The quantile probability model

A novel approach to Bayesian variable selection for clinical prediction

Rachel Heyard and Leonhard Held

Department of Biostatistics - EBPI - University of Zurich

22.06.2018

Clinical prediction models

Clinical rediction models used for decision making

Clinical prediction models

Clinical rediction models used for decision making

$\Rightarrow$ to identify low and high risk patients for a specific disease / pathogen / event /...

Clinical prediction models

Clinical rediction models used for decision making

- $\Rightarrow$ to identify low and high risk patients for a specific disease / pathogen / event /...
- $\Rightarrow$ to better target treatment and prevention

Clinical prediction models

Clinical rediction models used for decision making

- ⇒ to identify low and high risk patients for a specific disease / pathogen / event /...
- ⇒ to better target treatment and prevention
- ⇒ to improve medical practice

Objective Bayesian variable selection for GLMs

Select among p variables $\rightarrow$ 2^p candidate models $\mathcal{M}_j$

[1] Hu and Johnson. *Bayesian model selection using test statistics*. Journal of the Royal Statisical Society, 2009

[2] Held, Sabanés Bové and Gravestock. *Approximate Bayesian model selection with the deviance statistic*. Statistical Science, 2015

Objective Bayesian variable selection for GLMs

Select among p variables $\rightarrow$ 2^p candidate models $\mathcal{M}_j$

Posterior model probability:

$$\Pr(\mathcal{M}_j \mid \text{data}) = \frac{\text{TBF}_{j,0}\Pr(\mathcal{M}_j)}{\sum_{k \in \mathcal{J}} \text{TBF}_{k,0}\Pr(\mathcal{M}_k)}$$

[1] Hu and Johnson. *Bayesian model selection using test statistics*. Journal of the Royal Statisical Society, 2009

[2] Held, Sabanés Bové and Gravestock. *Approximate Bayesian model selection with the deviance statistic*. Statistical Science, 2015

Objective Bayesian variable selection for GLMs

Select among p variables $\rightarrow 2^p$ candidate models $\mathcal{M}_j$

Posterior model probability:

$$\Pr(\mathcal{M}_j \mid \text{data}) = \frac{\text{TBF}_{j,0}\Pr(\mathcal{M}_j)}{\sum_{k \in \mathcal{J}} \text{TBF}_{k,0}\Pr(\mathcal{M}_k)}$$

Generalized g-prior [1,2] leads to

$$\text{TBF}_{j,0} = \frac{\text{p}(z_j \mid \mathcal{M}_j)}{\text{p}(z_j \mid \mathcal{M}_0)} = (g+1)^{-d_j/2} \exp\left(\frac{g}{g+1}\frac{z_j}{2}\right)$$

z_j is the deviance statistic of model $\mathcal{M}_j$ with d_j degrees of freedom

---

[1] Hu and Johnson. *Bayesian model selection using test statistics*. Journal of the Royal Statisical Society, 2009

[2] Held, Sabanés Bové and Gravestock. *Approximate Bayesian model selection with the deviance statistic*. Statistical Science, 2015

Objective Bayesian variable selection for GLMs

Posterior inclusion probability of x_k, with $k \in \{1, \ldots, p\}$:

$$\Pr(x_k \in \mathcal{M} \mid \text{data}) = \sum_{j \in \mathcal{J} : x_k \in \mathcal{M}_j} \Pr(\mathcal{M}_j \mid \text{data})$$

[3] Barbieri and Berger. *Optimal predictive model selection*. Annals of Statistics, 2004

Objective Bayesian variable selection for GLMs

Posterior inclusion probability of x_k, with $k \in \{1, \dots, p\}$:

$$\Pr(x_k \in \mathcal{M} \mid \text{data}) = \sum_{j \in \mathcal{J} : x_k \in \mathcal{M}_j} \Pr(\mathcal{M}_j \mid \text{data})$$

Median probability model (MPM) [3]:
Include x_k if $\Pr(x_k \in \mathcal{M} \mid \text{data}) \geq 0.5$

[3] Barbieri and Berger. *Optimal predictive model selection*. Annals of Statistics, 2004

Objective Bayesian variable selection for GLMs

Posterior inclusion probability of x_k, with $k \in \{1, \ldots, p\}$:

$$\Pr(x_k \in \mathcal{M} \mid \text{data}) = \sum_{j \in \mathcal{J} : x_k \in \mathcal{M}_j} \Pr(\mathcal{M}_j \mid \text{data})$$

