۱۴۰۳ ،۲ ،۳۰

بسم الله الرحمن الرحیم

☆ Bayesian approach

Recall that : ① $\{D_i\}$, $i = 1 \ldots N$ ⁂ of observed Data Point

☆ As an Illustratu (1+1)-Dimensia

(x_i, y_i, σ_{y_i})

- x_i → Independent parameter
- y_i → Dependent parameter → Primary, Secondary
- σ_{y_i} → Associated error

☆ 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's)

$i = 1 \ldots N$

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

Diagonal → $\sigma_{y_i}^2$

and according to a given theoretical model, we have

② $\{\theta_j\}$, $j=1..M$ # of model free Parameters

· Definition Degree of freedom (DoF) $\boxed{\mathcal{N} \equiv N-M}$

$$\boxed{\chi^2 \longrightarrow \chi^2_\nu \equiv \frac{\chi^2}{\mathcal{N}}}$$ we will see later

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

Some types: Gaussian, Top-hat, fixed

④ Posterior Model probability is given by:

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

according to the Bayes theorem

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

$\mathcal{L}(D|\{\theta\})$ → likelihood of the Data given the Model

→ Denominator serves the Normalization factor

⑤ By Maximization of Posterior, we achieve to

$\{\theta\}_{Best}$ in such that $P(\{\theta\} = \{\theta\}_{Best} | D) \equiv$ 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 inference 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 → $P(M_i)$

  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:
- arXiv: 2403.02120
- arXiv: 2310.06747
- arXiv: 1509.08864

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

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

$$\Theta = \bigcup_{j=1}^{K} \Theta_j$$

☆ $$P(\Theta|D) = \sum_{j=1}^{K} P(\Theta, M_j | D)$$

$$= \sum_{j=1}^{K} P(\Theta | D, M_j)\, P(M_j | D)$$

Common Posterior Probability

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

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 focuse on $P(M_i|D)$ → $P(M_i)$ ← specified by operator; → $\mathcal{L}(D|M_i) = ?$

(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 $\mathcal{L}(D|M_i)$ which is called

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

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

To Compute above estimator (Un-biased estimator) we need to

☆ $\{\theta_i^{(t)}\}$ for McMc (Canonical Hamiltonian Mc, microcanonical Hamiltonian Mc, Canonical Langevin Hamiltonian Mc ..) chain

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

(13) In Summary for (BMA)

☆ $$P(M_i|D) = \frac{\mathcal{L}(D|M_i)\, P(M_i)}{\sum_{j=1}^{K} \mathcal{L}(D|M_j)\, P(M_j)}$$ Common Posterior

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

☆ $\hat{\mathcal{L}}(D|M_i) \longrightarrow \hat{\mathcal{L}}(D|M_i) = \dfrac{N_i}{\sum_{t=1}^{N_i} \dfrac{1}{\mathcal{L}(D|M_i, \theta_i^{(t)})}}$

(14) Numerical Algorithm includes :

☆ Based on the common algorithm for Bayesian model one should run MCMC (HMC, CAHMC, CALHMC) yielding $\{\theta_1\}_{N_1}, \{\theta_2\}_{N_2} \ldots \{\theta_k\}_{N_k}$

$$\underbrace{\{\theta_1(1, t=1 \ldots N_1), \theta_1(2, t=1 \ldots N_1) \ldots\}}_{M_1}$$

$$\underbrace{\{\theta_2(1, t=1 \ldots N_2), \theta_2(2, t=1 \ldots N_2) \ldots\}}_{M_2}$$

$$\underbrace{\{\theta_K(1, t=1 \ldots N_k), \ldots \theta_K(M_k, t=1 \ldots N_k)\}}_{M_k}$$

☆ Consequently we achieve to likelihood:

$$\mathcal{L}(D|M_1, \theta_1^{(t)}) : \{\mathcal{L}(D|M_1, \{\theta_1(1, t=1 \ldots N_1), \ldots \theta_1(M_1, t=1 \ldots N_1)\}$$

$$K \begin{cases} \mathcal{L}(D|M_2, \theta_2^{(t)}) = \{\mathcal{L}(D|M_2, \{\theta_2(1, t=1 \ldots N_2), \ldots \theta_2(M_2, t=1 \ldots N_2)\} \\ \vdots \end{cases}$$

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

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

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

Common Posterior