this time: MCMC

next time: hierarchical models

read: (G) ch. (5)

one widely-used diagnostic for comparing Bayesian models:

the **D**eviance **I**nformation **C**riterion (DIC)

Spiegelhalter et. al. (2002)

$$(\underset{\sim}{\theta}_j \mid \mathcal{B}) \sim p(\underset{\sim}{\theta}_j \mid \mathcal{B})$$

prior dist.

$$(y_i \mid \underset{\sim}{\theta}_j \, M_j \, \mathcal{B}) \overset{IID}{\sim} p(y_i \mid \underset{\sim}{\theta}_j \, M_j \, \mathcal{B})$$

$\in \mathbb{R}$

← model j $(i = 1, \dots, n)$

Compute log-likelihood function as usual:

let $\underset{\sim}{y} = (y_1, \dots, y_n)$

$$\ell\ell(\underset{\sim}{\theta}_j \mid \underset{\sim}{y} \, M_j \, \mathcal{B})$$

we want this to be big if model M_j fits the data well

define the deviance

$$D(\underset{\sim}{\theta}_j \mid \underset{\sim}{y} \, M_j \, \mathcal{B}) = -2 \, \ell\ell(\underset{\sim}{\theta}_j \mid \underset{\sim}{y} \, M_j \, \mathcal{B})$$

since we want $\ell\ell$ to be big, we want D to be small if M_j is good

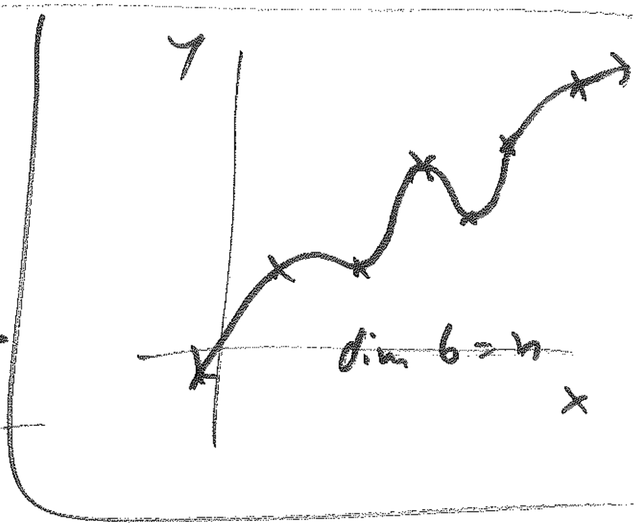
problem: you can make D (almost) arbitrarily small, by building a model with so many parameters (2)

$\underset{\sim}{\theta}_{k_j}$

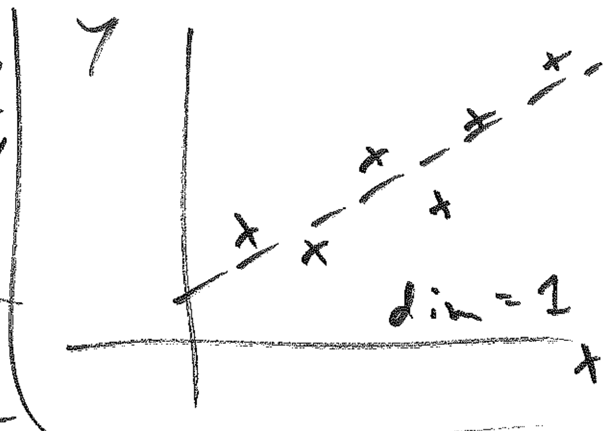
k_j = model dimension of M_j

that you fit the current data set (almost) perfectly, $\boxed{\text{but}}$ such a model will perform badly on new data because it is <u>overfit</u> to $\underset{\sim}{y}_n$

solution: <u>penalize</u> the log likelihood / deviance by a term that captures model complexity



$$\underline{DIC}\,(M_j \mid \underset{\sim}{y}\, \mathcal{B}) \triangleq \underbrace{\overset{\text{want small}}{\bar{D}_j}}_{\text{model } j \text{ fit}} + \overset{\text{also want small}}{(p_D)_j}$$



$\bar{D}_j$ = mean deviance value with respect to the posterior ~~for~~ $\underset{\sim}{\theta}_j$ for model M_j

$(p_D)_j$ = estimate of effective <u>complexity</u> of model M_j

if M_j is not hierarchical (more about this soon) $(p_D)_j$ will just be $= k_j$ (length of $\underset{\sim}{\theta}_j$ vector) 3

if M_j is hierarchical, need to do something fancier to estimate its complexity

2 main methods for estimating $(p_D)_j$

$(p_D)_j \geq 0$

① (winBUGS)

$$(\hat{p}_D)_j = \bar{D}_j - D(\underset{\sim}{\bar{\theta}}_j \mid y\, M_j\, \mathcal{B})$$

