Extra notes (1 hr)

1

$$y_{ij} = \beta_0 + u_j + e_{ij}$$

School ($J = 48$)

pupil

ess = 887

$u_j \overset{iid}{\sim} N(0, \sigma_u^2)$ ← ess = 48

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

$N = 887$ → ess = 887

(~~well-calibrated~~)

good diffuse priors

not difficult

β_0 : $N(0,$ huge variance$)$ ← tiny precision e.g. 1.0E-6

σ_e^2 : $\Gamma(\epsilon, \epsilon)$ on $\frac{1}{\sigma_e^2}$ (precision), $\epsilon = .001$

σ_u^2 :

BD. $\Gamma(\epsilon, \epsilon)$ on $\frac{1}{\sigma_u^2}$

→ BD. $U(0, \frac{1}{\epsilon})$ on σ_u^2

Gelman → $U(0, C)$ on σ_u ← just big enough to avoid truncation of likelihood

ess = effective sample size

②

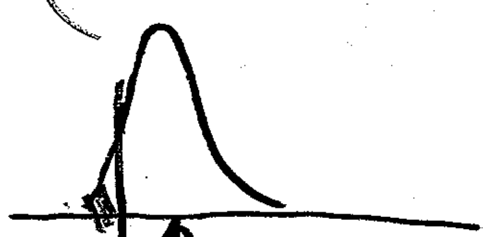
$Y = \hat{y} + \hat{e}$ ← residuals (estimated errors)

observed data

predicted data

if $\theta \in \mathbb{R}$ unbiased est. could be ok

(I small)



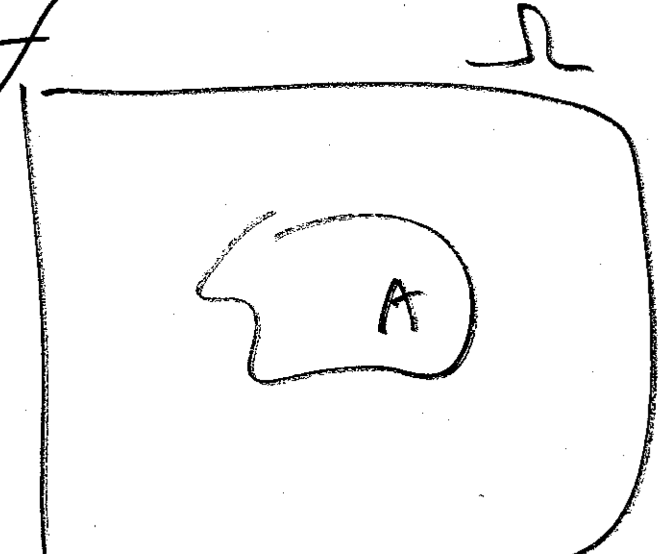
but boundary data separation ≈ trouble

$\tilde{\sigma}_H$

RS dist of $(\hat{\sigma}^2_{H})$

$Y_{ij} = \mu + a_i^H + a_{ij}^P$

iid $N(0, \sigma^2_H)$ $i = 1, \dots, I$



$P(A) = \dfrac{A}{1}$

$$P(A \mid B) = \begin{cases} \dfrac{P(A \,\&\, B)}{P(B)} & \text{if } P(B) > 0 \\ \text{undefined} & P(B) = 0 \end{cases}$$

$$y_{ij} = \mu + a_i^H + a_{ij}^P \qquad i = 1, \ldots, I \quad j = 1, \ldots, J$$

$$a_i^H \overset{iid}{\sim} N(0, \sigma_H^2) \qquad a_{ij}^P \overset{iid}{\sim} N(0, \sigma_P^2)$$

$$M$$

$$p(\mu \mid D\, B)$$

$$\wedge$$

$$p(\mu \mid D\, B\, M)$$

this structure

Y_i exchangeable $\longrightarrow$ $\left\{ \begin{array}{l} (\theta \mid \mathcal{B}) \sim p(\theta \mid \mathcal{B}) \\ (Y_i \mid \theta \mathcal{B}) \overset{iid}{\sim} \text{Bernoulli}(\theta) \end{array} \right\}$

