# Lecture 4

Friday, July 8, 2022 1:46 PM

Parameter inference :

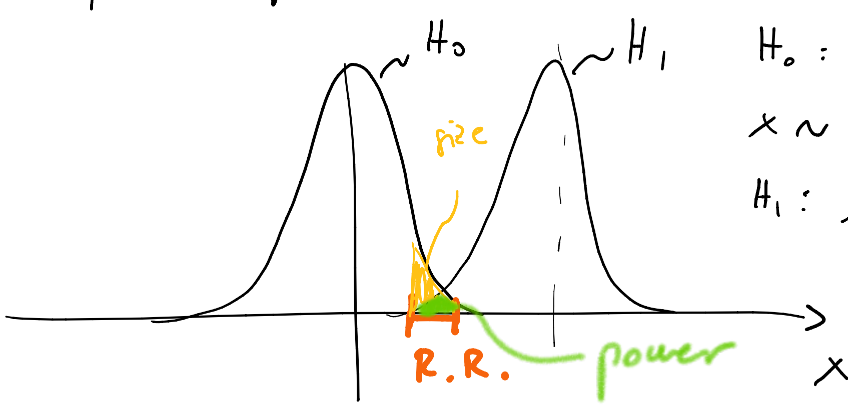
$$P(\Theta \mid d, \underbrace{M}_{\text{model}}) = \frac{P(d \mid \Theta, M)\, P(\Theta \mid M)}{\boxed{P(d \mid M)}}$$

evidence, model likelihood

Aside: <u>Hypothesis testing</u>

Null hypothesis $H_0$ ; alternative $H_1$

- Rejection region : a region where we reject $H_0$ and accept $H_1$
- The size : the prob. of an outcome in the rejection region if $H_0$ is true ( low size)
- The power of the rejection region : prob. of getting an outcome in the rejection region if $H_1$ is true ( high power)




$H_0 : \mu = 0$

$x \sim N(\mu, \sigma^2)$

$H_1 : \mu = 1$

Type I error : if $H_0$ is true , reject $H_0$ .

$\underline{\text{II}}$ - : if $H_1$ is true, accept $H_0$ .

Level of significance of the test $\alpha$ : prob. of committing a type I error

$$\alpha(R) = P_{H_0}(\text{accept } H_1) = P_{H_0}(x \in R)$$

Power of the test $= 1 -$ Prob( type II error)

**Neyman - Pearson lemma :**

$$L(x) = \frac{P_{H_1}(x)}{P_{H_0}(x)}$$

Set the significance $\alpha$ $(\alpha = 0.05)$

Accept $H_0$ if $L(x) < \alpha$

Reject $H_0$ if $L(x) > \alpha$

This is the most powerful test.

**For $\alpha$ :**

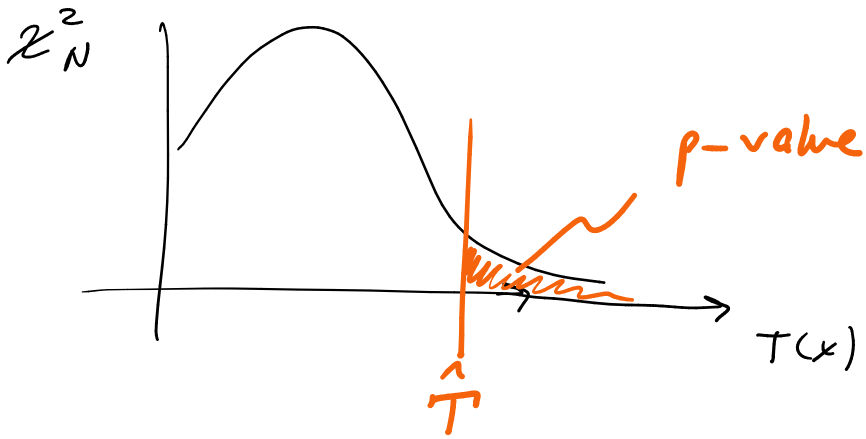
$H_0 : \mu = 0$

$x_i \sim N(\mu, \sigma^2)$, $N$ draws

$T(x) = \sqrt{N} \frac{\bar{x}}{\sigma}$, distributed as $\chi^2_N$

Size of the test :

$$P_{\mu=0}(T(x) \geq c) = P_{\mu=0}\left(\sqrt{N}\frac{\bar{x}}{\sigma} \geq c\right)$$




