8 Mar 13 (extra notes)

①

your model for your uncertainty about θ:

$$\{ p(\theta \mid \mathcal{B}),\ p(y \mid \theta \mathcal{B}),\ \mathcal{A},\ U(a, \theta) \} = M$$

real-world problem $\mathcal{P}$ $\xrightarrow{?}$ M

(not $\mathcal{P}$ = population)

except in rare cases, this mapping is not unique, even among reasonable people

how characterize good mappings?

how tell when m_1 (result of mapping 1) is better than m_2? ②

θ = pop. mean
QofC at DH
for HA
patients in
2012

want (*) $p(\theta \mid \mathcal{D}\, \mathcal{B})$ ← only by context & design

$\mathcal{D} = (y_1, \ldots, y_n)$

not

(**) $p(\theta \mid \mathcal{D}\, \mathcal{B}) \quad \mathcal{B} = \{ (y_i \mid \mu, \sigma^2) \overset{iid}{\sim} N(\mu, \sigma^2),\ \mu, \sigma^2 \text{ unknown} \})$

$\theta = \mu$

$$y_i = \beta_0 + \underbrace{\beta_1 x_{i1} + \dots + \beta_k x_{ik}} + e_i \quad (i = 1, \dots, n)$$

$$e_i \overset{iid}{\sim} N(0, \sigma_e^2)$$

$$\underset{n}{\begin{bmatrix} Y \end{bmatrix}} = \begin{bmatrix} 1 & \uparrow & & \uparrow \\ \vdots & x_1 & \cdots & x_k \\ 1 & \downarrow & & \downarrow \end{bmatrix} \cdot \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$$

$$\underset{n \times 1}{Y} = \underset{n \times (k+1)}{X} \; \underset{(k+1) \times 1}{\beta} + \underset{n \times 1}{e}$$

$$\underset{(k+1) \times 1}{\beta} = \underset{(k+1) \times p}{\boxed{Z}} \; \underset{p \times 1}{\gamma} + \underset{(k+1) \times 1}{\epsilon}$$

which x's are likely to have big (small) coeff.

hierarchical regression model

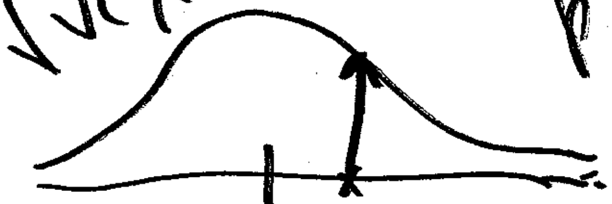
priors

(continuous alternative to (discrete) BMA

④

SD $\sqrt{V(y_0 | x \mathcal{B})}$

$p(y_0 | D, M_j, \mathcal{B})$ $D = y = (y_1, \dots, y_n)$



y_0

$E(y_0 | D \mathcal{B})$

z

$$\frac{y_0 - E}{\sqrt{V}} = z_0$$

$$LS_{FS}(M_j | D \mathcal{B}) = \frac{1}{n} \sum_{i=1}^{n} \log p(y_i | D \mathcal{B})$$

