Homework to Week 9

Statistics: Principle, Methods and R (II)

Gao Fengnan

27th May 2017

The homework is due on Monday, 5th June 2017. Please hand in the solutions to the teaching assistant He Siyuan at the beginning of the lecture.

1. Create an example like Example 16.2 on page 253 on *All of Statistics* in which $\alpha > 0$ and $\theta < 0$.

2. Suppose you are given data $(X_1, Y_1), \ldots, (X_n, Y_n)$ from an observational study, where $X_i \in \{0, 1\}$ and $Y_i \in \{0, 1\}$. Although it is not possible to estimate the causal effect θ, it is possible to put bounds on θ. Find upper and lower bounds on θ that can be consistently estimated from the data. Show that the bounds have width 1. Hint: Note that $\mathbb{E}(C_1) = \mathbb{E}(C_1|X = 1)P(X = 1) + \mathbb{E}(C_1|X = 0)\mathbb{P}(X = 0)$.

3. In this exercise, we use proof by induction to show that the linear projection onto an M-dimensional subspace that maximizes the variance of the projected data is defined by the M eigenvectors of the data covariance matrix S, corresponding to the M largest eigenvalues. In class, this result was proven for the case of $M = 1$. Now suppose the result holds for some general value of M and show that it consequently holds for dimensionality $M + 1$. To do this, first set the derivative of the variance of the projected data with respect to a vector u_{M+1} defining the new direction in data space equal to zero. This should be done subject to the constraints that u_{M+1} be orthogonal to the existing vectors $u_1, \ldots, u_M$, and also that it be normalized to unit length. Use Lagrange multipliers to enforce these constraints. Then make use of the orthonormality properties of the vectors $u_1, \ldots, u_M$ to show that the new vector is an eigenvector of S. Finally, show that the variance is maximized if the eigenvector is chosen to be the one corresponding to eigenvector λ_{M+1} where the eigenvalues have been ordered in decreasing value.