Akaike Information Criterion (AIC)

for model selection in ARMA(p,q):

minimize with respect to p, q

$$AIC(p,q) = -2\,\underbrace{\ell}_{\text{log-likelihood}}\big(\underbrace{\hat{\underline{\phi}}_p, \hat{\underline{\theta}}_q, \hat{\sigma}_\varepsilon^2}_{\text{MLE estimates}};\ \underbrace{X_1^n}_{\text{data}}\big) + 2(p+q)$$

or

$$\underbrace{2(p+q+1)}_{2\cdot(\text{no. of parameters})}$$

in general, for any reasonable parametric model

$$AIC = -2\,\ell\big(\hat{\beta}_{MLE}; X_1^n\big) + 2\dim(\beta)$$

interpretation:

$-2\,\ell(\hat{\beta}_{MLE}; X_1^n)$: in-sample goodness of fit measure

$\searrow$ as $\dim(\beta)$ increases

$2\dim(\beta)$: penalty term (for complexity of model)

Reasoning for AIC

- data are realizations of $X_1, X_2, \ldots, X_n$ from stochastic process with probability distribution P^*

- class of ARMA(p,q) models:
  probability distributions $\{ P_{\beta_{p,q}} : \beta_{p,q} = (\underline{\phi}_p, \underline{\theta}_q, \sigma_\varepsilon^2) \in A \subset \mathbb{R}^{p+q+1}, \; p,q \in \mathbb{N}_0 \}$

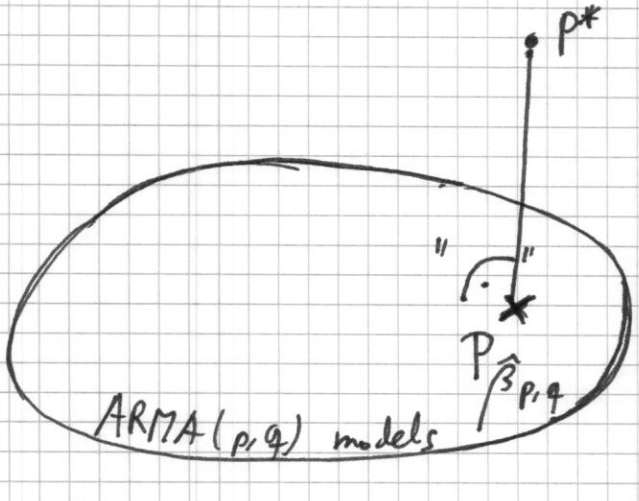
- pseudo-distance between probability distributions:

$$I_n(P, Q) = \int_{\mathbb{R}^n} \log\left(\frac{dP(y_1^n)}{dQ(y_1^n)}\right) dP(y_1^n)$$ Kullbach-Leibler information

aim: search for model such that

$I_n(P^*, P_{\hat{\beta}_{p,q}})$ minimal

$\hat{\beta}_{p,q}$: MLE in ARMA(p,q)



<u>interpretation</u>:

$$I_n(P^*, P_{\hat{\beta}_{p,q}}) = \int_{\mathbb{R}^n} \log\left(dP^*(y_1^n)\right) dP^*(y_1^n) - \int_{\mathbb{R}^n} \log\left(dP_{\hat{\beta}_{p,q}}(y_1^n)\right) dP^*(y_1^n)$$

$$= \underset{\substack{| \\ \text{w.r.t. model selection}}}{\text{const.}} + \mathbb{E}_{Y_1^n}\Big[-\log\big(dP_{\hat{\beta}_{p,q}}(Y_1^n)\big)\Big]$$

$\hat{\beta}_{p,q}$ based on <u>training data</u> $X_1, \dots, X_n$

where $Y_1^n \sim P^*$ is an independent copy of the data

to be interpreted as new <u>test data</u>

that is: $\underbrace{\mathbb{E}_{Y_1^n}\Big[-\log\big(dP_{\hat{\beta}_{p,q}}(Y_1^n)\big)\Big]}_{=T}$ = expected negative log-likelihood evaluated at test data

which is to be minimized

and it holds: AIC is a reasonable estimate of $2T$

(proof is a bit longer...)

$\rightsquigarrow$ $AIC(p,q)$ "small" $\Longleftrightarrow$ $I_n(P^*, P_{\hat{\beta}_{p,q}})$ "small"