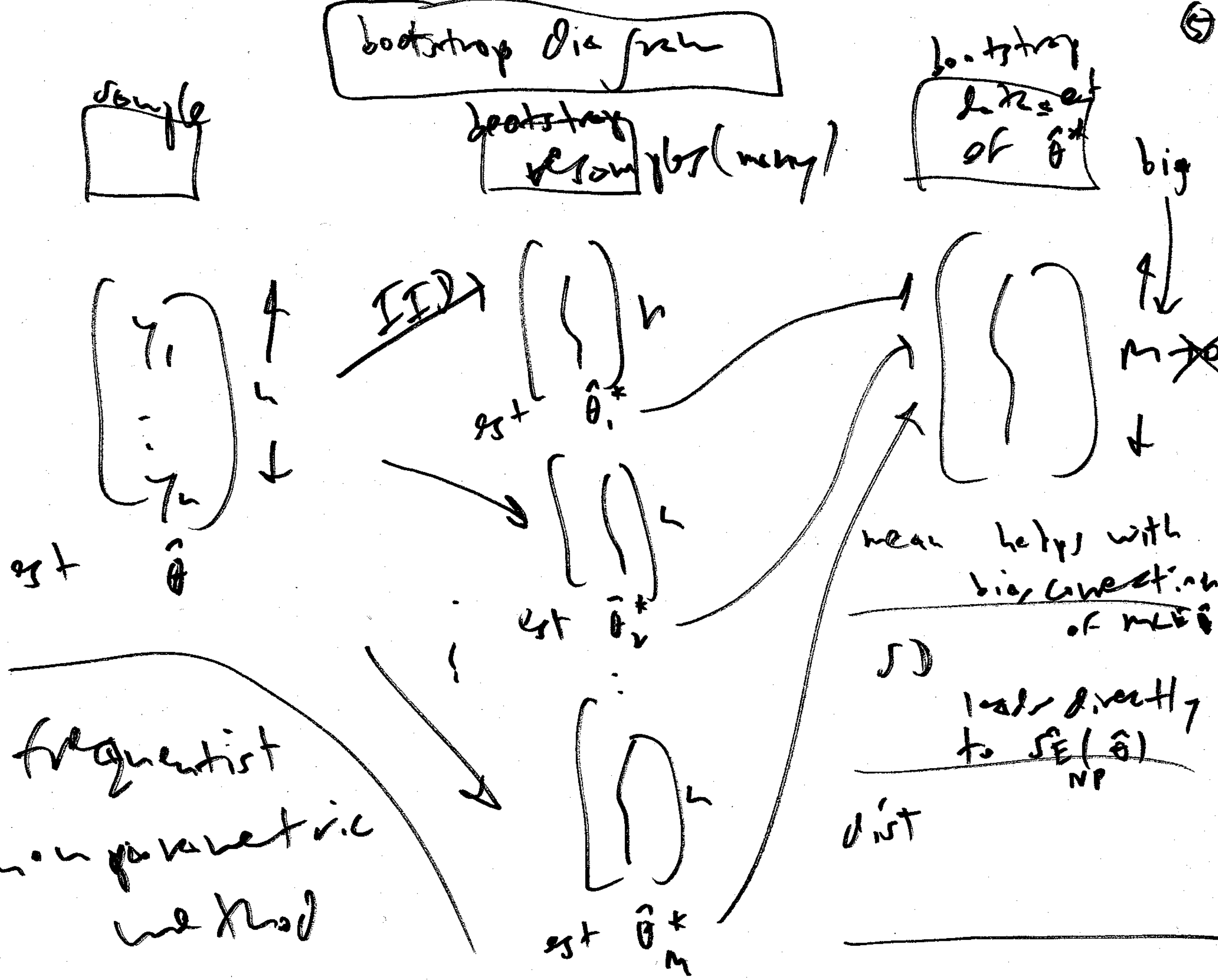
M_2 is better than M_1 if

$LS(M_2) > LS(M_1)$

$p(y_0 | D \mathcal{B} M_3)$

overfit

bootstrap diagram



mean helps with bias correction of MLE $\hat{\theta}$

SD leads directly to $\widehat{SE}_{NP}(\hat{\theta})$

dist

frequentist nonparametric method

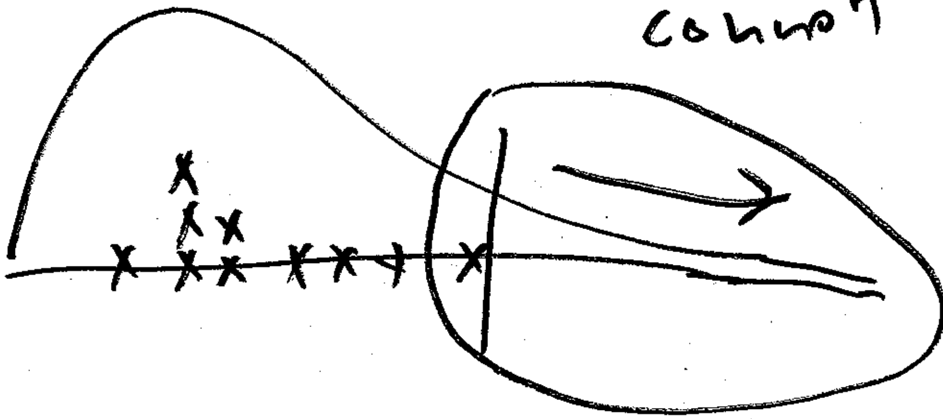
$(Y_i \mid \theta) \overset{iid}{\sim} p(y_i \mid \theta)$

