STAT 3690 Lecture 18

zhiyanggeezhou.github.io

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca)

Mar 11, 2022

Testing for nested models

- $H_0 : \mathrm{E}(\mathbf{Y} \mid \mathbf{X}) = \mathbf{X}_{(0)}\mathbf{B}_{(0)}$ (nested model) vs. $H_1 : \mathrm{E}(\mathbf{Y} \mid \mathbf{X}) = \mathbf{X}_{(0)}\mathbf{B}_{(0)} + \mathbf{X}_{(1)}\mathbf{B}_{(1)}$ (full model)
  - $\mathbf{X} = [\mathbf{X}_{(0)}, \mathbf{X}_{(1)}]$
  - When $\mathbf{X}_{(0)}$ has only the column of ones, we are testing the empty model (i.e., only the intercept) against the full model.
  - When $\mathbf{X}_{(1)}$ only contains one column, we are testing for the significance of that variable.
- Likelihood ratio

$$\lambda = \left(\frac{\det \widehat{\mathbf{\Sigma}}_{\mathrm{ML},H_0}}{\det \widehat{\mathbf{\Sigma}}_{\mathrm{ML}}} \right)^{-n/2} = \left[\det \left\{ (\widehat{\mathbf{\Sigma}}_{\mathrm{ML},H_0} - \widehat{\mathbf{\Sigma}}_{\mathrm{ML}})\widehat{\mathbf{\Sigma}}_{\mathrm{ML}}^{-1} + \mathbf{I} \right\} \right]^{-n/2}$$

- Alternatives to likelihood ration
  - Suppose $\eta_1 \geq \cdots \geq \eta_p$ are eigenvalues of $(\widehat{\mathbf{\Sigma}}_{\mathrm{ML},H_0} - \widehat{\mathbf{\Sigma}}_{\mathrm{ML}})\hat{\mathbf{\Sigma}}_{\mathrm{ML}}^{-1}$
  - Wilks' lambda: $\prod_i (1 + \eta_i)^{-1}$
  - Pillai's trace: $\sum_i \{\eta_i (1 + \eta_i)^{-1}\}$
  - Hotelling-Lawley trace: $\sum_i \eta_i$
  - Roy's largest root: $\eta_1 (1 + \eta_1)^{-1}$
  - When $\mathbf{X}_{(1)}$ has only one column, all four tests are equivalent; as n increases, all four tests give similar results.

Information criteria

- Akaike's information criterion (AIC)

$$-\ln \mathit{Likelihood} + 2 \times \text{number of parameters to estimate}$$

  - Number of parameters to estimate in $\mathbf{B}$ and $\mathbf{\Sigma}$: $p(q+1) + p(p+1)/2$
  - Smaller is better
- Bayesian information criterion (BIC)

$$-\ln \mathit{Likelihood} + \ln n \times \text{number of parameters to estimate}$$

- Model selection using information criteria proceeds as follows
  - Select models of interest $M_1, \ldots, M_K$. They do not need to be nested.
    - Candidate models should be selected using domain-specific expertise, if possible. Or, you can go through all possible models.
  - Compute the specific information criterion for each model.
  - Select the model with the smallest value of the information criterion.