Issues with Bayes Factors

- **Note**: When an **improper prior** is used for θ, the Bayes Factor is not well-defined.
- Note $B(\mathbf{x}) = \frac{\text{Posterior odds for } M_1}{\text{Prior odds for } M_1}$, and the “prior odds” is meaningless for an improper prior.
- Gill's Sec. 7.3.2 suggests several methods (Local Bayes factors, Intrinsic Bayes Factors, Partial Bayes Factors, Fractional Bayes Factors), none of them ideal, to define types of Bayes Factors with improper priors.
- One criticism of Bayes Factors is the (implicit) assumption that one of the competing models (M_1 or M_2) is correct.
- Another criticism is that the Bayes Factor depends heavily on the choice of prior.

The Bayesian Information Criterion

- The Bayesian Information Criterion (BIC) can be used (as a substitute for the Bayes factor) to compare two (or more) models.
- Conveniently, the BIC does **not** require specifying priors.
- For parameters $\boldsymbol{\theta}$ and data $\mathbf{x}$:

$$BIC = -2\ln L(\hat{\boldsymbol{\theta}}|\mathbf{x}) + p\ln(n)$$

  where p is the number of free parameters in the model, and $L(\hat{\boldsymbol{\theta}}|\mathbf{x})$ is the **maximized likelihood**, given observed data $\mathbf{x}$.
- Good models have relatively small BIC values:
  - A small value of $-2\ln L(\hat{\boldsymbol{\theta}}|\mathbf{x})$ indicates good fit to the data;
  - a small value of the "overfitting penalty" term $p\ln(n)$ indicates a simple, parsimonious model.

The Bayesian Information Criterion

- To compare two models M_1 and M_2, we could calculate

$$S = -\frac{1}{2}[BIC_{M_1} - BIC_{M_2}]$$
$$= \ln L(\hat{\boldsymbol{\theta}}_1|\mathbf{x}) - \ln L(\hat{\boldsymbol{\theta}}_2|\mathbf{x}) - \frac{1}{2}(p_1 - p_2)\ln(n)$$

- A small value of S would favor M_2 here and a large S would favor M_1.
- As $n \to \infty$,

$$\frac{S - \ln(B(\mathbf{x}))}{\ln(B(\mathbf{x}))} \to 0$$

and for large n,

$$BIC_{M_1} - BIC_{M_2} = -2S \approx -2\ln(B(\mathbf{x})).$$

The Bayesian Information Criterion

- Note that differences in BIC's can be used to compare several nonnested models.
- They should be trusted as a substitute for Bayes Factors only when (1) no reliable prior information is available and (2) when the sample size is **quite large**.
- See `R` examples: (1) Calcium data example and (2) Regression example on Oxygen Uptake data set.

CHAPTER 10 SLIDES BEGIN HERE

Hierarchical Models

- In **hierarchical Bayesian estimation**, we not only specify a prior on the data model's parameter(s), but specify a further prior (called a **hyperprior**) for the hyperparameters.
- This more complicated prior structure can be useful for modeling **hierarchical data structures**, also called *multilevel data*.
- Multilevel data involves a hierarchy of nested populations, in which data could be measured for several levels of aggregation.

Examples:

- We could measure white-blood-cell counts for numerous patients within several hospitals.
- We could measure test scores for numerous students within several schools.

Hierarchical Bayes Estimation

- Assume we have data **x** from density $f(\mathbf{x}|\theta)$ with a parameter of interest $\boldsymbol{\theta}$.
- Typically we would choose a prior for θ that depends on some hyperparameter(s) $\boldsymbol{\psi}$.
- Instead of choosing fixed values for $\boldsymbol{\psi}$, we could place a **hyperprior** $p(\boldsymbol{\psi})$ on it.
- Note that this hierarchy could continue for any number of levels, but it is rare to need more than two levels for the prior structure.

Hierarchical Bayes Estimation

- Our posterior is then:

$$\pi(\theta, \boldsymbol{\psi}|\mathbf{x}) \propto L(\theta|\mathbf{x})p(\theta|\boldsymbol{\psi})p(\boldsymbol{\psi})$$

- Posterior inference about θ is based on the **marginal** posterior for θ:

$$\pi(\theta|\mathbf{x}) = \int_{\boldsymbol{\psi}} \pi(\theta, \boldsymbol{\psi}|\mathbf{x})\, \mathrm{d}\boldsymbol{\psi}$$

- Except in simple situations, such analysis typically requires MCMC methods.