Median probability model (MPM) [3]:
Include x_k if $\Pr(x_k \in \mathcal{M} \mid \text{data}) \geq 0.5$

$\Rightarrow$ MPM = optimal for prediction under the squared error loss when selecting among normal linear models

[3] Barbieri and Berger. *Optimal predictive model selection*. Annals of Statistics, 2004

Objective Bayesian variable selection for GLMs

Posterior inclusion probability of x_k, with $k \in \{1, \ldots, p\}$:

$$\Pr(x_k \in \mathcal{M} \mid \text{data}) = \sum_{j \in \mathcal{J} : x_k \in \mathcal{M}_j} \Pr(\mathcal{M}_j \mid \text{data})$$

Median probability model (MPM) [3]:
Include x_k if $\Pr(x_k \in \mathcal{M} \mid \text{data}) \geq 0.5$

$\Rightarrow$ MPM = optimal for prediction under the squared error loss when selecting among normal linear models

BUT MPM also highly sentitive to prior choices

[3] Barbieri and Berger. *Optimal predictive model selection*. Annals of Statistics, 2004

Prior choices

[4] Scott and Berger. *Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem*. Annals of Statistics, 2010

Prior choices

Model prior of $\mathcal{M}_j$ with p_j variables

[4] Scott and Berger. *Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem*. Annals of Statistics, 2010

Prior choices

Model prior of $\mathcal{M}_j$ with p_j variables

$\Rightarrow$ fixed prior inclusion probability at $q = 1/2$

$\rightarrow$ uniform prior on the model space: $\Pr(\mathcal{M}_j) = 1/2^p$

[4] Scott and Berger. *Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem*. Annals of Statistics, 2010

Prior choices

Model prior of $\mathcal{M}_j$ with p_j variables

$\Rightarrow$ fixed prior inclusion probability at $q = 1/2$

$\rightarrow$ uniform prior on the model space: $\Pr(\mathcal{M}_j) = 1/2^p$

$\Rightarrow$ beta distibution on the prior inclusion probability $q \sim \mathrm{Be}(a, b)$

$$\rightarrow \Pr(\mathcal{M}_j \mid a, b) \quad = \quad \frac{\mathrm{B}(a + p_j, b + p - p_j)}{\mathrm{B}(a, b)}$$

[4] Scott and Berger. *Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem*. Annals of Statistics, 2010

Prior choices

Model prior of $\mathcal{M}_j$ with p_j variables

$\Rightarrow$ fixed prior inclusion probability at $q = 1/2$

$\rightarrow$ uniform prior on the model space: $\Pr(\mathcal{M}_j) = 1/2^p$

$\Rightarrow$ beta distibution on the prior inclusion probability $q \sim \mathrm{Be}(a, b)$

$$\rightarrow \Pr(\mathcal{M}_j \mid a, b) \quad = \quad \frac{\mathrm{B}(a + p_j, b + p - p_j)}{\mathrm{B}(a, b)}$$

$a = b = 1 \rightarrow$ multiplicity-corrected model prior [4]

[4] Scott and Berger. *Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem*. Annals of Statistics, 2010

Prior choices

Model prior of $\mathcal{M}_j$ with p_j variables

$\Rightarrow$ fixed prior inclusion probability at $q = 1/2$

$\rightarrow$ uniform prior on the model space: $\Pr(\mathcal{M}_j) = 1/2^p$

$\Rightarrow$ beta distibution on the prior inclusion probability $q \sim \text{Be}(a, b)$

$$\rightarrow \Pr(\mathcal{M}_j \mid a, b) = \frac{\text{B}(a + p_j, b + p - p_j)}{\text{B}(a, b)}$$

$a = b = 1 \rightarrow$ multiplicity-corrected model prior [4]

Definition of g

[4] Scott and Berger. *Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem*. Annals of Statistics, 2010

Prior choices

Model prior of $\mathcal{M}_j$ with p_j variables

$\Rightarrow$ fixed prior inclusion probability at $q = 1/2$

$\rightarrow$ uniform prior on the model space: $\Pr(\mathcal{M}_j) = 1/2^p$

$\Rightarrow$ beta distibution on the prior inclusion probability $q \sim \mathrm{Be}(a, b)$

$$\rightarrow \Pr(\mathcal{M}_j \mid a, b) \quad = \quad \frac{\mathrm{B}(a + p_j, b + p - p_j)}{\mathrm{B}(a, b)}$$

