Lecture 4

## Model fit diagnostics

Let $k$ = # parameters

$\hat{L}$ = maximized likelihood

$$AIC = 2k - 2\ln(\hat{L})$$

$$BIC = k\ln(n) - 2\ln(\hat{L})$$

Smaller is better!

penalty

reward

Both criteria

- reward goodness of fit
- penalize complexity
- note: BIC more severly penalizes complexity than AIC

Consider Bayes' Thm

$$\underbrace{p(m \mid \bar{y})}_{\text{posterior prob of model}} = \underbrace{p(m)}_{\text{prior prob of } m} \times \underbrace{\frac{p(\bar{y} \mid m)}{p(\bar{y})}}_{\text{updating factor}}$$

Model comparison:

$$\underbrace{\frac{p(m_1 \mid \bar{y})}{p(m_2 \mid \bar{y})}}_{\text{posterior odds}} = \underbrace{\frac{p(m_1)}{p(m_2)}}_{\text{prior odds}} \times \underbrace{\frac{p(\bar{y} \mid m_1)}{p(\bar{y} \mid m_2)}}_{\text{updating factor}}$$

updating factor "Bayes Factor" (Kass & Raftery, '95)

Note: the notation hides some complexity:

This is a "marginal likelihood":

$$p(\bar{y} \mid m_i) = \int p(\bar{y} \mid \theta, m_i)\, p(\theta \mid m_i)\, d\theta$$

## Bayes factor

$$B_{12} = \frac{p(\vec{y} \mid M_1)}{p(\vec{y} \mid M_2)} = \text{ratio of marginal likelihoods.}$$

**Interp:** $B_{12}$ is the extent to which data $\vec{y}$ is better predicted by $M_1$ than $M_2$.

**Note:** if $B_{12} < 1$, then $M_2$ better predicts data.

**Note:** $B_{12}$ is **hard** to compute! in general.

**But**, there is a useful approximation for our purposes:

**Fact**

$$B_{12} \approx \exp\left(\frac{BIC_2 - BIC_1}{2}\right)$$

Example: Murdock (1961) data.

$\mathcal{M}_1$: power function

$$\ln(\hat{L}) = -313.365$$

$\mathcal{M}_2$: exponential function

$$\ln(\hat{L}) = -305.306$$

$$BIC_1 = k \ln(n) - 2 \ln(\hat{L})$$

$$= 2 \ln(5) - 2(-313.365)$$

$$= 629.95$$

$$BIC_2 = 2 \ln 5 - 2(-305.306)$$

$$= 613.83$$

better fit!

So $$B_{21} = \exp\left(\frac{629.95 - 613.83}{2}\right)$$

$$= 3165.29$$

$\Rightarrow$ The observed data are 3165 times more likely under the exponential model than the power model.

# Exercises

1. Compute BICs for power/exponential models based on the Rubin & Baddeley data. Estimate a Bayes factor for the winning model. Interpret?

2. (we'll need this tomorrow) Write a Python function that takes data as input and returns

   1) MLEs for parameters a, b

   2) BIC for the fit.

   Hint: think about the structure of the returned object. Consider a tuple or dict.