

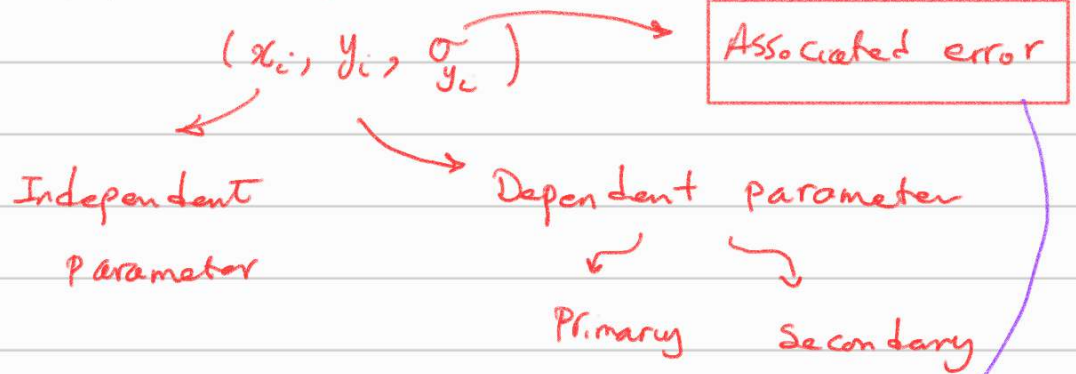
14.3, 2, 3.

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ

☆ Bayesian approach

Recall that : ① $\{D_i\}$, $i=1 \dots N$ * of observed data point

☆ As an Illustrative (1+1)-Dimension



☆ Statistical Error

☆ Systematic

$$\sigma_{y_i}^2 = (\sigma_{y_i}^{stat})^2 + (\sigma_{y_i}^{sys})^2$$

↑ ↑
Independent Value

☆ Another example (1+1)-Dimension experiment

x_i, y_i and Cov matrix (Including the error for y_i)
 ↑ ↑
 $i=1 \dots N$

$$Cov = \langle \delta y \otimes \delta y \rangle = \begin{bmatrix} \langle \delta y_1^2 \rangle & \langle \delta y_1 \delta y_2 \rangle & \dots & \langle \delta y_1 \delta y_N \rangle \\ & \langle \delta y_2^2 \rangle & & \\ & & & \\ & & & \langle \delta y_N^2 \rangle \end{bmatrix}_{N \times N}$$

→ $\sigma_{y_i}^2$

and according to a given theoretical model, we have

② $\{\theta_j\}, j=1, \dots, M$ ~~N~~ of model free Parameters

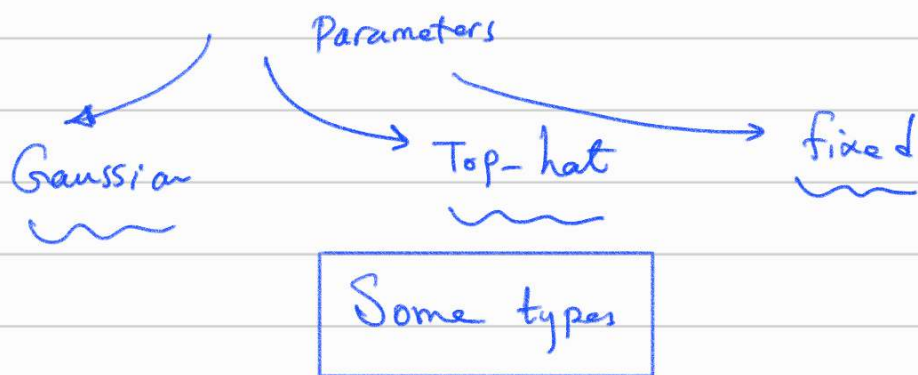
• Definition Degree of freedom (Dof)

$$\mathcal{N} \equiv N - M$$

$$\chi^2 \rightarrow \chi^2_{\nu} \equiv \frac{\chi^2}{\mathcal{N}}$$

we will see later

③ $P(\{\theta\})$: Prior Probability of model free



④ Posterior Model Probability is given by:

$P(\{\theta\} | D)$: Probability of model given Data

according to the Bayes theorem

$$P(\{\theta\} | D) = \frac{L(D | \{\theta\}) P(\{\theta\})}{\int d\{\theta\} L(D | \{\theta\}) P(\{\theta\})}$$

likelihood of the Data given the model

Denominator Serves the Normalization factor

⑤ By Maximization of Posterior, we achieve to

$$f(\theta)_{\text{Best}} \text{ in such that } \underline{P(f(\theta) = f(\theta)_{\text{Best}} | D) \equiv \text{Maximized}}$$

⑥ The Confidence interval for free parameters will be determined usually numerically.

