Bayesian Hypothesis Comparison

Vaccine vs control treatment

f_V = prob of malaria if V

f_C = " " if C

H_1 : V is better than control $f_V < f_C$

H_0 : $V \simeq C$: $|f_V - f_C| < \epsilon$ [$\epsilon = 0.01$]

$$P(H_1 | D) = \frac{P(D | H_1) P(H_1)}{P(D)}$$

$$P(H_0 | D) = \frac{P(D | H_0) P(H_0)}{P(D)}$$

$$P(D) = \sum_{\text{all hypoths}} P(D | H_i) P(H_i)$$

but

$$\frac{P(H_1 | D)}{P(H_2 | D)} = \frac{P(D | H_1) \cdot P(H_1)}{P(D | H_0) \cdot P(H_0)}$$

if $P(H_1) = P(H_0)$ priors

$$\frac{P(H_1 | D)}{P(H_0 | D)} = \frac{P(D | H_1)}{P(D | H_0)}$$

$H_1 \Leftrightarrow H_0 \longrightarrow$ compare the evidences of D under the two hypothesis

$P(D|H) \rightarrow$ by integrating to all vals of the parameters that define the H

$$P(D|H_1) = \int_{f_c}\int_{f_v} P(D|f_v f_c)\, P(f_c f_v)\, df_c\, df_v$$

$f_v < f_c$ marginalization

$$P(D|H_0) = \int_{f_c}\int_{f_v} P(D|f_c f_v)\, P(f_c f_v)\, df_c\, df_v$$

$f_c - \epsilon \le f_v \le f_c + \epsilon$

$$\frac{P(D|H_1)}{P(D|H_0)} = \frac{\int_0^1 df_c \int_0^{f_c} df_v\; f_v^{n_v} f_c^{n_c} (1-f_v)^{N_v-n_v} (1-f_c)^{N_c-n_c}}{\int_0^1 df_c \int_{f_c-\epsilon}^{f_c+\epsilon} df_v\; f_v^{n_v} f_c^{n_c} (1-f_v)^{N_v-n_v} (1-f_c)^{N_c-n_c}}$$

$N_c = 10 - n_c = 3$

$N_v = 30 - n_v = 9$

$\longrightarrow$ go to code

$$\frac{P(D|H_1)}{P(D|H_0)} =$$

Occam's razor

Why more complicated models can turn out less probable

if several models can explain the observations,
always go with the simpler model

why? simplicity
aesthetics
Bayesian inference

→ BREAKROOMS

Consider 2 nested hypotheses

H_1 # param H_1 > # param H_2

H_2 $H_2 \subset H_1$

(x_i, y_i) $H_1 : y_i = \underline{w_0} + \underline{w_1} x_i$

$H_2 : y_i = \underline{w_0}$

* If ML fit to parameters
  H_1 will allways be more favorable

* If bayesian comparison?

$$\frac{P(H_1 \mid D)}{P(H_2 \mid D)} = \frac{P(D \mid H_1)}{P(D \mid H_2)}$$

$$P(D \mid H_1) = \int_{P_1} P(D \mid P_1 H_1)\, P(P_1)\, dP_1$$

$$P(D \mid H_2) = \int_{P_2} P(D \mid P_2 H_2)\, P(P_2)\, dP_2$$

$$\log P(D \mid P_1 H_1) \simeq \log P(D \mid P_1^*) + \frac{1}{2} \left. \frac{d^2 \log P(D \mid P_1)}{dP_1^2} \right|_{P_1^*} (P - P_1)^2$$

$$= \log P(D \mid P_1^*) - \frac{1}{2} \frac{(P - P_1)^2}{\sigma_1^2}$$

$$P(D \mid P_1 H_1) \simeq P(D \mid P_1^*)\, e^{-\frac{1}{2}\frac{(P-P_1)^2}{\sigma_1^2}}$$

$$P(P_1) = \begin{cases} \dfrac{1}{P^+ - P^-} \equiv \dfrac{1}{\sigma_1} & P^- \le P_1 \le P^+ \\ 0 & \text{otherwise} \end{cases}$$

