Model selection criteria

- Cross-validation is great for large datasets, but can’t be applied for small datasets.

- Bayes factors are hard to compute for complex models.

- There are several model-selection criteria to fill these gaps.

- We will explore several approaches for choosing between models:
  - LPML
  - DIC
  - WAIC
  - Prediction criteria

- Code is available at
  http://www4.stat.ncsu.edu/~reich/ST740/code/DIC_LMPL_PD.R.
Log pseudo marginal likelihood (LPML)

- LPML is leave-one-out ($n$-fold) cross-validation with log likelihood as the criteria,

\[ LPML = \sum_{i=1}^{n} \log(CPO_i) \quad \text{and} \quad CPO_i = f(y_i | y_{(-i)}). \]

- $CPO_i$ is the conditional predictive ordinate and $y_{(-i)}$ is the data set without $y_i$.

- We pick the model with largest LPML.

- Gelfand and Day show that you can compute LPML with a single MCMC chain:
Deviance information criteria (DIC)

- Many model selection criteria are based on the deviance $D(y|\theta) = -2 \log[f(y|\theta)]$:

  \[
  AIC = D(y|\hat{\theta}) + 2 \dim(\theta) \\
  BIC = D(y|\hat{\theta}) + \log(n) \dim(\theta)
  \]

  where $\hat{\theta}$ is the MLE.

- The deviance $D(y|\hat{\theta})$ penalizes lack of fit and $\dim(\theta)$ penalizes complexity.

- Smaller values are preferred.

- Problems for Bayesians:
Deviance information criteria (DIC)

- DIC handles these issues. It is
  \[ DIC = \bar{D} + p_D. \]

- \( \bar{D} = E_{\theta|y} [D(y|\theta)] \) is the posterior mean of the deviance and penalizes lack of fit.

- \( \hat{D} = D(y|\hat{\theta}) \) is the deviance evaluated at the posterior mean (or median) of \( \theta \).

- \( p_D = \bar{D} - \hat{D} \) is the effective model size and penalizes complexity.

- We choose the model with smallest DIC.

- Where does it come from?
Watanabe-Akaike information criteria (WAIC)

- WAIC is another criteria that is decomposed as terms for fit and complexity.

- Fit is measured by
  \[
  DEV_W = \sum_{i=1}^{n} \log E_{\theta|y} [f(y_i|\theta)].
  \]

- Complexity is measured by
  \[
  p_W = \sum_{i=1}^{n} V_{\theta|y} [\log p(y_i|\theta)].
  \]

- Then WAIC = $DEV_W + p_W$ and small WAIC is preferred.

- Where does it come from?
Posterior predictive model selection

- Laud and Ibrahim propose a class of criteria based on sampling many replicate datasets.
- Let $Y^*$ be a posterior sample data set drawn at the same design points as $Y$.
- If the model is correct, $Y^*$ should be similar to $Y$.
- To quantify the difference, define discrepancy measure $d(Y, Y^*)$.
- The final criteria is the posterior mean discrepancy $\bar{d}$; models with small $\bar{d}$ are preferred.
- Computed in MCMC as:

- Example discrepancy measures: