Lecture 9

02/15/2022

- Bayesian parameter estimation examples

① Samples from a Gaussian distribution, but sampled data have errors.

↳ errors blur the distribution ... it won't necessarily be Gaussian anymore

⇒ Noise = measurement uncertainties + underlying distribution spread

$$\therefore \ln \mathcal{L} \propto \frac{-1}{2} \sum_{i=1}^{N} \left\{ \ln(\sigma^2 + \sigma_i^2) + \frac{(x_i - \mu)^2}{(\sigma^2 + e_i^2)} \right\}$$

where $[x_i, e_i]$ = [data, error]
$[\mu, \sigma]$ = [mean, std] of underlying distribution

$\boxed{\text{NOTE}}$ posterior of μ, σ are NON-GAUSSIAN!

② Gaussian distribution embedded in a uniform noise background.

Likelihood = Mixture of Gaussian + Uniform

$$p(x_i | A, \mu, \sigma) = \frac{A}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x_i-\mu)^2}{2\sigma^2}} + \frac{1-A}{W}$$

⇒ A, and $(1-A)$ are mixture probs.

- ## Bayesian Model Comparison

$\Rightarrow$ Rank models based on how well $\underbrace{\text{(and how efficiently)}}_{\text{Occam's Razor}}$ they explain data

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}}$$

$$\text{Evidence} \equiv \underbrace{p(d)}_{\substack{\text{sometimes} \\ \text{“}Z\text{”}}} = \int \text{Likelihood} \times \text{Prior}\, d\theta$$

$$= \int p(d|\theta)\, p(\theta)\, d\theta$$

$$\boxed{\text{Bayesian Odds}} = \frac{p(M_1|d)}{p(M_2|d)}$$

think of betting odds

$$O_{12} = \frac{p(d|M_1) \times p(M_1)}{p(d|M_2) \times p(M_2)}$$

$$= \underbrace{\frac{Z_1}{Z_2}}_{\text{Bayes factor, } B_{12}} \times \frac{p(M_1)}{p(M_2)}$$

* How do we interpret these?

B_{12}	Strength
$< 1:1$	negative
$10:1$	substantial
$30:1$	strong
$> 100:1$	decisive

subjective + problem dependent

Approximate Bayesian Model Comparison

⇒ Computing the ODDS RATIO and BAYES FACTOR are hard... any shortcuts?

(i) Akaike Information Criterion (AIC)

$$AIC = -2\ln\mathcal{L}_{max} + 2k + \frac{2k(k+1)}{N-k-1}$$

parameters data points.

(ii) Bayesian Information Criterion (BIC)

$$BIC = -2\ln\mathcal{L}_{max} + k\ln N$$

↳ can be derived from Bayesian odds ratio assuming $\mathcal{L}$ is Gaussian + data is very ~~strong~~ + informative.

⇒ These approximations are essentially χ^2 modified by a penalty on model complexity.