

Start from the likelihood,  $L_{\text{model}} = P(\text{data} | \text{model})$

Maximum likelihood estimation: find parameters that maximize  $L$ , within a given model

We can compare 2 models by taking the ratio of their maximum likelihoods / difference of their log likelihoods,

log likelihood ratio test statistic

$$TS = 2 \ln \left( \frac{\max(L_M)}{\max(L_{M_0})} \right)$$

Typical application: models are nested, want to understand if adding additional parameters significantly improves the model (e.g. extra parameter could be coefficient of a new template)

If  $M = \text{model}$ ,  $M_0 = \text{same model with parameters restricted to specific values (could be zero)}$

$$\max(L_M) \geq \max(L_{M_0})$$

$$TS = 2 \ln \left( \frac{\max(L_M)}{\max(L_{M_0})} \right)$$

Question: what does a given value of TS mean, in terms of e.g. confidence level for exclusion of  $M_0$ ?

Wilks' theorem: as sample size  $\rightarrow \infty$ , distribution of TS under the hypothesis  $M_0$  approaches a  $\chi^2$  distribution with degrees of freedom equal to difference in dimensionality of  $M, M_0$

- assumes values of parameters lie in interior of supported parameter space (i.e. may not hold if maximum  $L$  occurs on boundary)

( $\chi^2$ : distribution of sum of squares of  $k$  independent

standard normal random variables,  $k = \# \text{ of dof}$ )

If WT does not apply, need to run simulations assuming  $M_0$  to figure out the expected distribution of TS.

### Bayesian statistics:

Work with posterior probability distribution,

$$P(\text{model} | \text{data}) = \frac{P(\text{data} | \text{model}) \cdot P(\text{model})}{P(\text{data})} \xrightarrow{\text{posterior}} \text{likelihood} \quad \xrightarrow{\text{prior}} \text{overall normalization factor}$$

If the model is described by a set of parameters  $\{\theta\}$ , we can define prior probability distributions for the parameters (what I called  $P(\text{model})$ ).

$$\text{Bayesian evidence: } P(\text{data}) = \int P(\text{data} | \{\theta\}) P(\{\theta\}) d\theta$$

(can think of this as like the total probability of finding the observed data given any allowed set of parameters in model)

Posterior probability distributions tell us about preferred values of model parameters (if we can define "credible intervals" within which a certain fraction of the probability distribution lies)

Bayes factor between two models = ratio of their evidences