$\uparrow$

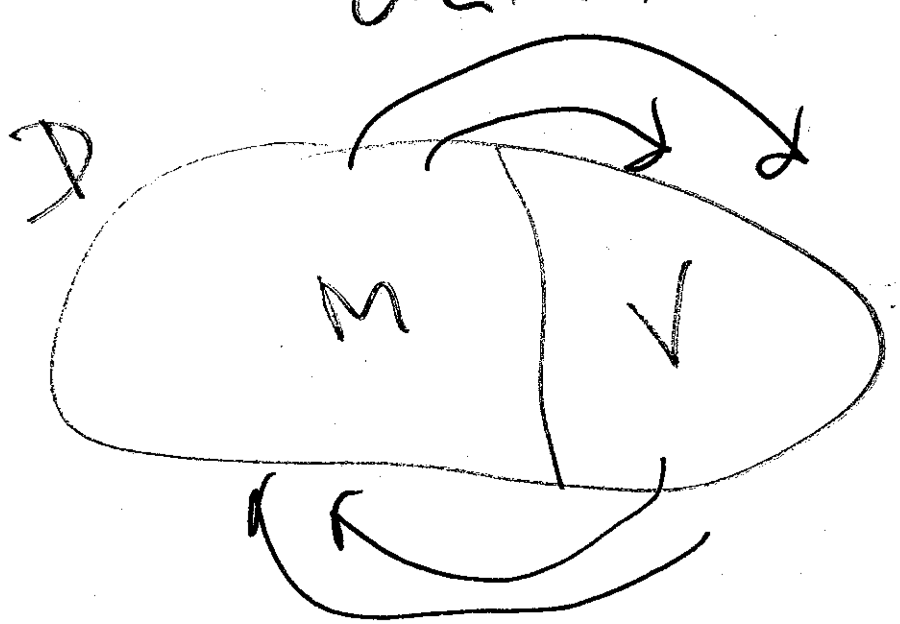
1/0

de Finetti's thm

problem

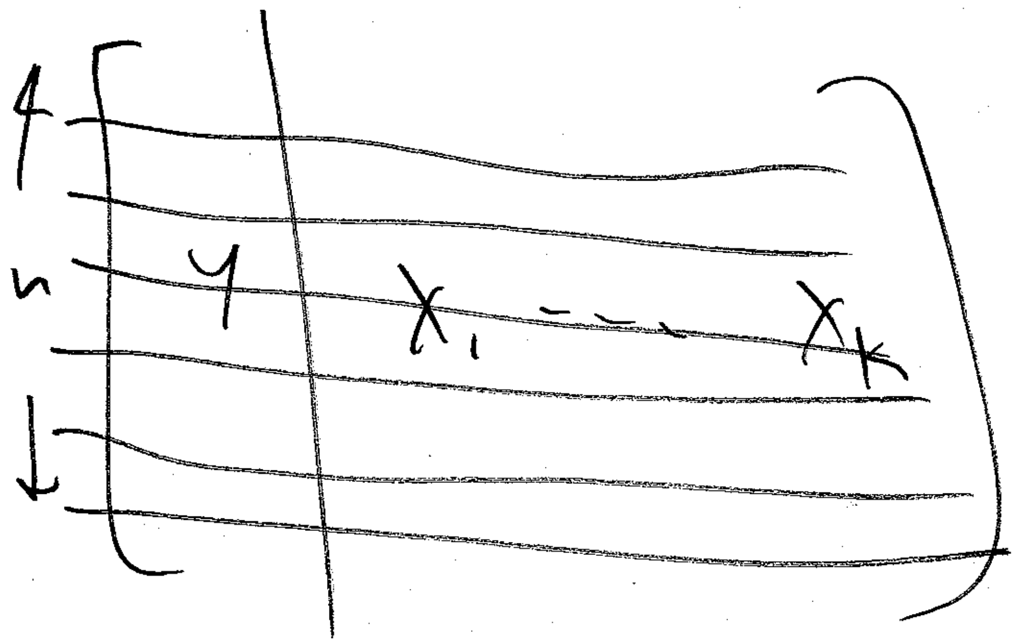
unique sampling dist.

