Bayesian model Comparison.

• **Basics:**

$$B_{12} = \frac{p(d|\mathcal{M}_1)}{p(d|\mathcal{M}_2)}$$ assumes implicitly $p(\mathcal{M}_1) = p(\mathcal{M}_2)$

ratio of evidences. ($\neq$ ratio of likelihoods in frequentist stats.)

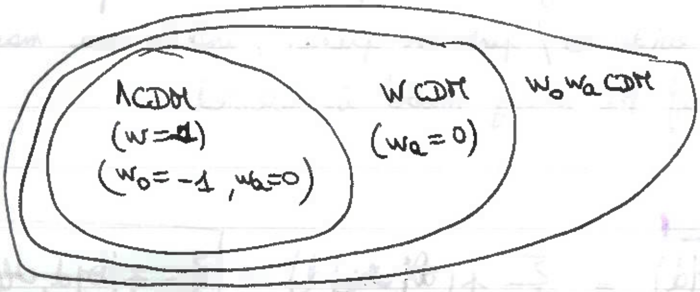
Bayes factor.

$$\underbrace{p(d|\mathcal{M}_1)}_{\text{evidence}} = \int_{\text{param space of } \mathcal{M}_1} \underbrace{p(d|\theta,\mathcal{M}_1)}_{\text{likelihood}} \, \underbrace{p(\theta|\mathcal{M}_1)}_{\text{prior}} \, d\theta .$$

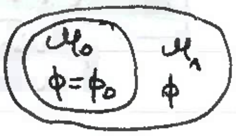
= marginal likelihood

• **Nested models and the Savage-Dickey Density Ratio (SDDR):**

Example: nested models.



suppose



ϕ subset of all params ψ.

$$B_{01} = \frac{p(d|\mathcal{M}_0)}{p(d|\mathcal{M}_1)}$$

$$p(d|\mathcal{M}_0) = \int p(d|\psi,\mathcal{M}_0)\, p(\psi|\mathcal{M}_0)\, d\psi .$$

continuity condition. $p(\psi|\mathcal{M}_0) = p(\psi|\phi=\phi_0,\mathcal{M}_1)$.

$$\Rightarrow p(d|\mathcal{M}_0) = \int p(d|\psi,\phi=\phi_0,\mathcal{M}_1)\, p(\psi|\phi=\phi_0,\mathcal{M}_1)\, d\psi = p(d|\phi=\phi_0,\mathcal{M}_1)$$

$$= \frac{p(\phi=\phi_0|d,\mathcal{M}_1)\, p(d|\mathcal{M}_1)}{p(\phi=\phi_0|\mathcal{M}_1)}$$

$$\Rightarrow \boxed{B_{01} = \frac{p(\phi=\phi_0|d,\mathcal{M}_1)}{p(\phi=\phi_0|\mathcal{M}_1)}}$$ (SDDR). These are usually low-d probabilities.

• **Bayesian model selection as a decision analysis**

$\{\mathcal{M}_k\}_{1\leq k\leq N_m}$: set of models $\{a_k\} = \{\text{"selecting } \mathcal{M}_k\text{"}\}_{1\leq k\leq N_m}$: set of actions.

$\Rightarrow$ Bayesian decision theory says: take action a^*:

$$a^* = \underset{k}{\text{argmax}}\ U(a_k|d) \qquad \boxed{U(a_k|d) = \sum_{k'} G(a_k|\mathcal{M}_{k'})\, p(\mathcal{M}_{k'}|d)}$$

With this formalism, instead of the posterior odds of pairs of models,

$\mathcal{P}_{kk'} = p(\mathcal{M}_k|d) / p(\mathcal{M}_{k'}|d)$, one can quote the ratios

$\mathcal{U}_{kk'} = U(\mathcal{M}_k|d) / U(\mathcal{M}_{k'}|d) \equiv$ Bayes utility factors.

Simplest choice: 0-1 gain functions: $G(a_k|\mathcal{M}_k) = \delta_K^{kk'} \Rightarrow U(a_k|\mathcal{M}_k) = p(\mathcal{M}_k|d)$.

$\mathcal{P}_{kk'} = \mathcal{U}_{kk'}$, we are back to usual Bayesian model comparison.

More generally: one may want to favour models for reasons that are not encoded in their evidences / posterior proba., and/or one may not want to "have lost everything" if the wrong model is selected.

- **Model averaging.**

$$p(\theta|d) = \sum_k p(\theta, \mathcal{M}_k|d) = \boxed{\sum_k \underbrace{p(\theta|d,\mathcal{M}_k)}_{\text{posterior of } \theta \text{ in } \mathcal{M}_k}\ \underbrace{p(\mathcal{M}_k|d)}_{\text{posterior of } \mathcal{M}_k}}$$

in particular:

$$E(\theta|d) = \sum_k E(\theta|d,\mathcal{M}_k)\, p(\mathcal{M}_k|d).$$

$$\mathrm{Var}(\theta|d) = \sum_k \left[\mathrm{Var}(\theta|d,\mathcal{M}_k) + E(\theta|d,\mathcal{M}_k)^2\right] p(\mathcal{M}_k|d) - E(\theta|d)^2$$

for cases where all the $p(\mathcal{M}_k)$ are equal: $\boxed{p(\theta|d) \propto \sum_k \underbrace{p(\theta|d,\mathcal{M}_k)}_{\text{posterior of } \theta \text{ in } \mathcal{M}_k}\ \underbrace{p(d|\mathcal{M}_k)}_{\text{evidence of } \mathcal{M}_k}}$

- **Model selection with insufficient summary statistics.**

within a model $S(d)$ is sufficient for parameter θ if and only if:

$$p(\theta|S(d)) = p(\theta|d,S(d)) \iff p(d|S(d),\theta) = p(d|S(d)).$$
$$= p(\theta|d)$$

$$\underbrace{\boxed{B_{12}}}_{\text{true Bayes factor for full data}} = \frac{p(d|\mathcal{M}_1)}{p(d|\mathcal{M}_2)} = \frac{p(d|S(d),\mathcal{M}_1)}{p(d|S(d),\mathcal{M}_2)}\ \frac{p(S(d)|\mathcal{M}_1)}{p(S(d)|\mathcal{M}_2)} = \boxed{\underbrace{\frac{p(d|S(d),\mathcal{M}_1)}{p(d|S(d),\mathcal{M}_2)}}_{\text{not bounded, a priori!}}\ \underbrace{B_{12}^S}_{\text{approximate Bayes factor from statistics summary.}}}$$

$S(d)$ is sufficient for comparing $\mathcal{M}_1$ and $\mathcal{M}_2$ if and only if

$p(d|S(d),\mathcal{M}_1) = p(d|S(d),\mathcal{M}_2) \quad (\Rightarrow B_{12} = B_{12}^S)$

|| Suffiency for $\mathcal{M}_1$, $\mathcal{M}_2$ or even both does not imply sufficiency for comparing $\mathcal{M}_1$ and $\mathcal{M}_2$. (Didelot et al 2011)

Danger in part. for ABC: B_{12}^S can be biased, and the approximation error unrelated to the computational effort (Robert et al. 2011)