⑦ The Goodness of fit is also computed via

☆ Akaike Information Criterion (AIC)

☆ Bayesian Information Criterion (BIC)

☆ In other word: we are trying to do model selection via Information theory

☆ In this regard, we are dealing with the trade-off between the Goodness of fit and the Simplicity of the Selected Model.

☆ By introducing a Penalty term to investigate overfitting and underfitting.

⑧ Finally our inference will be drawn with respect to the value of AIC and/or BIC

☆ A main point: The main challenge related to the

Common Bayesian approach is that the inferences are typically drawn as if the model was

Chosen a priori - that is

and it does not account for uncertainty in the choice of model (model selection)

⑨ To mitigate (reduce) above challenge a systematic approach can be introduced by.

Bayesian Model Averaging (BMA)

□ offer a principled approach to deal with the

choice of various model. $\longrightarrow P(M_i)$

\longleftarrow Prior Probability of model selection

which is specified by Analyst

□ Weighted average of model's free parameters

□ Robustness: Less sensitivity to the specific choice

□ Improved Predictive Performance:

by averaging we always-achieving accurate

Prediction for best value compared to

relying one model.

Some Refs: $\left\{ \begin{array}{l} \text{arXiv: 2403.02120} \\ \text{arXiv: 2310.06747} \\ \text{arXiv: 1509.08864} \end{array} \right\}$

BMA

⑩ To start suppose that we have

$M_i, i=1, \dots, \underline{K}$ # of models

Each model has $\{ \theta_i \}$ Parameters
 $i=1, \dots, M_i$

The Posterior model Probability reads as

almost new to compute: we are looking for an efficient estimator

$$\star P(M_i | D) = \frac{\mathcal{L}(D | M_i) \phi(M_i)}{\sum_{j=1}^K \mathcal{L}(D | M_j) \phi(M_j)}$$

$$\theta = \bigcup_{j=1}^K \theta_j$$

$$\begin{aligned} \star P(\theta | D) &= \sum_{j=1}^K P(\theta, M_j | D) \\ &= \sum_{j=1}^K P(\theta | D, M_j) P(M_j | D) \end{aligned}$$

Common Posterior Probability

$$P(\theta | D, M_j) = \frac{\mathcal{L}(D | \theta, M_j) P(\theta, M_j)}{\int d\theta \mathcal{L}(D | \theta, M_j) P(\theta, M_j)}$$

$$\int d\theta \mathcal{L}(D | \theta, M_j) P(\theta, M_j)$$

We usually consider that $P(\theta, M_j) = P(\theta) P(M_j)$

$$P(\theta | D, M_j) = \frac{\mathcal{L}(D | \theta, M_j) P(\theta)}{\int d\theta \mathcal{L}(D | \theta, M_j) P(\theta)}$$

Common Posterior Probability.

⑪ Let's us focus on $P(M_i | D)$ $\left\{ \begin{array}{l} P(M_i) \leftarrow \text{specified by operator} \\ \mathcal{L}(D | M_i) = ? \end{array} \right.$

⑫ To compute $\mathcal{L}(D | M_i)$ [marginal likelihood], we have

$$\mathcal{L}(D | M_i) = \int d\theta_i \mathcal{L}(D | M_i, \theta_i)$$

a proper Estimator for $L(D|M_i)$ which is called

$\hat{L}(D|M_i)$ is given by:

$$\hat{L}(D|M_i) = \frac{N_i}{\sum_{t=1}^{N_i} L(D|M_i, \theta_i^{(t)})}$$

To compute above estimator (un-biased estimator) we need to

★ $\{\theta_i^{(t)}\}$ for MCMC (Canonical Hamiltonian Mc,

Micro Canonical Hamiltonian Mc, Canonical Langevin

Hamiltonian Mc . . .) chain

★ The full likelihood : $L(D|M_i, \theta_i^{(t)})$

(13) In Summary for (BMA)

$$\star P(M_i|D) = \frac{L(D|M_i) P(M_i)}{\sum_{j=1}^K L(D|M_j) P(M_j)}$$

Common Posterior

$$\star P(\theta|D) = \sum_{i=1}^K P(\theta, M_i|D) = \sum_{i=1}^K P(\theta|D, M_i) P(M_i|D)$$