context

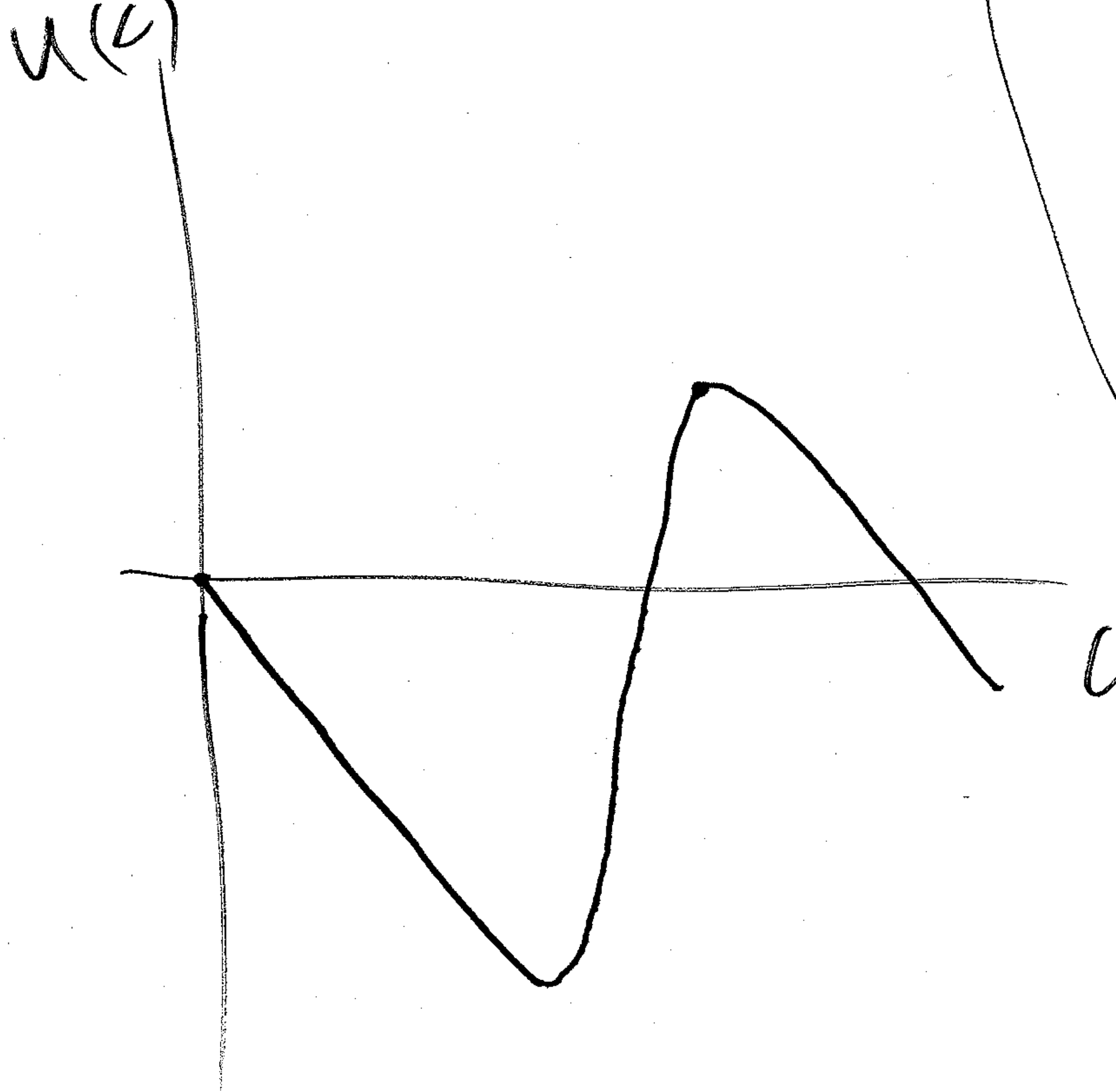


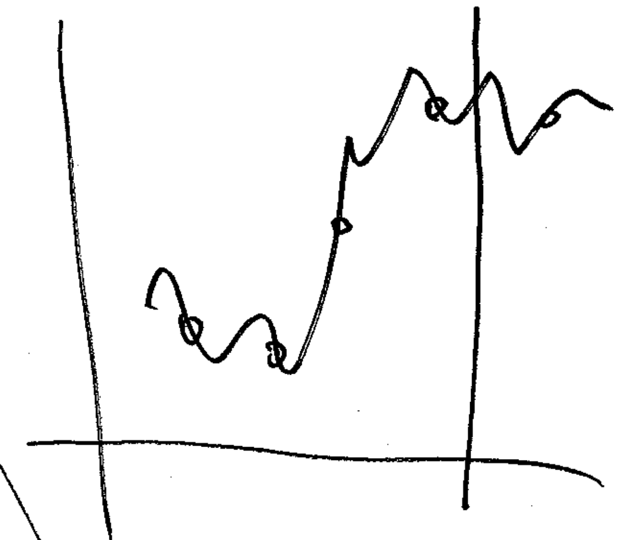
2 way cross-validation

CCV
calibration
cross-
validation



regression





$Y_1 = 1$ w.p. p

$\ 0 \quad 1-p$