$a = b = 1 \rightarrow$ multiplicity-corrected model prior [4]

Definition of g

$\Rightarrow$ Local empirical Bayes: $\hat{g} = \max\{z_j/d_j - 1, 0\}$

$\Rightarrow$ hyper-g/n: $\frac{g/n}{g/n+1} \sim \mathrm{U}(0, 1)$

$\Rightarrow$ Zellner-Siow adapted

[4] Scott and Berger. *Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem*. Annals of Statistics, 2010

Case study

Application on **GUSTO-I** data using logistic regression

$\Rightarrow$ Randomized study for comparison of 4 different treatments in over 40‘000 acute myocardial infarction patients

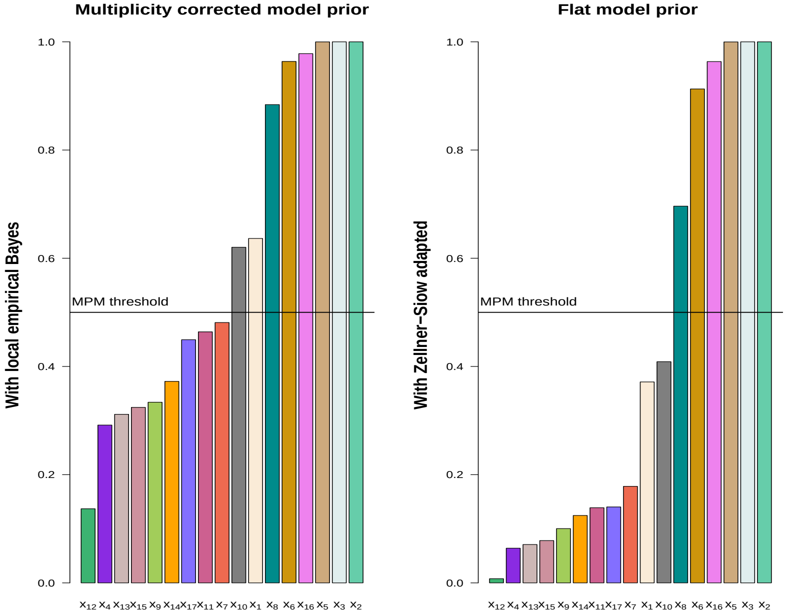
$\Rightarrow$ We use a publicly available subgroup from the Western region of the USA ($n = 2188$)

$\Rightarrow$ prediction of **30-days survival**

We focus on **17 covariates**

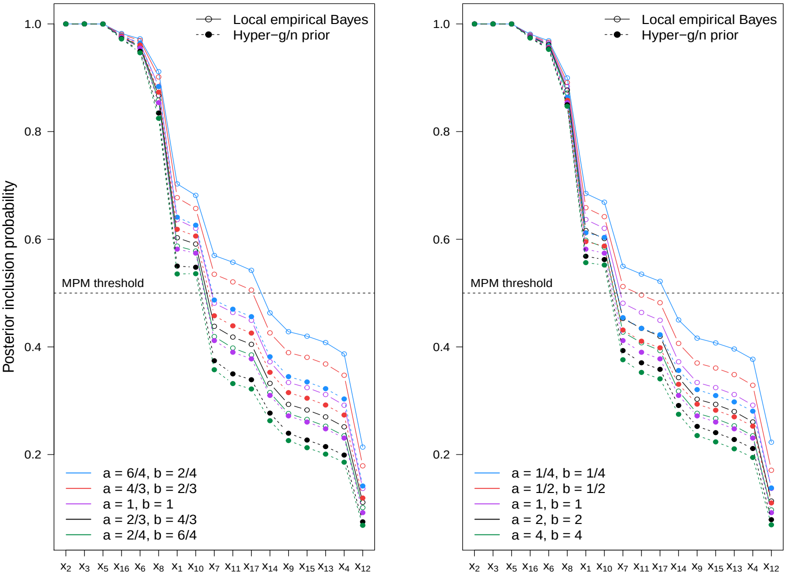
Sensitivity of the MPM

with respect to the definition of *g* and the model prior



Sensitivity of the MPM

with respect to the hyper-parameters *a* and *b*



The quantile probability model

The quantile probability model

- Define $p + 1$ candidate models $\{\mathcal{M}_0, \mathcal{M}_1, \ldots, \mathcal{M}_p\}$ based on inclusion probabilities