$$p(y_i \mid \theta) = \frac{p(y_i, \theta)}{p(\theta)}$$

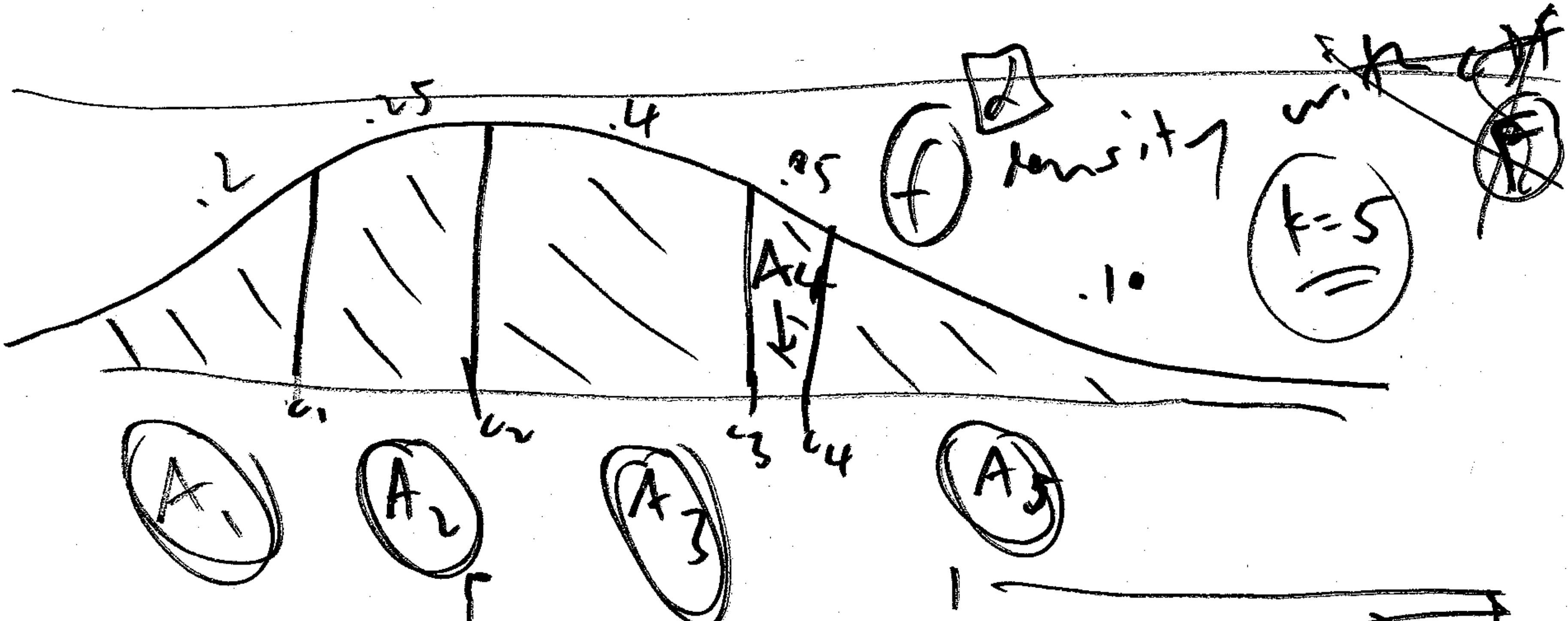
$$\cancel{p(\theta)} \quad \boxed{p(y_1, \ldots, y_n, \theta)} = p(\theta)\, p(y_i \mid \theta)$$