n	model space dim
1	1
2	3

		Y_2	
		1	0
Y_1	1	p_1	p_2
	0	p_3	$1-p_1-p_2-p_3$

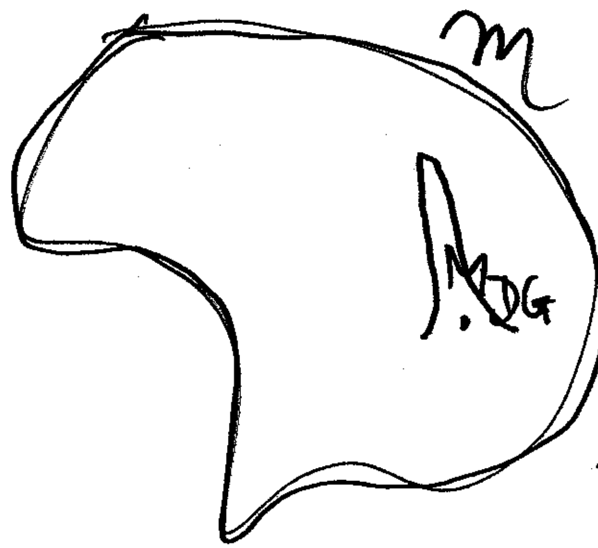
n	dim of model space
1	$1 = 2^n - 1$
2	$3 = 2^n - 1$
3	$7 = 2^n - 1$
⋮	