The quantile probability model

- Define $p + 1$ candidate models $\{\mathcal{M}_0, \mathcal{M}_1, \ldots, \mathcal{M}_p\}$ based on inclusion probabilities

- Compute predictive model selection criterion:
  DIC, WAIC or LOO-IC

The quantile probability model

- Define $p+1$ candidate models $\{\mathcal{M}_0, \mathcal{M}_1, \dots, \mathcal{M}_p\}$ based on inclusion probabilities
- Compute predictive model selection criterion:

  DIC, WAIC or LOO-IC
- Select the model with lowest information criterion $\rightarrow$ QPM

Predictive model selection information criteria (IC)

The **deviance information criterion** (DIC):

$$-2 \log f(\mathbf{y} \mid \hat{\theta}) \quad + \quad 2p_{\text{DIC}}$$

[5] Gelman, Hwang and Vehtari. *Understanding predictive information criteria for Bayesian models*. Statistics and Computing, 2014

Predictive model selection information criteria (IC)

The **deviance information criterion** (DIC):

$$
\begin{array}{ccc}
-2 \log f(\mathbf{y} \mid \hat{\theta}) & + & 2p_{\text{DIC}} \\
\text{[adjustement of the log pred. density]} & + & \text{[approximate bias correction]}
\end{array}
$$

[5] Gelman, Hwang and Vehtari. *Understanding predictive information criteria for Bayesian models*. Statistics and Computing, 2014

Predictive model selection information criteria (IC)

The **deviance information criterion** (DIC):

$$-2 \log f(\mathbf{y} \mid \hat{\theta}) \quad + \quad 2p_{\text{DIC}}$$

$$[\text{adjustement of the log pred. density}] \quad + \quad [\text{approximate bias correction}]$$

Similarly, Watanabe-Akaike and leave-one-out cross-validation IC (WAIC and LOO-IC) can be defined [5]

[5] Gelman, Hwang and Vehtari. *Understanding predictive information criteria for Bayesian models*. Statistics and Computing, 2014

Predictive model selection information criteria (IC)

The **deviance information criterion** (DIC):

$$-2 \log f(\mathbf{y} \mid \hat{\theta}) \quad + \quad 2p_{\text{DIC}}$$

[adjustement of the log pred. density] + [approximate bias correction]

Similarly, Watanabe-Akaike and leave-one-out cross-validation IC (WAIC and LOO-IC) can be defined [5]

→ computation possible with `loo`-package in `R`

[5] Gelman, Hwang and Vehtari. *Understanding predictive information criteria for Bayesian models*. Statistics and Computing, 2014

Predictive model selection information criteria (IC)

The **deviance information criterion** (DIC):

$$\begin{array}{ccc} -2 \log f(\mathbf{y} \mid \hat{\theta}) & + & 2p_{\text{DIC}} \\ [\text{adjustement of the log pred. density}] & + & [\text{approximate bias correction}] \end{array}$$

Similarly, Watanabe-Akaike and leave-one-out cross-validation IC (WAIC and LOO-IC) can be defined [5]

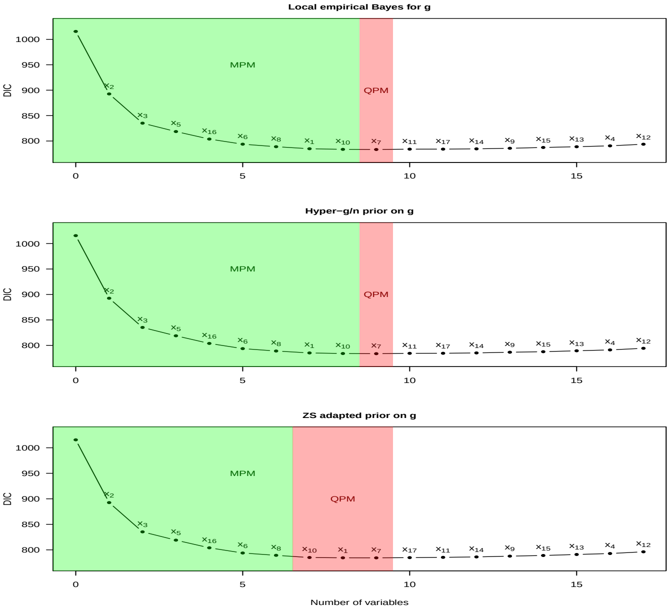
→ computation possible with `loo`-package in `R`

→ IC does not depend on model prior !