(mean deviance) − (deviance evaluated at posterior mean of $\underset{\sim}{\theta}_j$) (2002)

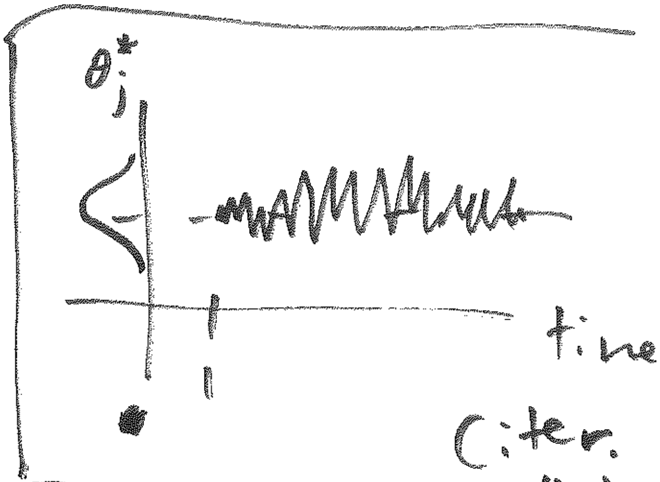
(this $(\hat{p}_D)_j$ can be negative: bad)

② $$(\hat{p}_D)_j = \frac{1}{2} \underbrace{\hat{V}\left[D(\underset{\sim}{\theta}_j \mid y\, M_j\, \mathcal{B}) \right]}_{\text{variance of deviance value with respect to the posterior for } \underset{\sim}{\theta}_j} \geq 0$$

(rjags) (2004)

$$\left[\;\cdots\; \theta_j \;\cdots\;\right]$$ ↑ initial, burn-in discarded

$$m \left[\begin{array}{c} (\theta_j)_1^* \\ \vdots \\ (\theta_j)_m^* \end{array}\right]$$ monitoring



mean $\bar{\theta}_j^* = \frac{1}{m}\sum_{i=1}^{m} (\theta_j^*)_i$

SD $\hat{\sigma}_j$

if $(\theta_j^*) \sim AR_1(\hat{\rho})$

then

$$MCSE(\bar{\theta}_j^*) = \underbrace{\frac{\hat{\sigma}_j}{\sqrt{m}}}_{IID}\;\underbrace{\sqrt{\frac{1+\hat{\rho}}{1-\hat{\rho}}}}$$

if $\rho = 0$ (IID) MCMC = MC ✓

if chain not mixing well, $\hat{\rho}$ will be close to +1

$\sqrt{\frac{1+\hat{\rho}}{1-\hat{\rho}}}$ will get really big!

IHGA case study

(C)		(T) = (E)	
0 ⋮ 0 / 1 ⋮ 1 ⋮	$n_C = 287$	0 ⋮ 0 / 1 ⋮ 1 ⋮	$n_T = 285$

MCMC accuracy can be orders of magnitude smaller than IID accuracy

The difference $\frac{\bar{y}_T - \bar{y}_C}{\bar{y}_C} = -0.19$ is

large in practical terms (is practically significant (practsig))

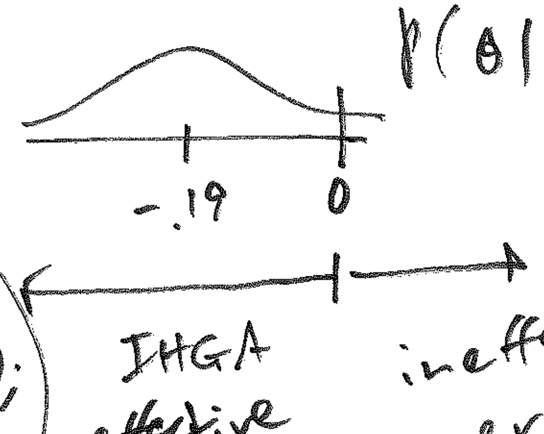
Q: Is this difference large in statistical terms (statistically significant (statsig))?

p	pop.
μ_C	
μ_T	

devil's advocate: $\frac{\mu_T - \mu_C}{\mu_C} = 0$

θ

$p(\theta \mid D\, \mathcal{B})$



we compute $p(\theta < 0 \mid D\, \mathcal{B})$; if this is big enough, (> .95) statsig