if $M_{DG} \in \mathcal{M}$

then $p(M \mid DB)$

→ point mass on "nearest" (closest) model in $\mathcal{M}$

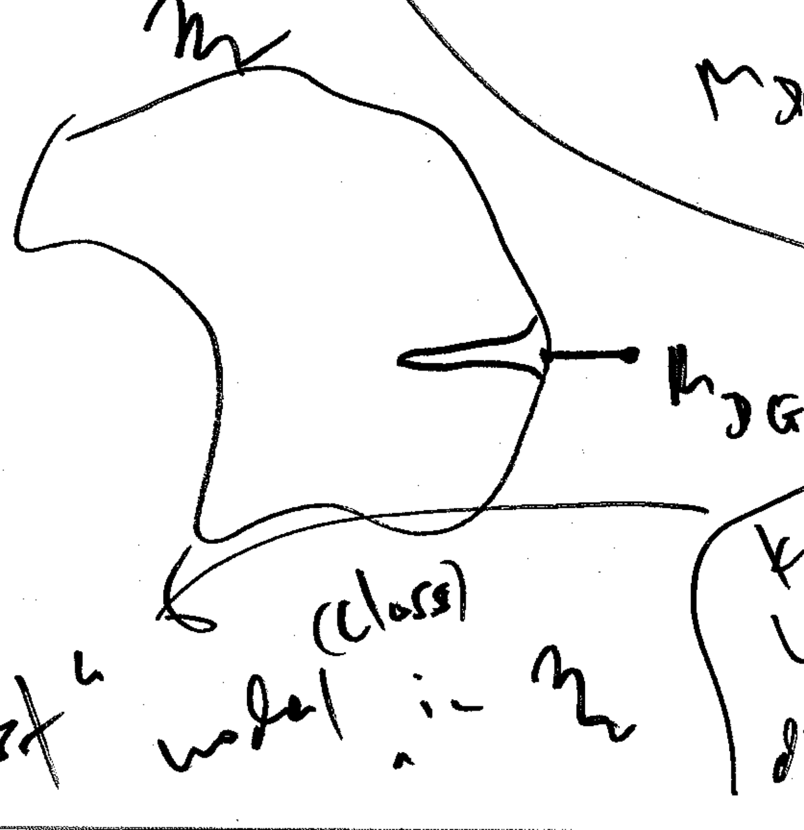
Kullback-Leibler divergence



as $n \to \infty$

$p(M \mid DB)$

→ point mass on M_{DG} if $M_{DG} \in \mathcal{M}$



$(M_1, \ldots, M_k) = \mathcal{M}$

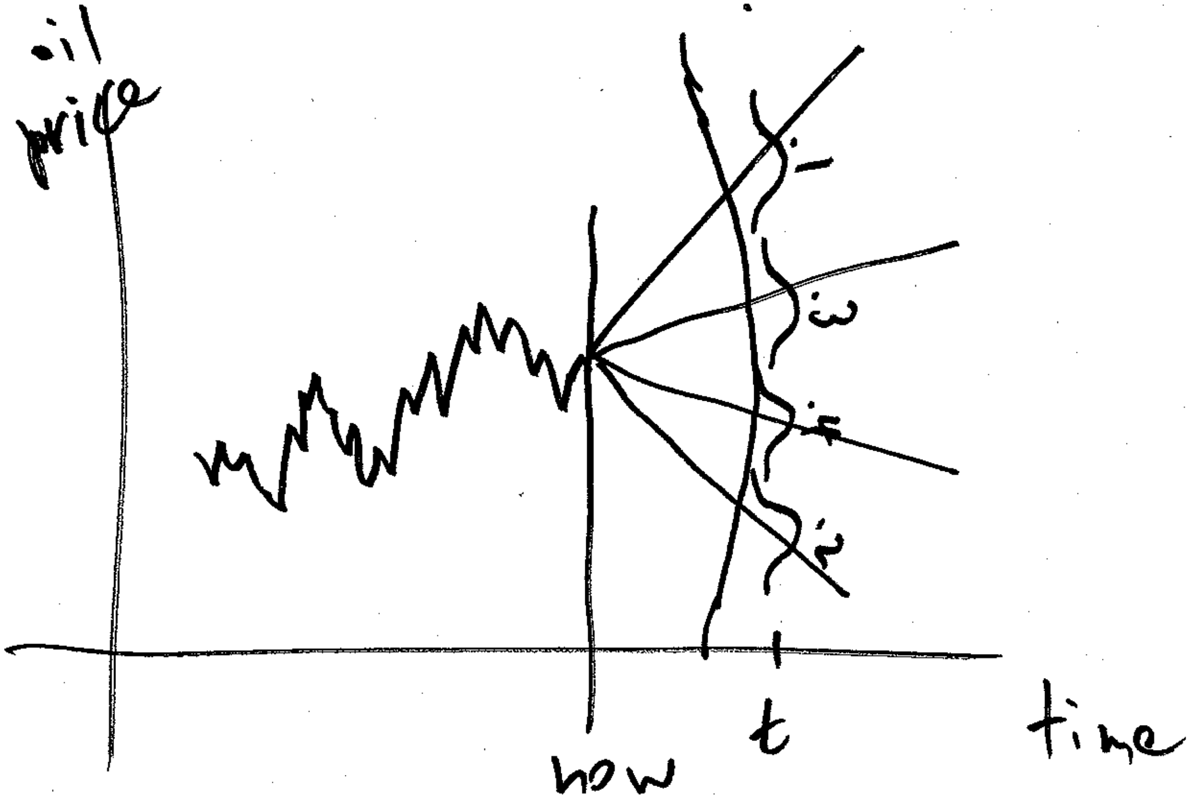
$p(y^* \mid D\, B\, \mathcal{M})$

future data

$p(y^* \mid D\, B\, M_3) \quad \cdot k$

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

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