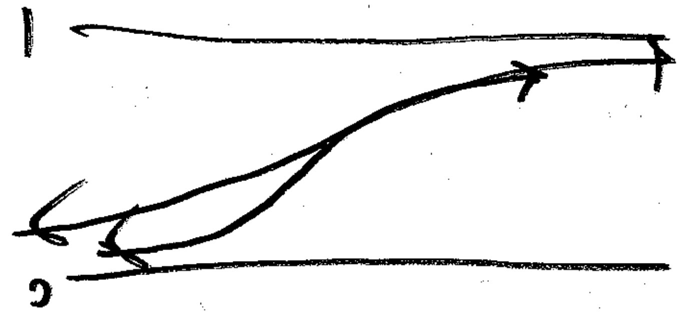
(MCMC slow) Bayesian: learn about $p(\theta \mid y)$ by perturbing θ in

boot (faster) freq (bootstrap): learn about $p(\theta \mid y)$ by perturbing y_i in



cannot give values in bootstrap samples > max





V_1 V_2 beta(1,1)

beta(10,10)

0 W_1 W_2 1

$\frac{1}{2}$

$A = \{\text{all } \alpha_\epsilon\}$

center pt at

$N(0,1)$

$C_0 \sim \text{beta}(\alpha_0, \alpha_1)$ $C_1 = 1 - C_0$

0 (left)

1 (right)

2^1 B_0 $\cancel{2}$ $\cancel{4}$ B_1

$C_{10} \sim \text{beta}(\alpha_{10}, \alpha_{11})$ $C_{11} = 1 - C_{10}$

2^2 $\cancel{4}$ $\cancel{4}$ $\boxed{0}$ $\cancel{4}$ $\cancel{4}$

B_{00} $\Phi^{-1}(.25)$ $-.67$ B_{01} 0 B_{10} $\Phi^{-1}(.75)$ $+.67$ B_{11}

etc.!

how many layers? call this integer M

choose m finite (popular choice $m = 6$ → $2^6 = 64$ partition sets at bottom

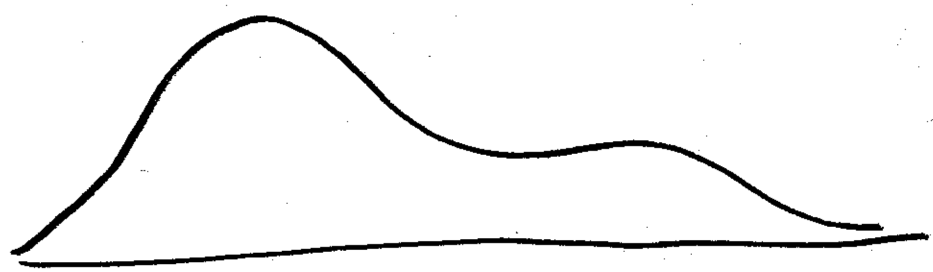
I(x_i's obs comes from component 1 or 2)