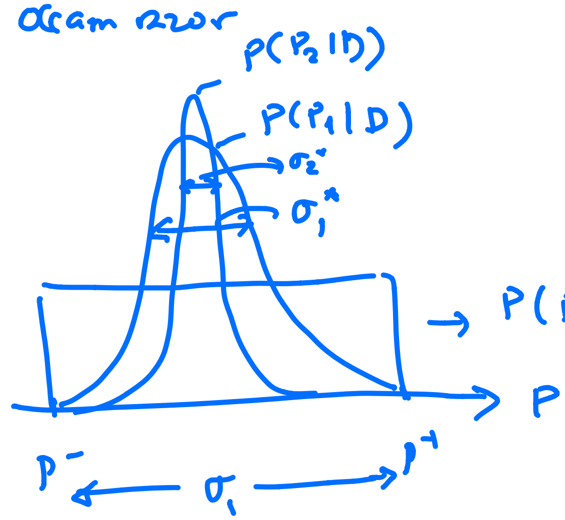
$$P(D \mid H_1) \simeq P(D \mid P_1^* H_1) \int_{P^-}^{P^+} e^{-\frac{(P_1 - P_1^*)^2}{2\sigma_1^{*2}}} \frac{1}{\sigma_1}\, dP_1$$

$$\simeq \overbrace{P(D \mid P_1^* H_1)}^{\text{ML fit}} \cdot \overbrace{\sqrt{2\pi}\, \sigma_1^* \frac{1}{\sigma_1}}^{\text{occam razor term}}$$

$$\frac{P(D|H_1)}{P(D|H_2)} = \frac{P(D|p_1^* H_1)}{P(D|p_2^* H_2)} \cdot \frac{\sigma_1^*/\sigma_1}{\sigma_2^*/\sigma_2}$$

if $H_1 \supset H_2$ $\quad P(D|p_1^* H_1) \gtrsim P(D|p_2^* H_2)$

- Bayesian Occam razor



Example

$$P(t|\lambda_1 H_1) = \frac{e^{-t/\lambda_1}}{Z(\lambda_1)}$$

$$P(t|\lambda_1 \lambda_2 H_2) \propto \eta e^{-t/\lambda_1} + (1-\eta) e^{-t/\lambda_2}$$

$$P(t|\lambda_1 \lambda_2 H_2) = \frac{\eta e^{-t/\lambda_1} + (1-\eta) e^{-t/\lambda_2}}{\eta Z(\lambda_1) + (1-\eta) Z(\lambda_2)}$$

$$P(D|\lambda_1 H_1) = \frac{e^{-\Sigma_i t_i/\lambda_1}}{Z(\lambda_1)}$$

$$P(D|\lambda_1 \lambda_2 H_2) = \prod_i \frac{\eta e^{-t_i/\lambda_1} + (1-\eta) e^{-t_i/\lambda_2}}{\eta Z(\lambda_1) + (1-\eta) Z(\lambda_2)}$$

$$= \frac{\prod_i (\eta e^{-t_i/\lambda_1} + (1-\eta) e^{-t_i/\lambda_2})}{[\eta Z(\lambda_1) + (1-\eta) Z(\lambda_2)]^N}$$

$$P(D|H_1) = \int_{\lambda^-}^{\lambda^+} \frac{e^{-\sum_i t_i/\lambda_1}}{Z^N(\lambda_1)} \cdot \frac{1}{\lambda^+ - \lambda^-} \, d\lambda_1$$

$$P(D|H_2) = \int_{\lambda_2^-}^{\lambda_2^+} \int_{\lambda_1^-}^{\lambda_1^+} \frac{\prod_i \left[\eta e^{-t_i/\lambda_1} + (1-\eta) e^{-t_i/\lambda_2}\right]}{\left[\eta Z(\lambda_1) + (1-\eta) Z(\lambda_2)\right]^N} \left(\frac{1}{\lambda^+ - \lambda^-}\right)^2 d\lambda_1 d\lambda_2$$

$$\lambda_1^+ - \lambda_1^- = \sigma_1$$

$$\lambda_2^+ - \lambda_2^- = \sigma_2$$

$$P(D|H_1) = \frac{1}{\sigma_1} \int_{\lambda_1^-}^{\lambda_1^+} \frac{e^{-\sum_i t_i/\lambda_1}}{Z^N(\lambda_1)} \, d\lambda_1$$

$$P(D|H_2) = \frac{1}{\sigma_1} \frac{1}{\sigma_2} \int_{\lambda_1^-}^{\lambda_1^+} d\lambda_1 \int_{\lambda_2^-}^{\lambda_2^+} d\lambda_2 \times \frac{\prod_i \left[\eta e^{-t_i/\lambda_1} + (1-\eta) e^{-t_i/\lambda_2}\right]}{\left[\eta Z(\lambda_1) + (1-\eta) Z(\lambda_2)\right]^N}$$

$N = 6 \quad t_1 .. t_6 = 1.2, 2.1, 3.4, 4.1, 7, 11$

$\eta = 0.5 \quad \lambda^- = 0.05 \quad \lambda^+ = 80$

$\underline{t} = 1 \quad t_+ = 20$

$P(H_1|D) = e^{-16.24}$

$P(H_2|D) = e^{-16.51}$

$$\frac{P(H_1|D)}{P(H_2|D)} = 1.30 = \frac{0.56}{0.44}$$