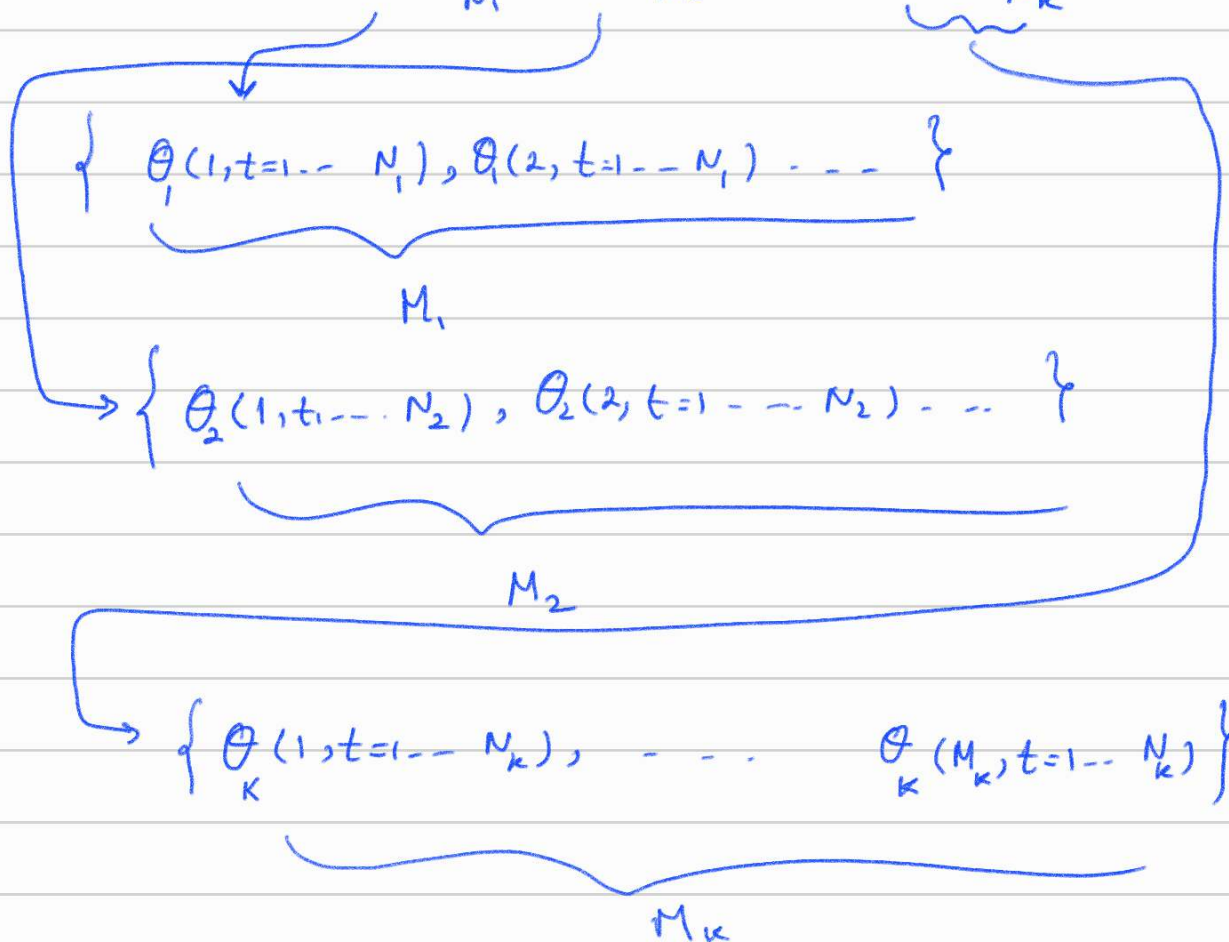
$$\star \hat{L}(\text{DIM}_i) \rightarrow \hat{L}(\text{DIM}_i) = \frac{N_i}{\sum_{t=1}^{N_i} L(\text{DIM}_i, \theta_i^{(t)})}$$

(14) Numerical Algorithm includes :

★ Based on the Common algorithm for Bayesian model

one should: run MCMC (HMC, GHMC, CALHMC)

yielding $\{\theta_1\}_{N_1}, \{\theta_2\}_{N_2} \dots \{\theta_k\}_{N_k}$



★ Consequently we achieve to likelihood:

$$L(\text{DIM}_i, \theta_i^{(t)}) : \left\{ L(\text{DIM}_i, \{\theta_1(1, t=1 \dots N_1), \dots, \theta_1(M_1, t=1 \dots N_1)\}) \right\}$$

$$K \left\{ \begin{array}{l} \mathcal{L}(\text{DIM}_2, \theta_2^{(t)}) = \mathcal{L}(\text{DIM}_2, \{\theta_2(1, t=1 \dots N_2), \dots, \theta_2(M_2, t=1 \dots N_2)\}) \\ \vdots \\ \vdots \end{array} \right.$$

$$\star \hat{\mathcal{L}}(\text{DIM}_i) = \frac{N_i}{\sum_{t=1}^{N_i} \left(\frac{1}{\mathcal{L}(\text{DIM}_i, \theta_i^{(t)})} \right)}$$

$$\hat{\mathcal{L}}(\text{DIM}_i) P(M_i)$$

$$\star P(M_i | D) = \frac{\hat{\mathcal{L}}(\text{DIM}_i) P(M_i)}{\sum_{j=1}^K \hat{\mathcal{L}}(\text{DIM}_j) P(M_j)}$$

$$\star P(\theta | D) = \sum_{i=1}^K P(\theta | D, M_i) P(M_i | D)$$

Common Posterior

