Parallel MCMC

multi-core (single machine)

AMS 206
22 Mar 19

①

distributed systems

GPU / TPU

Bayesian model averaging

length k_j

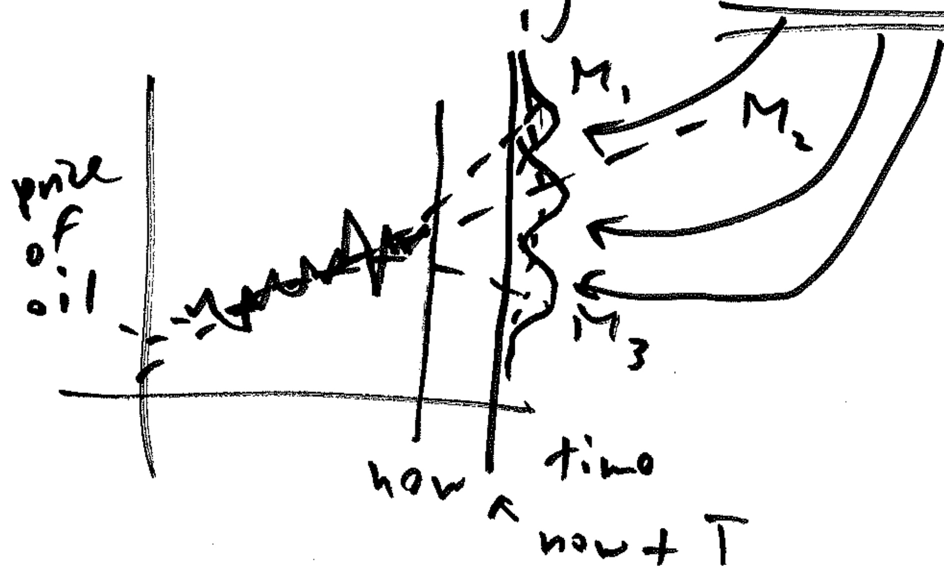
$$M_j : \begin{cases} (\underset{\sim}{\theta}_j \mid \ldots \mathcal{B}) \sim p(\underset{\sim}{\theta}_j \mid \ldots \mathcal{B}) \\ (Y_i \mid \underset{\sim}{\theta}_j \ldots \mathcal{B}) \stackrel{IID}{\sim} p(y_i \mid \underset{\sim}{\theta}_j \ldots \mathcal{B}) \end{cases}$$

$j = 1, \ldots, m$ $\quad y = (y_1, \ldots, y_n)$

$$\mathbb{M} = \{ M_1, \ldots, M_m \}$$

$$p(y^* \mid y\, \mathbb{M}\, \mathcal{B}) = \sum_{j=1}^{m} p(y^*, M_j \mid y\, \mathbb{M}\, \mathcal{B})$$

$$= \sum_{j=1}^{m} \underline{p(y^* \mid y\, M_j\, \mathcal{B})}\, p(M_j \mid y\, \mathbb{M}\, \mathcal{B})$$



Draper (1995)
Raftery & co.

(2)

$$BIC(M_j \mid y\,\mathcal{B}) = -2\,\ell\ell(\hat{\theta}_j \mid y\,M_j\,\mathcal{B}) + k_j \log(n)$$

$$AIC(M_j \mid y\,\mathcal{B}) = -2\,\ell\ell(\hat{\theta}_j \mid y\,M_j\,\mathcal{B}) + 2k_j$$

same

not same

warning: neither BIC nor AIC works for model comparison if one or more models have random effects in them

ex.
$(\mu\,\sigma \mid \mathcal{B}) \sim p(\mu\,\sigma)$
$(\theta_i \mid \mu\,\sigma\,\mathbb{N}\,\mathcal{B}) \overset{IID}{\sim} N(\mu, \sigma^2)$
$(Y_i \mid \theta_i\,V_i\,\mathcal{B}) \overset{I}{\sim} N(\theta_i, V_i)$
$(i = 1, \dots, k)$

q: how many parameters in this model?

A_1 (naively) $k = 2$ (μ, σ)

A_2: but what about the random effects?

model complexity here $\neq$ # parameters (because random effects θ_i are also unknown)

$\boxed{Q:}$ when AIC & BIC differ in choice of best model, which one do you believe, or should you not believe either?

2 goals in defining "good" model comparison method

$\boxed{\text{goal 1}}$ pretend data had to come from one of (say M_j)

$\mathbb{M} = \{ M_1, \dots, M_m \}$; "true" model

a model comparison method (MC) is called consistent if P_F(MC picks M_j) $\uparrow 1$ as $n \uparrow \infty$

$\boxed{\text{fact:}}$ BIC was designed to be consistent. (this is inference

goal 2 M_2 is better than M_1 if M_2 makes better predictions of future data than M_1
(this is prediction)

as $n \uparrow \infty$

AIC picks the best-predicting model more often than BIC does, but AIC is not (necessarily) consistent.

approach 1: conclude that M_2 is better than M_1 if $BIC(M_2) < BIC(M_1)$

$$-2\,\ell\ell(\hat{\theta}_1) + k_1 \log(n) < -\ell\ell(\hat{\theta}_2) + k_2 \log(n)$$

approach 2: conclude that M_2 is better than M_1 if $AIC(M_2) < AIC(M_1)$

$$-2\,\ell\ell(\hat{\theta}_1) + 2k_1 < -2\,\ell\ell(\hat{\theta}_2) + 2k_2$$

if $\log(n) > 2$ (i.e., $n > e^2 = 7.4$), $n \geq 8$

ln

complexity penalty for BIC bigger than for AIC; i.e., if $n \geq 8$ BIC will choose simpler model than AIC

~~meta~~ fact: ~~(conjecture)~~

optimal prediction requires (slightly) more complicated models than optimal inference about a hypothetical "true" model