akaike74
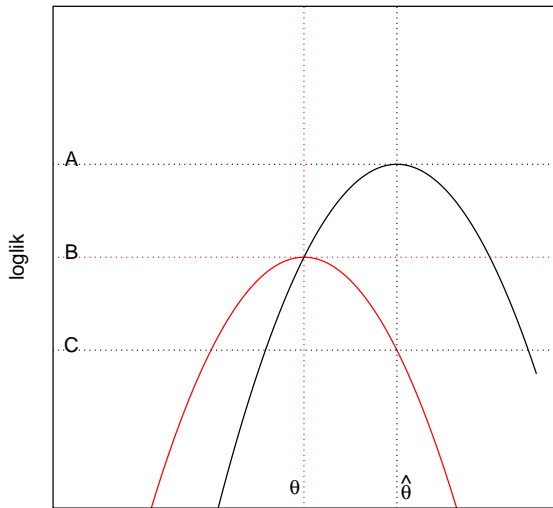
STATS 631, Winter 2026

- over 70,000 citations on Google Scholar
- Akaike has a sequence of papers leading to AIC and then investigating it. This paper focuses on time series and gives perhaps the clearest derivation of AIC.

- ▶ A : loglik at MLE
- ▶ B : loglik at truth, estimates $E_{\theta_0}[\log f(X; \theta_0)]$
- ▶ A-B : expected over-fitted likelihood for this dataset is $\dim(\theta)/2$
- ▶ B-C : expected log-likelihood penalty for using $\hat{\theta}$ on a new dataset is the same as the over-fitting.
- ▶ C : estimates $E_{\theta_0}[\log f(X; \hat{\theta})]$

- ▶ What are the alternatives to AIC?
  - ▶ BIC
  - ▶ WAIC
  - ▶ GCV

Some derivations are at

https://dept.stat.lsa.umich.edu/~kshedden/Courses/Regression_Notes/model_selection.pdf

"Although the present author has no proof of optimality of MAICE it is at present. the only procedure applicableto every situation where the likelihood can be properly defined and itis actually producing very reasonable results without very much amount of help of subjective judgement. The successful results of numerical experiments suggest almost unlimited applicability of MAICE in the fields of modeling, prediction, signal detection, pattern recognition. and adaptation."

▶ AIC penalizes model complexity so far as it affects expected predictive skill measured by log-likelihood. Additional parsimony may have scientific value.