**Bayesian approach :** (BMC)

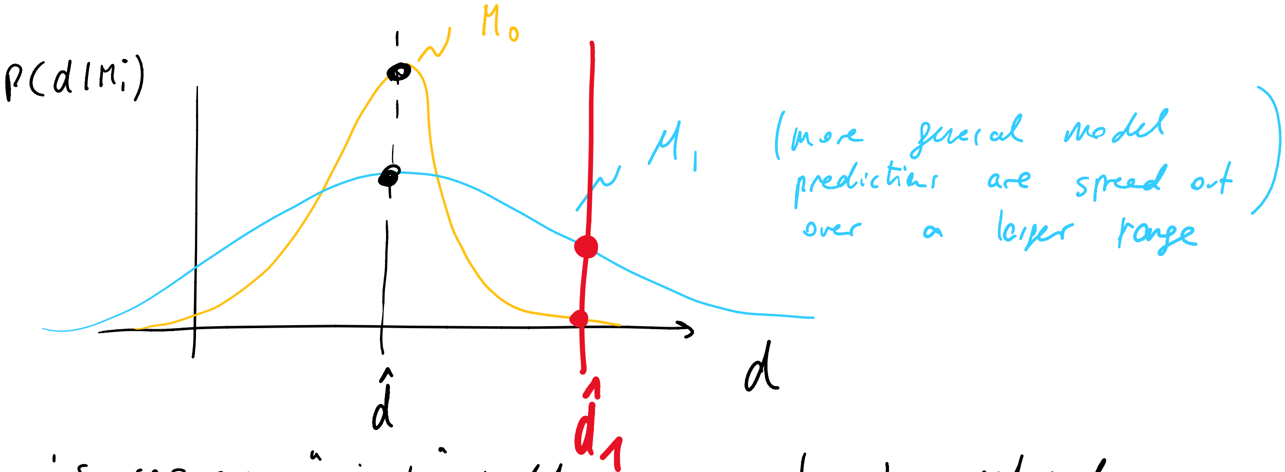
## Bayesian approach: (BMC)

$$P(M|d) = \frac{\boxed{p(d|M)}\, P(M)}{P(d)}$$

$$p(d|M) = \int_{\Omega} p(d|\theta, M)\, p(\theta|M)$$

$$\underbrace{\frac{P(M_0|d)}{P(M_1|d)}}_{\text{posterior odds}} = \underbrace{\boxed{\frac{p(d|M_0)}{p(d|M_1)}}}_{\substack{\text{BAYES FACTOR} \\ B_{01}}} \cdot \underbrace{\boxed{\frac{P(M_0)}{P(M_1)}}}_{\text{prior odds}}$$

Evidence is a predictive prob. for $d$:




Occam's razor: "simple" models are to be preferred.

## 1D example:

$M_0$ no free parameters: $\theta = 0 \rightarrow \delta(\theta)$

$M_1$ with 1 parameter, $p(\theta)$, $x \sim N(\theta, a^2)$

observed $\hat{x}$.

