臺灣大學數學系
99 學年度上學期博士班資格考試題
科目：迴歸分析 2010.09.17

1. (25%) Suppose that, conditional on the covariates $\mathbf{X} \in R^p$, the Y's are independent $0-1$ variables, with logit $P(Y_i = 1|\mathbf{X}_i) = \mathbf{X}_i\boldsymbol{\beta}$, i.e., the logit model holds.

   (a) (8%) Show that the log likelihood function can be written as

   $$L_n(\boldsymbol{\beta}) = -\left(\sum_{i=1}^{n} \log[1+\exp(\mathbf{X}_i\boldsymbol{\beta})]\right) + \left(\sum_{i=1}^{n} \mathbf{X}_i Y_i\right)\boldsymbol{\beta}.$$

   (b) (12%) Show that $L_n(\boldsymbol{\beta})$ is a concave function of $\boldsymbol{\beta}$, and strictly concave if the design matrix formed by the $\mathbf{X}_i$ has full rank.

   (c) (5%) Discuss the existence of maximum likelihood estimate of $\boldsymbol{\beta}$ using (a) and (b).

2. (25%) Consider data $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$. Suppose that the postulated regression model is

   $$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad 1 \le i \le n,$$

   where

   $$\epsilon_1, \cdots, \epsilon_n \overset{i.i.d.}{\sim} N(0, \sigma^2).$$

   In fact, the true model is

   $$y_i = \theta(x_i) + e_i, \quad 1 \le i \le n,$$

   where $\theta(\cdot)$ is bounded, $E(e_i) = 0$, $Var(e_i) = \sigma_0^2$, and

   $$Cov(e_i, e_j) = \begin{cases} \rho\sigma_0^2 & \text{when } |i-j| = 1, \\ 0 & \text{otherwise} \end{cases}$$

   Let $\hat{\beta}_0$ and $\hat{\beta}_1$ denote the least squares estimators of β_0 and β_1, respectively.

   (a) (7%) Determine $E(\hat{\beta}_1)$ and $Var(\hat{\beta}_1)$.

   (b) (5%) Describe the distribution of $\hat{\beta}_1$ if $(e_1, \cdots, e_n)$ has a multivariate normal distribution. Give reason to justify your result.

   (c) (13%) When $x_i = i/n$ and n is large, does $\hat{\beta}_1$ converge to a constant a_1 in probabiity? If it does, describe a_1 in terms of $\theta(\cdot)$. Justify your answer.

3. (25%) Let $Y_{ij} = \alpha_i + \beta x_{ij} + \epsilon_{ij}$, where $i = 1, 2, \ldots, I$; $j = 1, 2$; and the ϵ_{ij} are independently distributed as $N(0, \sigma^2)$.

   (a) (13%) Obtain the maximum likelihood estimate of β, $\hat{\beta}$, and give condition in terms of x_{ij} to ensure that $\hat{\beta}$ is a consistent estimate of β. Justify your answer.

   (b) (12%) Obtain a test statistic and testing procedure for testing the hypothesis

   $$H_0 : \alpha_i = \alpha, \quad \text{all } i.$$

4. (25%) Consider the multiple regression models

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i, \quad 1 \le i \le n,$$

where $\sum_{i=1}^n X_{i1} = \sum_{i=1}^n X_{i2} = \sum_{i=1}^n X_{i1}X_{i2} = 0$. It is known that $|\beta_1| + |\beta_2| \le 1$.

(a) (12%) Derive the least squares estimate of $(\beta_0, \beta_1, \beta_2)$, $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$, satisfying the constraint $|\beta_1| + |\beta_2| \le 1$.

(b) (13%) When $\beta_1 = 0$, $\beta_2 = 1$, $\sum_{i=1}^n X_{i1}^2 = \sum_{i=1}^n X_{i2}^2 = n$, and ϵ_i's are iid with distribution $N(0, \sigma^2)$, derive the joint distributions of $(\hat{\beta}_1, \hat{\beta}_2)$.