[5] Gelman, Hwang and Vehtari. *Understanding predictive information criteria for Bayesian models*. Statistics and Computing, 2014

The quantile probability model

QPM model $\mathcal{M}_9$ independent from the definition of g



Monte Carlo error of DIC

Sampling from the posterior distribution $\rightarrow$ **Monte Carlo** (MC) **error**

Monte Carlo error of DIC

Sampling from the posterior distribution → **Monte Carlo** (MC) **error**

Brute force approach:

Replicate the DIC calculation N times and compute sample standard deviation of the N DIC estimates

Monte Carlo error of DIC

Sampling from the posterior distribution → **Monte Carlo** (MC) **error**

Brute force approach:

Replicate the DIC calculation N times and compute sample standard deviation of the N DIC estimates

⇒ very time-consuming

New efficient approach for Monte Carlo error of DIC

New efficient approach for Monte Carlo error of DIC

Suppose DIC has been computed using MC sample of size S

New efficient approach for Monte Carlo error of DIC

Suppose DIC has been computed using MC sample of size *S*

New efficient approach for Monte Carlo error of DIC

Suppose DIC has been computed using MC sample of size S

New efficient approach for Monte Carlo error of DIC

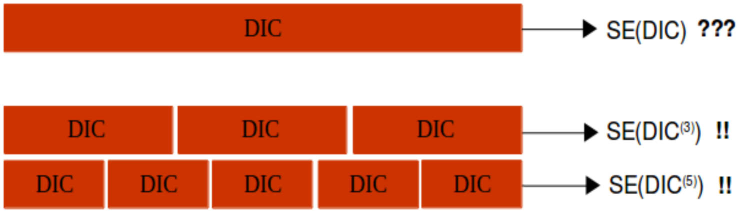
Suppose DIC has been computed using MC sample of size S

New efficient approach for Monte Carlo error of DIC

Suppose DIC has been computed using MC sample of size S

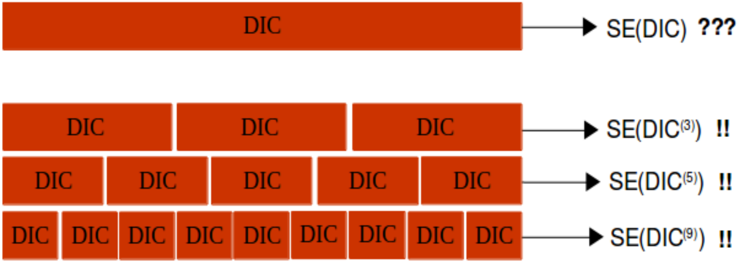
New efficient approach for Monte Carlo error of DIC

Suppose DIC has been computed using MC sample of size *S*



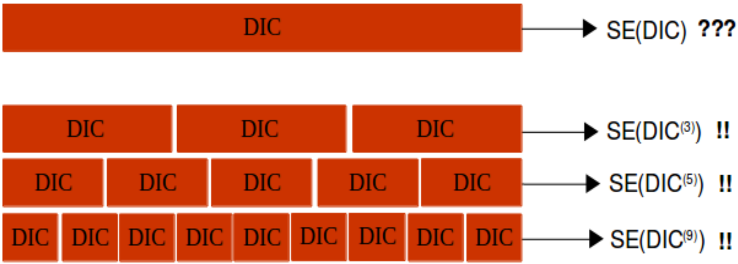
New efficient approach for Monte Carlo error of DIC

Suppose DIC has been computed using MC sample of size S



New efficient approach for Monte Carlo error of DIC

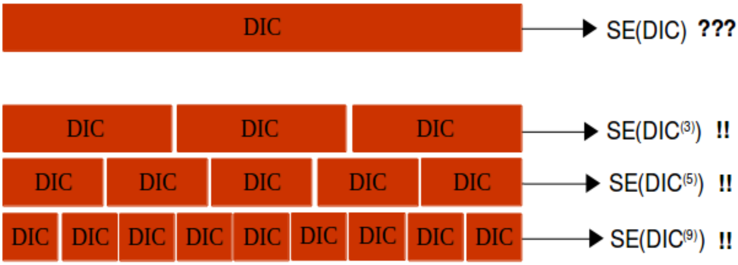
Suppose DIC has been computed using MC sample of size S



with the square-root law we obtain $\mathrm{SE}(\mathrm{DIC}^{(k)}) \propto \sqrt{\frac{1}{S/k}} \propto \sqrt{k}$

