Optional practice problems

Stat 151A, Fall 2017

October 19, 2017

1. For the following statements determine whether they are true of false. In each case provide a reason behind your choice.

   (a) AIC tends to favor model selection of models with larger number of parameters.

   (b) Say the true model is

   $$y_i = x_i^T\beta + z_i^T\delta + e_i, \quad i = 1, \ldots, n.$$

   Then if we use only the x variables, we get an unbiased estimator of β.

   (c) The studentized residuals have variance equal to one.

   (d) Cook distance can only detect outliers on x.

   (e) Any high leverage point is also an influential point.

2. The following problems from the book: 11.1, 11.3, 12.2, 12.3, 12.4, 12.5.

3. Recall that the delta method tells us that

   $$Var(h(y)) \approx h'(Ey)^2 Var(y).$$

   Prove that $sin^{-1}(\sqrt{Y})$ is approximately variance stabilizing when $Y_i = X_i/m_i$ is a proportion distributed as binomial with X_i successes and m_i trials.

   $$\frac{d}{dz} sin^{-1}(z) = \frac{1}{\sqrt{1-z^2}}$$

   $$\int sin^{-1}(z)dz = z sin^{-1}(z) + \sqrt{1-z^2} + C$$

4. Let h_i be the leverage point for an observation i.

   (a) Show that $\sum h_i = tr(H) = p + 1$.
   Hint: Recall that $tr(ABC) = tr(BCA) = tr(CAB)$ when the dimension of the matrices are such that all of those operations are well defined.

   (b) What implications does this have for interpreting large leverage values?