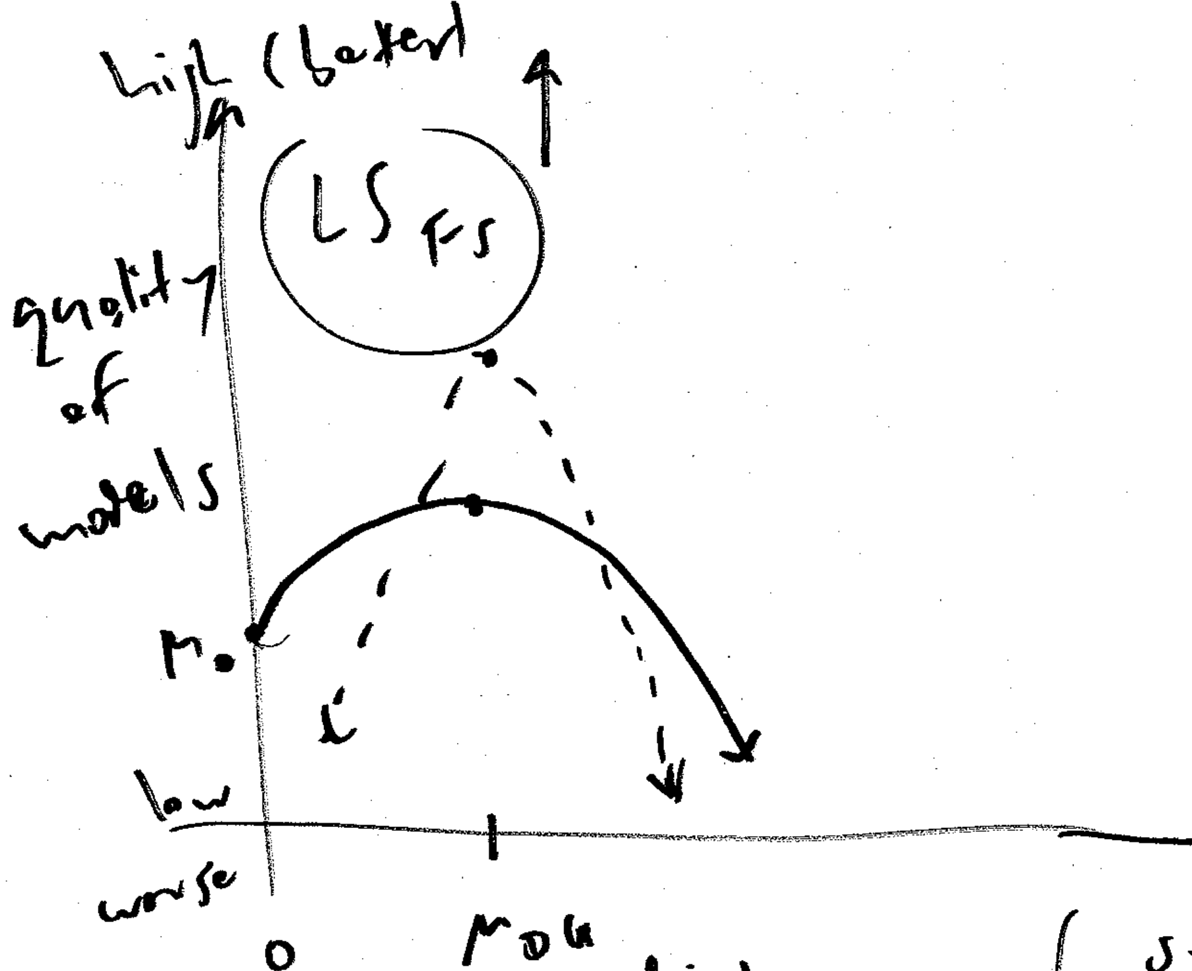
$$\begin{bmatrix} d_1 & T_1 \\ d_2 & T_2 \\ \vdots & \vdots \\ d_N & T_N \end{bmatrix}$$

latent group indicators

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \qquad (Y_i \mid F) \overset{iid}{\sim} F \, ?$$

$$M_0 : \left\{ \begin{array}{l} (\mu, \sigma^2) \sim \text{diffuse} \\ (Y_i \mid \mu, \sigma^2) \overset{iid}{\sim} N(\mu, \sigma^2) \end{array} \right.$$





distance between $\cancel{\text{models}}$ M_0 & M_1

(symmetrized Kullback-Leibler divergence)

$$M_0 : Y_i \overset{iid}{\sim} \underline{\underline{N}}(0,1)$$

11

$$M_1 : \left\{ (Y_i \mid \underline{\underline{\mu}}) \overset{iid}{\sim} N(\mu, 1) \right\} \quad \mu \sim \text{diffuse}$$

$$M_2 : \left\{ (Y_i \mid \mu, \sigma^2) \overset{iid}{\sim} N(\mu, \sigma^2) \right\} \quad (\mu, \sigma) \sim \text{diffuse}$$