

Lecture 9

- 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~~ Noise = measurement uncertainties

underlying + distribution spread

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

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

NOTE posterior of μ, σ are **NON-GAUSSIAN!**

- ② Gaussian distribution embedded in a uniform noise background.

Likelihood = Mixture of Gaussian + Uniform

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

$\Rightarrow A$, and $(1-A)$ are mixture probs.

- Bayesian Model Comparison

→ Rank models based on how well
(and how efficiently) they explain data

Occam's Razor

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

$$\text{Evidence} \equiv p(d) = \int \text{Likelihood} \times \text{Prior } d\theta$$

sometimes
 "Z" = $\int p(d|\theta) p(\theta) d\theta$

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

think of
betting odds

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

$$= \left(\frac{Z_1}{Z_2} \right) \times \frac{p(M_1)}{p(M_2)}$$

* How do we interpret these?

Bayes factor, B_{12}

B_{12}	Strength	subjective + problem dependent
< 1:1	negative	
10 : 1	substantial	
30 : 1	strong	
> 100 : 1	decisive	

- Approximate Bayesian Model Comparison
 - ⇒ Computing the Odds Ratio and Bayes Factor are hard ... any shortcuts?

(i) Akaike Information Criterion (AIC)

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

↑ parameters ↑ data points

(ii) Bayesian Information Criterion (BIC)

$$BIC = -2 \ln L_{\max} + k \ln N$$

⇒ can be derived from Bayesian odds ratio assuming L is Gaussian + data is very strong + informative.

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