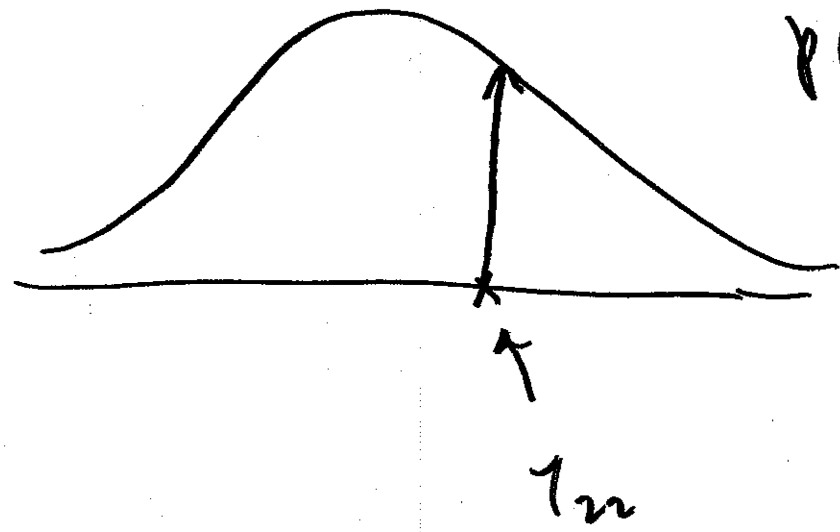


$\sigma_1^2 + x_i \sigma_2^2$

$$(Y_i \mid \lambda_i) \overset{ind}{\sim} \boxed{P}(\lambda_i)$$

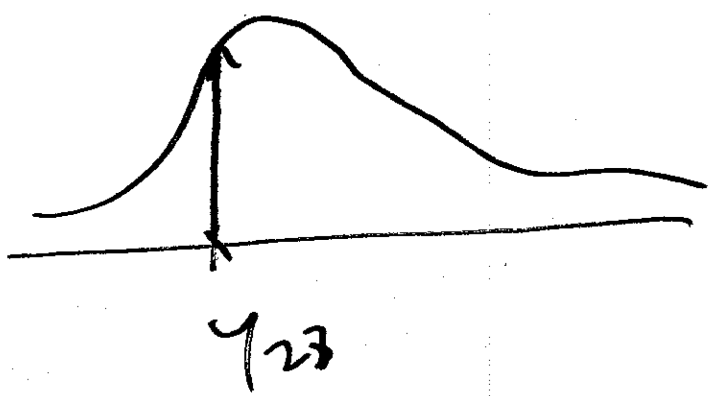
$$\boxed{N}(0, \sigma^2) \text{ iid}$$

$$\log(\lambda_i) = \gamma_0 + \gamma_1 x_i + e_i$$



$p(y_{22} \mid y_1, \ldots, y_{21}\ \mathcal{B}\ \mathcal{M})$

$\boxed{\text{scoring rule}}$



find M to make

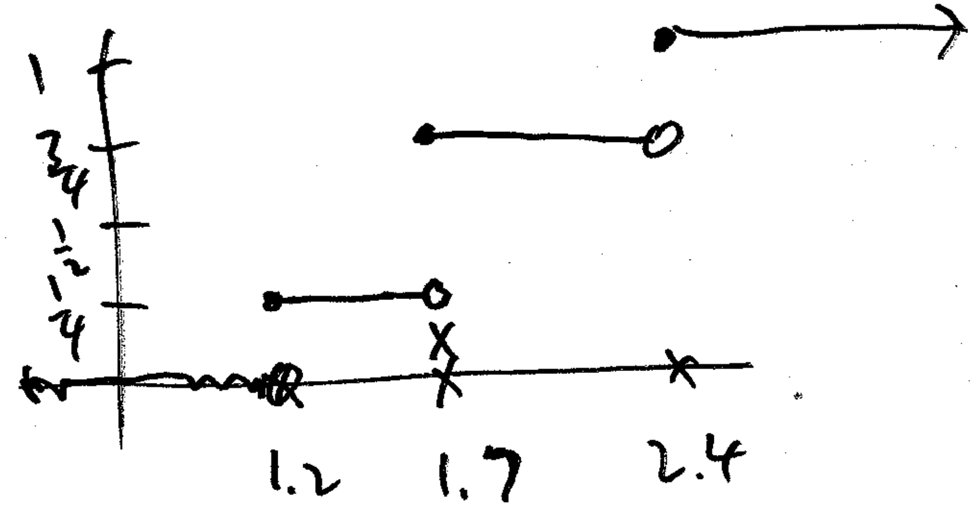
$$LS_{CV}(\mathcal{M} \mid y\ \mathcal{B}) = \frac{1}{n} \sum_{i=1}^{n} \log p(y_i \mid y_{-i}\ \mathcal{B}\ \mathcal{M})$$