New efficient approach for Monte Carlo error of DIC

Suppose DIC has been computed using MC sample of size S



with the square-root law we obtain $\mathrm{SE}(\mathrm{DIC}^{(k)}) \propto \sqrt{\frac{1}{S/k}} \propto \sqrt{k}$

$\Rightarrow \mathrm{SE}(\mathrm{DIC}^{(k)}) = c \cdot \sqrt{k} \quad \Rightarrow \quad \mathrm{SE}(\mathrm{DIC}^{(1)}) = c$

estimate c with weighted linear regression of $\mathrm{SE}(\mathrm{DIC}^{(k)})$ on $\sqrt{k}$ with weights k

Monte Carlo error of DIC

Application to case study

Case study: $S = 50,000$ and $k \in \{80, 100, 125, 160, 200\}$

Monte Carlo error of DIC

Application to case study

Case study: $S = 50,000$ and $k \in \{80, 100, 125, 160, 200\}$

The **Monte Carlo standard error** of $\Delta\text{DIC}_j = \text{DIC}_{j-1} - \text{DIC}_j$:

$$\text{SE}(\Delta\text{DIC}_j) = \sqrt{\text{SE}(\text{DIC}_{j-1})^2 - \text{SE}(\text{DIC}_j)^2}$$

Monte Carlo error of DIC

Application to case study

Case study: $S = 50,000$ and $k \in \{80, 100, 125, 160, 200\}$

The **Monte Carlo standard error** of $\Delta\text{DIC}_j = \text{DIC}_{j-1} - \text{DIC}_j$:

$$\text{SE}(\Delta\text{DIC}_j) = \sqrt{\text{SE}(\text{DIC}_{j-1})^2 - \text{SE}(\text{DIC}_j)^2}$$

	LEB		hyper-g/n		ZS adapted	
	ΔDIC_j	(SE)	ΔDIC_j	(SE)	ΔDIC_j	(SE)
			⋮			
$\mathcal{M}_8$ vs. $\mathcal{M}_9$	0.217	(0.031)	0.166	(0.034)	0.057	(0.033)
$\mathcal{M}_9$ vs. $\mathcal{M}_{10}$	-0.295	(0.031)	-0.342	(0.034)	-0.170	(0.032)
			⋮			

Monte Carlo error of DIC

Application to case study

Case study: $S = 50,000$ and $k \in \{80, 100, 125, 160, 200\}$

The **Monte Carlo standard error** of $\Delta\text{DIC}_j = \text{DIC}_{j-1} - \text{DIC}_j$:

$$\text{SE}(\Delta\text{DIC}_j) = \sqrt{\text{SE}(\text{DIC}_{j-1})^2 - \text{SE}(\text{DIC}_j)^2}$$

	LEB		hyper-g/n		ZS adapted	
	ΔDIC_j	(SE)	ΔDIC_j	(SE)	ΔDIC_j	(SE)
			⋮			
$\mathcal{M}_8$ vs. $\mathcal{M}_9$	0.217	(0.031)	0.166	(0.034)	0.057	(0.033)
$\mathcal{M}_9$ vs. $\mathcal{M}_{10}$	-0.295	(0.031)	-0.342	(0.034)	-0.170	(0.032)
			⋮			

Conclusion: The DIC of the QPM $\mathcal{M}_9$ is the unique minimum !

Conclusion

- Drastically reduced candidate model space: 2^p to $p + 1$

Conclusion

- Drastically reduced candidate model space: 2^p to $p+1$
- In the application: QPM not sensitve with respect to prior choices

Conclusion

- Drastically reduced candidate model space: 2^p to $p + 1$
- In the application: QPM not sensitve with respect to prior choices
- Model prior does not affect the information criterion

Conclusion

- Drastically reduced candidate model space: 2^p to $p + 1$
- In the application: QPM not sensitve with respect to prior choices
- Model prior does not affect the information criterion
- Efficient method introduced to compute Monte Carlo error of IC

Conclusion

- Drastically reduced candidate model space: 2^p to $p + 1$
- In the application: QPM not sensitve with respect to prior choices
- Model prior does not affect the information criterion
- Efficient method introduced to compute Monte Carlo error of IC

- Work revised for *Journal of Computational Statistics and Data Analysis*

Thank you

Questions? Comments?