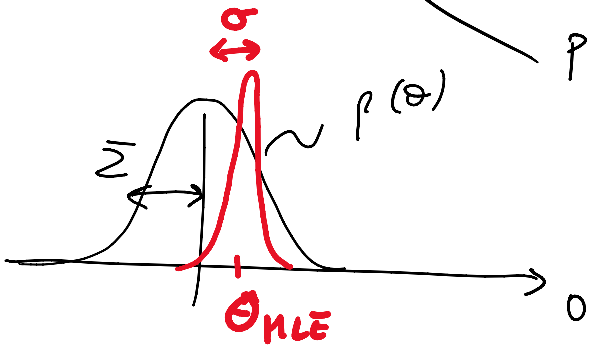
$$B_{01} = \frac{e^{-\frac{1}{2}(0-\hat{x})^2/a^2}}{\int d\theta\, e^{-\frac{1}{2}\frac{(\theta-\hat{x})^2}{a^2}}\, \boxed{p(\theta)}} \sim ? \qquad \left(\frac{\text{size } \mathcal{L}}{\text{size of prior}}\right)$$

$$B_{01} = \frac{e^{-\dots}}{\int d\theta \; e^{-\frac{1}{2}\frac{(\theta-\hat{x})^2}{\sigma^2}} \boxed{p(\theta)}} \sim (?)$$

(size of prior)

$=$

Model 1 :




$$P(d|M_1) \approx \underbrace{\mathcal{L}(\theta=\theta_{MLE})}_{\text{ML value}} \Big/ \left(\sigma \cdot \frac{1}{\Sigma}\right)$$

ML value (Quality of fit)

Occam's factor

penalizes "wasted" parameter space

$$\boxed{\text{Prior dependency in BMC } \underline{\text{never}} \text{ goes away!}}$$

⚠

$$\ln(B_{01}) = \ln\left[\sqrt{1+\left(\frac{\sigma}{\Sigma}\right)^{-2}} \exp\left(-\frac{1}{2}\frac{\lambda^2}{1+\left(\sigma/\Sigma\right)^2}\right)\right]$$

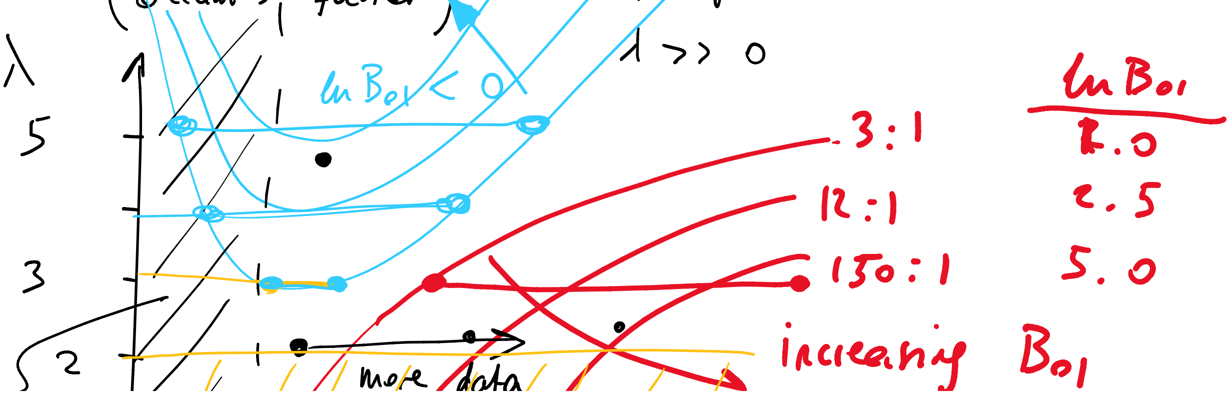
$$\lambda = \frac{|\theta_{MLE}|}{\sigma} \quad (\#\text{ of } \sigma \text{ "detection"})$$