jackknife

big

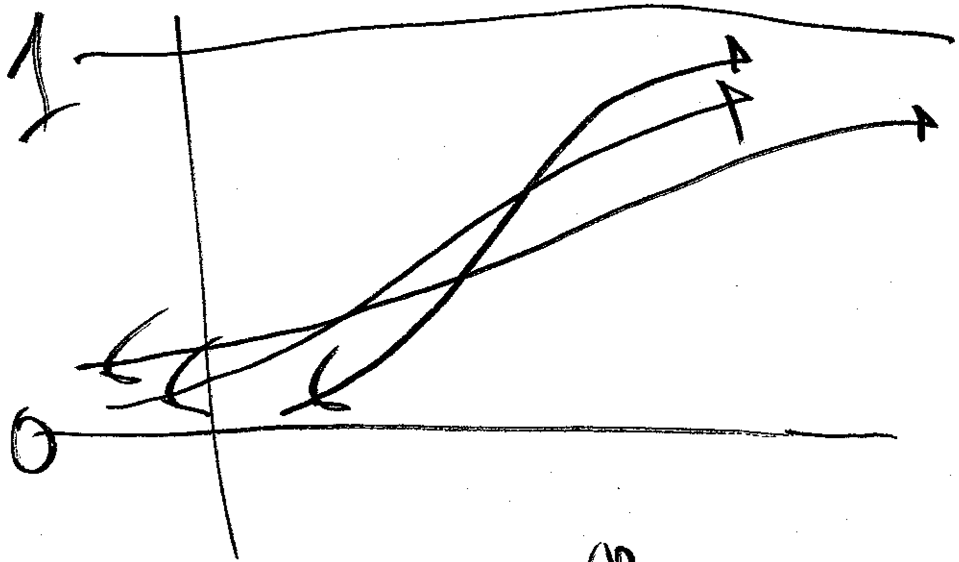
$$LS_{FS}(M \mid y\ \mathcal{B}) = \frac{1}{n} \sum_{i=1}^{n} \log p(y_i \mid y\ \mathcal{B}\ M)$$

empirical CDF of $\begin{Bmatrix} 1.2 \\ 1.7 \\ 1.7 \\ 2.4 \end{Bmatrix}$



$$\hat{F}_n(t) = \frac{\#(y_i \le t)}{n}$$

$$= \frac{1}{n} \sum_{i=1}^{n} I(y_i \le t)$$

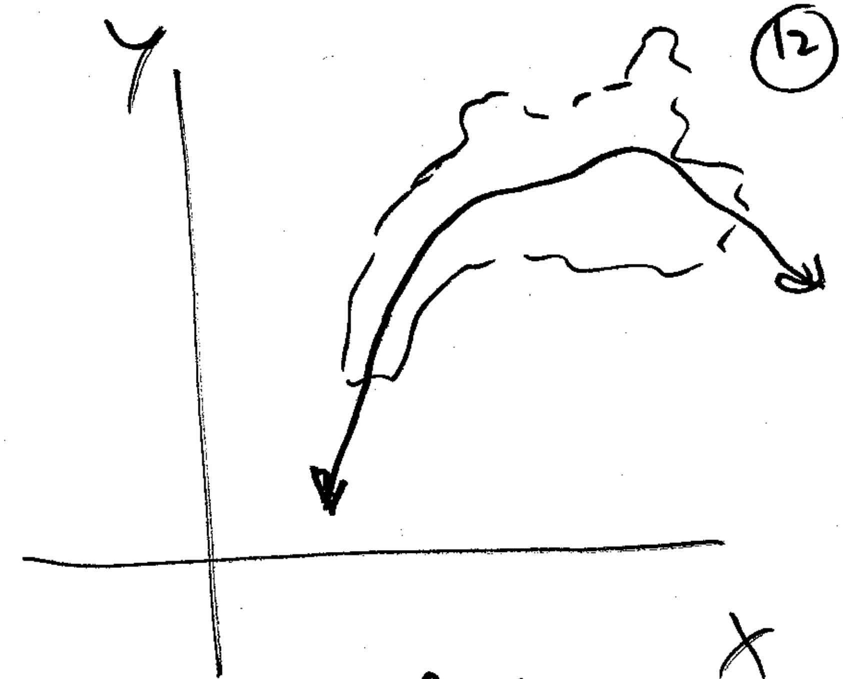


BNP

on CDFs

(Polya trees)

(DP priors)



BNP

on

regression

surfaces

(Gaussian

processes)