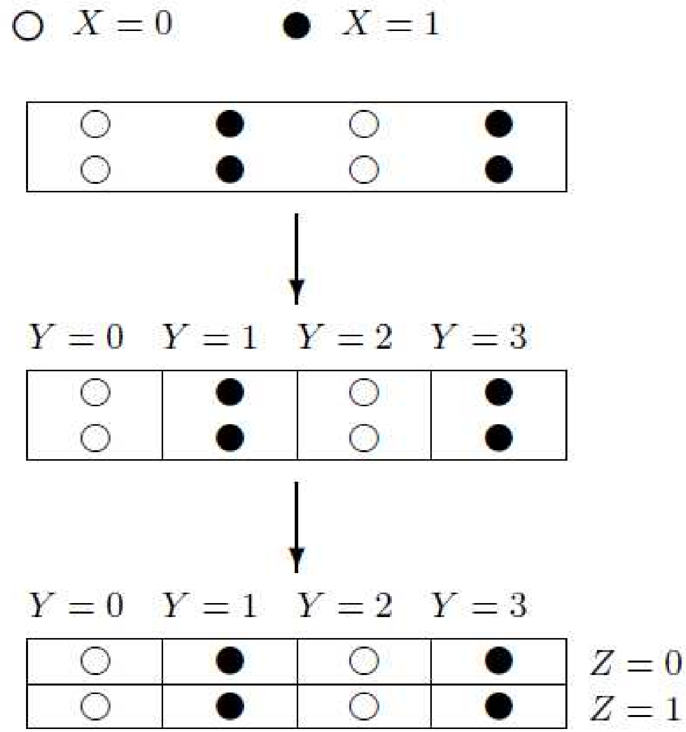
$$Q(x^n|y^n) = (\frac{1/8}{1/4} \cdot \frac{1+1/8}{1+1/4})^4 = (\frac{9}{10})^4 2^{-4}$$

$$Q(x^n|y^n, z^n) = (\frac{1/16}{1/8} \cdot \frac{1+1/16}{1+1/8})^4 = (\frac{17}{18})^4 2^{-4}$$

$$Q(x^n|y^n) < Q(x^n|y^n, z^n)$$

(Suzuki 2017)

X	Y	Z
0	0	0
1	1	0
0	2	1
1	3	1
1	3	1
0	2	1
1	1	0
0	0	0



Quotient + Jeffreys' satisfies regularity

$$Q(y^n) := \prod_{i=1}^{n} \frac{c_{i-1}(y_i) + 0.5}{i - 1 + 0.5\beta} \qquad Q(x^n, y^n) := \prod_{i=1}^{n} \frac{c_{i-1}(x_i, y_i) + 0.5}{i - 1 + 0.5\alpha\beta} \qquad Q(x^n, y^n, z^n) := \prod_{i=1}^{n} \frac{c_{i-1}(x_i, y_i, z_i) + 0.5}{i - 1 + 0.5\alpha\beta\gamma}$$

$$Q(y^n) = \frac{(3/4)^4}{2 \cdot 3 \cdots 9} \qquad Q(x^n, y^n) = Q(y^n, z^n) = \frac{(3/4)^4}{4 \cdot 5 \cdots 11} \qquad Q(x^n, y^n, z^n) = \frac{(3/4)^4}{8 \cdot 9 \cdots 15}$$

$$\frac{3}{55} = Q(x^n|y^n) = \frac{Q(x^n, y^n)}{Q(y^n)} > \frac{Q(x^n, y^n, z^n)}{Q(y^n, z^n)} = Q(x^n|y^n, z^n) = \frac{1}{39}$$

Theorem 1: $H(x^n|y^n) = 0 \Longrightarrow Q(x^n|y^n) \geq Q(x^n|y^n, z^n)$ for Quotient + Jeffreys

