# STAT 151A Lecture 29

Henry Liev

3 November 2025

**Remark 0.1** (Where Does AIC Come From)

$K-L$ divergence $\mathcal{J}(f,f_m) = \int \log\left[\frac{f(y)}{f_m(y)}\right] f(y)dy = \int f(y)\log f(y)dy - \int f(y)\log f_m(y)dy = C - \mathbb{E}[\log f_m(y)]$
Approximate $\mathbb{E}[\log f_m(y)] \to \frac{1}{n}\sum \log[f_m(y_i)] = \frac{1}{n}\log\prod f_m y_i = \frac{1}{n}\log\mathcal{L}(\beta_m|\vec{y})$
To minimize $K-L$, we can maximize the likelihood, approximate with $\hat{\beta}_m$
$\frac{1}{n}\log\mathcal{L}(\hat{\beta}_m|\vec{y})$
One problem is that we did two approximations using the same data $\to$ slightly baised, likelihood tends to be a little too high
Correct for bias $\frac{1}{n}\log\mathcal{L}(\hat{\beta}_m|\vec{y}) - \frac{(p_m+1)}{n}$
Goal: Pick lowest $K-L$ divergence
$\arg\min_m \mathcal{J}(f,f_m) \approx \arg\min_m\left[C - \frac{1}{n}\log\mathcal{L}(\hat{\beta}_m) + \frac{p_m+1}{n}\right] = \arg\min_m[-2\log\mathcal{L}(\hat{\beta}_m) + 2(p_m+1)]$
For NLM:
$\log\mathcal{L}(\hat{\beta}_m) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum(y_i - x_{i_m}^\intercal\hat{\beta}_m)^2$
$-2\log\mathcal{L}(\hat{\beta}_m) = n\log 2\pi + n\log\hat{\sigma}^2 + \frac{1}{\hat{\sigma}^2}ErrSS = C + n\log\frac{ErrSS}{n}$
$AIC = -2n\log\frac{ErrSS}{n} + 2(p_m+1) + C$

**Remark 0.2** (BIC)

$BIC = -2\log\mathcal{L}(\hat{\beta}_m) + \boxed{\log(n)}(p_m+1)$
Compare to AIC
$AIC = -2\log\mathcal{L}(\hat{\beta}_m) + \boxed{2}(p_m+1)$
Boxed is the difference between AIC and BIC
Where does BIC come from?
Maximize the posterior probability of $\hat{\beta}_m$ in a Bayesian $\leftrightarrow$ very close to a likelihood slightly different bias correction

**Remark 0.3** (We Have a Criterion, What Next?)

One way to proceed: All subsets/Best subset regression

(1) Start by fitting all possible models ($2^p$) using a subset of the $p$ predictors

(2) Figure out the best 5 models for each possible model size $1, \ldots, p-1$ According to our criterion (or ErrSS)

(3) Look at which variables are included in these models $\rightarrow$ diagnostic/exploratory step

(4) Choose the best model either using criterion or if step 3 suggests something a little different