Bayesian Model Comparison

$$P(\theta \mid d, M) = \frac{\mathcal{L}(d \mid \theta, M)\, \pi(\theta \mid M)}{Z_M}$$

where $Z_M = P(d \mid M)$.

$$Z_M = \int d\theta\, \mathcal{L}(d \mid \theta, M)\, \pi(\theta \mid M)$$

Suppose we have two ~~competing~~ competing models A, B.

params λ_A params λ_B.

Both models describe same d.

$$\mathcal{L}(d \mid \lambda_A, A)$$
$$\mathcal{L}(d \mid \lambda_B, B)$$

Both models making predictions for <u>same</u> data d.

What's wrong with using <u>Max Likelihood Ratio</u> as a model selection statistic?

$$MLR = \frac{\max_{\lambda_A} \mathcal{L}(d \mid \lambda_A, A)}{\max_{\lambda_B} \mathcal{L}(d \mid \lambda_B, B)}.$$

Model A: $y(t) = 0$ $\leftarrow$ 0 free params

Model B: $y(t) = \lambda$ $\leftarrow$ 1 free param.

$$\mathcal{L}(D \mid A) = \prod_{i=1}^{6} \frac{\exp\left(-\frac{1}{2} y_i^2\right)}{\sqrt{2\pi}}$$

$$\mathcal{L}(D \mid \overset{\lambda}{\underset{\wedge}{}} B) = \prod_{i=1}^{6} \frac{\exp\left(-\frac{1}{2} (y_i - \lambda)^2\right)}{\sqrt{2\pi}}$$

$$\max_{\lambda} \mathcal{L}(D \mid \lambda, B) = \boxed{?} > \mathcal{L}(D \mid A)$$

@ $\lambda =$ sample mean $(y_i) = \frac{1}{6} \sum_{i=1}^{6} y_i$

Bayesian solution

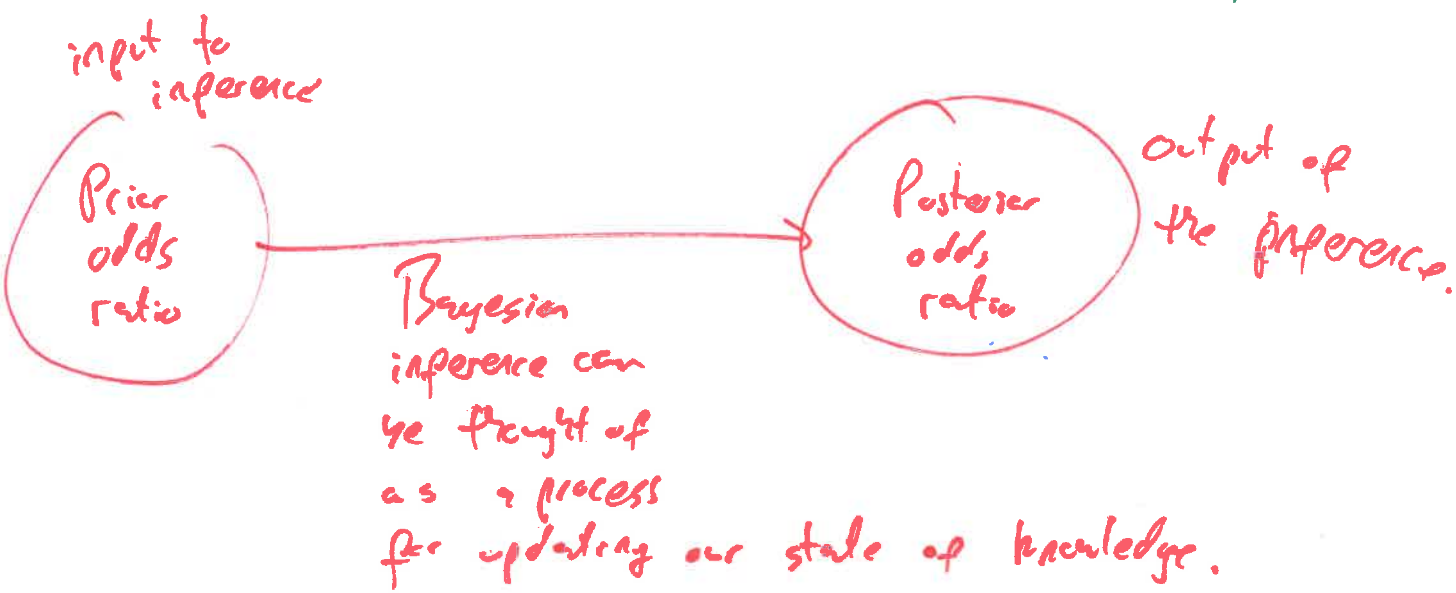
$$\text{Posterior Odds Ratio} =$$

$$O_{A,B} = \frac{P(\boxed{A} \mid D)}{P(B \mid D)}$$

A is the correct model.

$$= \frac{\cancel{\frac{1}{P(D)}}\, P(D \mid A)\, P(A)}{\cancel{\frac{1}{P(D)}}\, \underbrace{P(D \mid B)}_{\text{depends on data}}\, \underbrace{P(B)}_{}}$$

prior odds ratio.



$$O_{A,B} = \frac{P(D|A)}{P(D|B)} \cdot \frac{P(A)}{P(B)}$$

$$= \frac{\int d\lambda_A \, P(D, \lambda_A | A)}{\int d\lambda_B \, P(D, \lambda_B | B)} \; \frac{P(A)}{P(B)}$$

← law of total prob.

likelihood $\mathcal{L}(D|\lambda_A, A)$

$$= \frac{\int d\lambda_A \, \boxed{P(D|A, \lambda_A)} \, \boxed{P(\lambda_A | A)}}{\int d\lambda_B \, P(D|B, \lambda_B) \, P(\lambda_B|B)} \; \frac{P(A)}{P(B)}$$

← product rule.

$\pi(\lambda_A | A)$ prior PDF on the params of model A

$$= \frac{\int d\lambda_A \, \mathcal{L}(D|\lambda_A, A) \, \pi(\lambda_A|A)}{\int d\lambda_B \, \mathcal{L}(D|\lambda_B, B) \, \pi(\lambda_B|B)} \; \frac{P(A)}{P(B)}$$

$$= \boxed{\frac{Z_A}{Z_B}} \; \frac{P(A)}{P(B)}$$

multiply by the evidence ratio.

Choose equal prior odds ratio $\frac{P(A)}{P(B)} = 1$

$$O_{A,B} = \frac{Z_A}{Z_B} \qquad O_{A,B} = \frac{1}{O_{B,A}}$$

we need to calculate the two evidences

$$Z_A = \cancel{\int d\lambda_A}\, \mathcal{L}(D \mid \cancel{\lambda_A}, A)\ \cancel{\pi(\lambda_A \mid A)}$$

$$= \mathcal{L}(D \mid A)$$

$$\boxed{= \prod_{i=1}^{6} \frac{\exp\left(-\frac{1}{2} x_i^2\right)}{\sqrt{2\pi}}} \leftarrow$$

No free param, just evaluate the likelihood for A.

$$Z_A = 1.4 \times 10^{-8}$$

$$Z_B = \int d\lambda_B\, \mathcal{L}(D \mid \lambda_B, B)\ \pi(\lambda_B \mid B)$$

$$= \int d\lambda\, \mathcal{L}(D \mid \lambda, B)\ \pi(\lambda \mid B)$$

$$\mathcal{L}(D|\lambda, B) = \prod_{i=1}^{6} \frac{\exp\left(-\frac{1}{2}(y_i - \lambda)^2\right)}{\sqrt{2\pi}}$$

we need our final prior $\pi(\lambda | B) = \frac{\mathbb{1}_{(-\Lambda,\Lambda)}(\lambda)}{2\Lambda}$

"uniform prior"

$$Z_B^{(\Lambda)} = \frac{1}{2\Lambda} \int_{-\Lambda}^{\Lambda} d\lambda \prod_{i=1}^{6} \frac{\exp\left(-\frac{1}{2}(y_i - \lambda)^2\right)}{\sqrt{2\pi}}$$

$$Z_B = \frac{\lim_{\Lambda \to \infty} Z_B(\Lambda)}{2\Lambda} = \boxed{\frac{1.45 \times 10^{-3}}{\Lambda}}$$

$$\boxed{O_{\Lambda,B} = 0.1\Lambda.}$$

models that are more "flexible" have lower Z.

"Occam penalty".

makes less precise predictions for data.

A model is ~~more~~ less flexible if it has
- few parameters.
- narrow prior ranges.

Bayesian inference naturally penalises flexible/complicated models by applying Occam penalty.