Proof: Appendix A

MDL is regular

$$L(x^n|y^n) = H(x^n|y^n) + \frac{(\alpha-1)\beta}{2n}\log n$$

$$L(x^n|y^n, z^n) = H(x^n|y^n, z^n) + \frac{(\alpha-1)\beta\gamma}{2n}\log n$$

$$H(x^n|y^n) = 0 \Longrightarrow H(x^n|y^n, z^n) = 0 \Longrightarrow L(x^n|y^n) \leq L(x^n|y^n, z^n)$$

$$\Longrightarrow \text{MDL chooses } \{Y\} \text{ rather than } \{Y, Z\}$$

B & B for MDL (Suzuki, ICML-96)

$$L(x^n|y^n) = H(x^n|y^n) + \frac{(\alpha-1)\beta}{2n}\log n$$

$$L(x^n|y^n,z^n) = H(x^n|y^n,z^n) + \frac{(\alpha-1)\beta\gamma}{2n}\log n$$

Pruning Rule:

$$L(x^n|y^n) \le \frac{(\alpha-1)\beta\gamma}{2n}\log n \Longrightarrow L(x^n|y^n) \le L(x^n|y^n,z^n,\cdots)$$

B & B for MDL obtains the lowerbound

by assuming $H(x^n|y^n,z^n)=0$ and utilizes its regularity

B & B for Quotient + Jeffreys' (Proposed)

$$Q(x^n|y^n) = \frac{Q(x^n, y^n)}{Q(y^n)} \quad = \quad \prod_{i=1}^{n} \frac{i-1+0.5\beta}{i-1+0.5\alpha\beta} \cdot \prod_{i=1}^{n} \frac{c_{i-1}(x_i, y_i)+0.5}{c_{i-1}(y_i)+0.5} \, ,$$

$$Q(x^n|y^n, z^n) = \frac{Q(x^n, y^n, z^n)}{Q(y^n, z^n)} \quad = \quad \prod_{i=1}^{n} \frac{i-1+0.5\beta\gamma}{i-1+0.5\alpha\beta\gamma} \cdot \prod_{i=1}^{n} \frac{c_{i-1}(x_i, y_i, z_i)+0.5}{c_{i-1}(y_i, z_i)+0.5} \, ,$$

$$H(x^n|y^n, z^n) = 0 \Longleftrightarrow c_{i-1}(x_i, y_i, z_i) = c_{i-1}(y_i, z_i), \; i = 1, \cdots, n$$

Pruning Rule:

$$Q(x^n|y^n) \geq \prod_{i=1}^{n} \frac{i-1+0.5\beta\gamma}{i-1+0.5\alpha\beta\gamma} \geq Q(x^n|y^n, z^n, \cdots)$$

Quotient + Jeffreys' is close to MDL

$$Q(x^n|y^n) = \prod_{i=1}^{n} \frac{i-1+0.5\beta}{i-1+0.5\alpha\beta} \cdot \prod_{i=1}^{n} \frac{c_{i-1}(x_i, y_i) + 0.5}{c_{i-1}(y_i) + 0.5} \ ,$$

