STAT-511 **2018 Spring**

Lecture 7: Model Selection

Instructor: Naveen N. Narisetty Scribe: Yutong Dai Last Modified: 2018-05-07

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications, and they may contain factual and/or typographic errors.*

7.1 Criteria: AIC

Goal: Find the model that gives the best prediciton without assuming the any candidate model is correct.

Setup: Suppose $x_i \overset{iid}{\sim} f_\theta$ and the candidate set is $\mathcal{M} = \{M_j | j = 1, 2, ..., n\}$ where $M_j = \{p_{\theta_j}(x); \theta_j \in \Theta_j\}$. We wan to minimize the loss

$$\min_j KL(f_\theta || p_{\hat{\theta_j}}) = \int (f_\theta \log f_\theta - f_\theta \log p_{\hat{\theta_j}}) d\theta,$$

where $p_{\hat{\theta_j}}$ is an estimate of p_{θ_j} $(p_{\hat{\theta_j}} = p_{\hat{\theta}_j})$. This is equivalent to

$$\min_j - \int f_\theta \log p_{\hat{\theta_j}} d\theta.$$

A good estimator of $\int f_\theta \log p_{\hat{\theta_j}} d\theta$ is

$$\bar{K}_j = \frac{1}{n} \sum_i \log p_{\hat{\theta_j}}.$$

7.2 Criteria: BIC