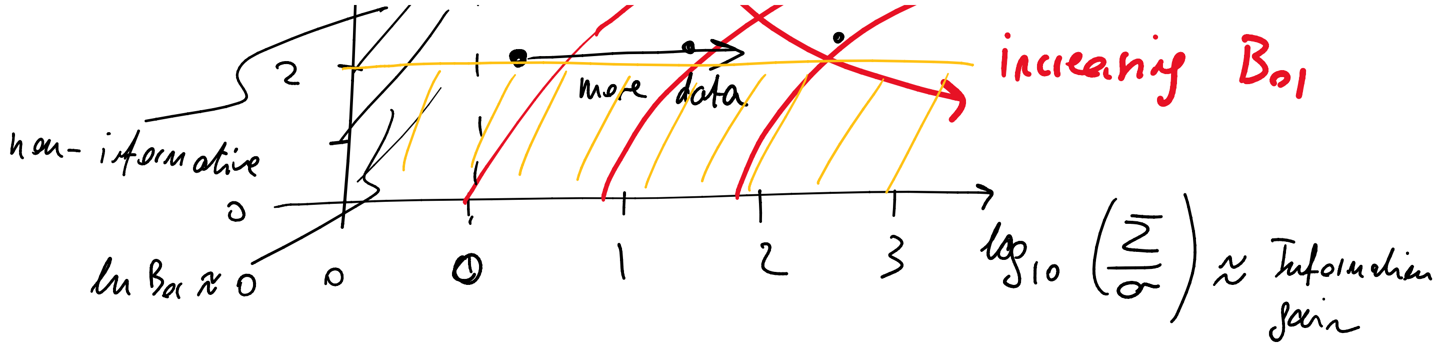
informative data $\left(\frac{\sigma}{\Sigma} \ll 1\right)$

$$\approx \underbrace{\ln\left(\frac{\Sigma}{\sigma}\right)}_{\geq 0 \text{ (Occam's factor)}} - \underbrace{\frac{\lambda^2}{2}}_{\text{evidence in favour of extra parameter}} = \ln B_{01}$$

$\lambda \gg 0$

(Lindley Paradox)






## Computation:

For nested models:

$M_0 : \theta, \; p_0(\theta)$

$M_1 : \{\theta, \omega\}, \; p_1(\theta, \omega) \longrightarrow M_0$ for $\omega = \omega_*$

Assuming separable priors: $p_1(\theta, \omega) = p_1(\omega) \cdot p_0(\theta)$

Then:

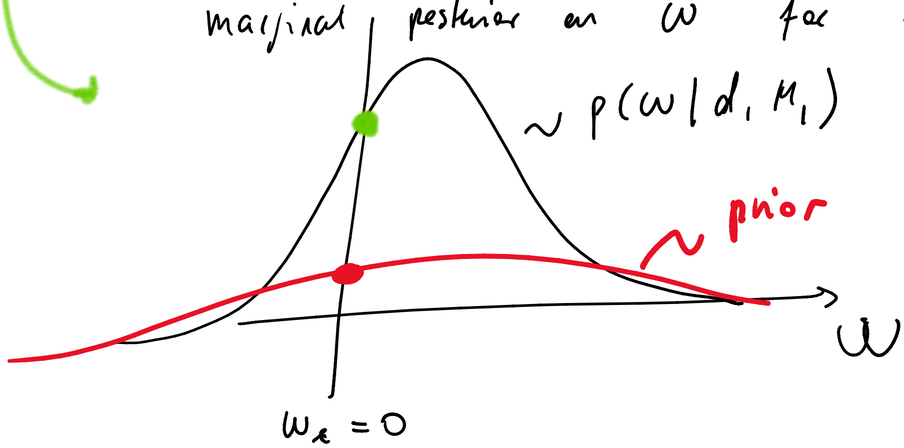
$$\boxed{B_{01} = \frac{p(\omega = \omega_* \mid d, M_1)}{p_1(\omega = \omega_* \mid M_1)}}$$

prior on additional parameter

Savage-Dickey density ratio (SDDR)

doesn't depend on $\theta$; does NOT depend on $p_0(\theta)$

marginal posterior on $\omega$ for Model 1




## Numerical tools:

nested sampling; designed to compute

$p(d \mid M) = \int p(\theta \mid d)\, p(\theta)\, d\theta$

$$P(d|M) = \int P(\theta|d)\,P(\theta)\,d\theta$$

$D \gtrsim 30$ (ellipsoidal) $\longrightarrow$ $\int_0^1 \mathcal{L}(X)\,dX$

For $D \gg 1$ : (PolyChord)

MCMC samples reweighting : $D \gtrsim 10$