$$L(x^n|y^n) = H(x^n|y^n) + \frac{(\alpha-1)\beta}{2n} \log n$$

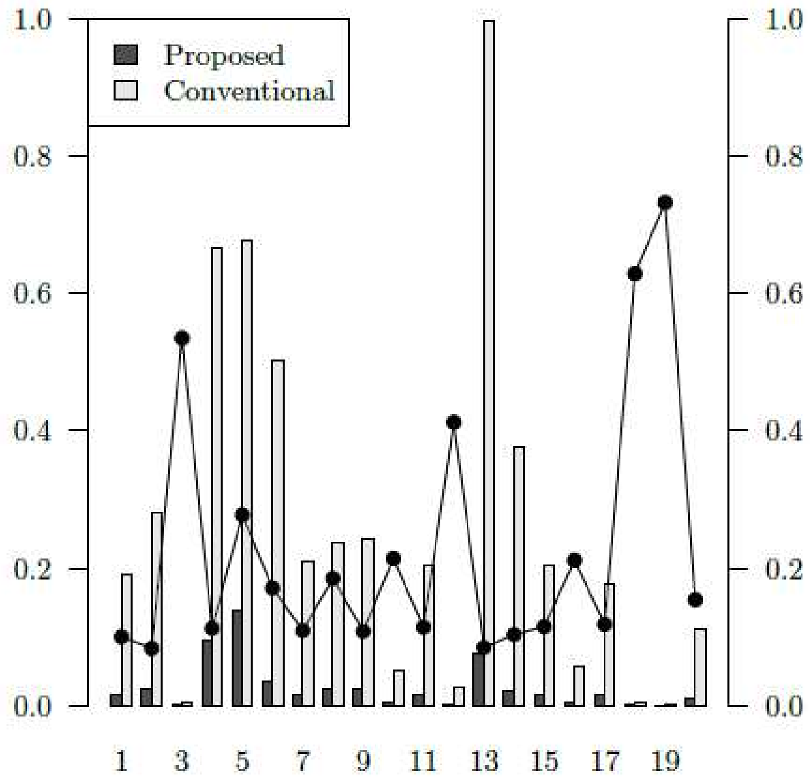
Theorem 4:

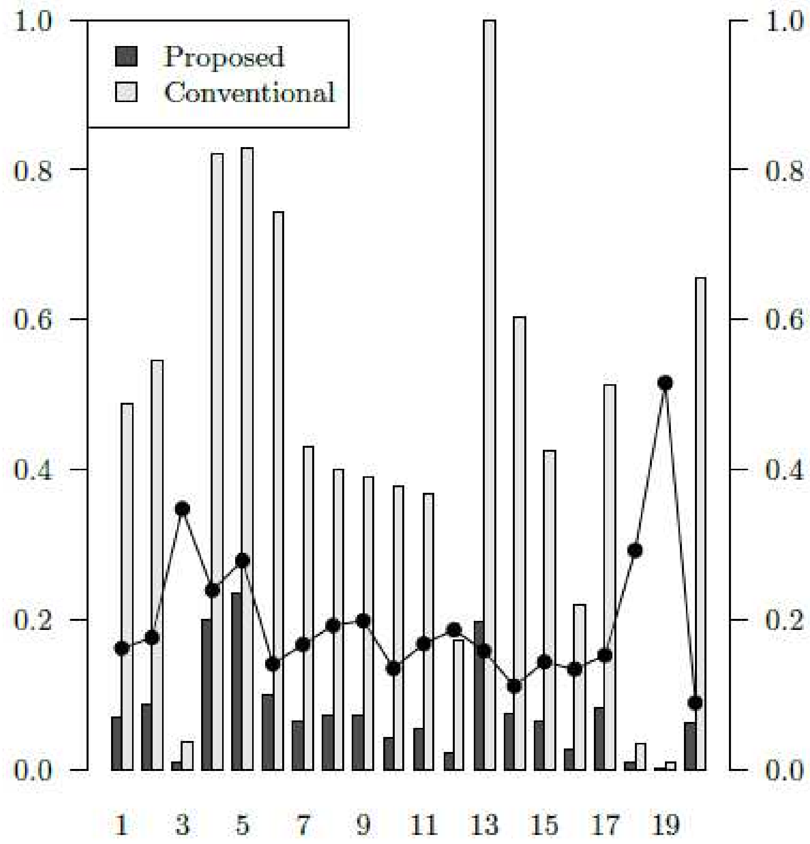
$$-\log\{\prod_{i=1}^{n} \frac{c_{i-1}(x_i, y_i) + 0.5}{c_{i-1}(y_i) + 0.5}\} = nH(x^n|y^n) + O(1)$$

$$-\log\{\prod_{i=1}^{n} \frac{i-1+0.5\beta}{i-1+0.5\alpha\beta}\} = \frac{1}{2}(\alpha-1)\beta \log n + O(1)$$

$$\text{maximizing } Q(x^n|y^n) \iff \text{minimizing } L(x^n|y^n)$$

The proposed B&B is more efficient than the existing one





Regular BNSL offers efficient B&B while BDeu does not