Ph.D. Qualifying Exam A

Probability Distributions and Regression Analysis

May 26, 2023

1. (20 points) Suppose $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, distribution, which is partitioned as $\begin{bmatrix} \mathbf{X_1} \\ \mathbf{X_2} \end{bmatrix}$, with mean and variance-covariance matrix, $\begin{bmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{bmatrix}$, $\begin{bmatrix} \boldsymbol{\Sigma_{11}} & \boldsymbol{\Sigma_{12}} \\ \boldsymbol{\Sigma_{21}} & \boldsymbol{\Sigma_{22}} \end{bmatrix}$, respectively. Let $\mathbf{X_1}$ be of dimension $m < n$ and $\mathbf{X_2}$ of dimension $n - m$. Assume that $\boldsymbol{\Sigma}$ is positive definite. Then derive the conditional distribution of $\mathbf{X}_1|\mathbf{X}_2$ to be $N_m(\boldsymbol{\mu_1} + \boldsymbol{\Sigma_{12}}\boldsymbol{\Sigma_{22}^{-1}}(\mathbf{X_2} - \boldsymbol{\mu_2}), \boldsymbol{\Sigma_{11}} - \boldsymbol{\Sigma_{12}}\boldsymbol{\Sigma_{22}^{-1}}\boldsymbol{\Sigma_{21}})$.

2. (20 points) Prove that

$$\lim_{n\to\infty}\left[1 + \frac{b}{n} + \frac{\psi(n)}{n}\right]^{cn} = e^{bc},$$

where b and c do not depend on n and

$$\lim_{n\to\infty}\psi(n) = 0.$$

3. (20 points) Let (X, Y) be bivariate random variables whose joint cumulative distribution function (c.d.f.) is continuous. Show how the computation of $P(x < X \leq x + \epsilon_1, y < Y \leq y + \epsilon_2)$, for all x, y and any $\epsilon_1, \epsilon_2 > 0$, gives rise to the joint probability density function (p.d.f.) of (X, Y). Give an example of the use of this result to compute the joint p.d.f. of the j^{th} and k^{th} order statistics, $j < k$, from a random sample $X_1, ..., X_n$ from some distribution with continuous c.d.f. F.

4. (20 points) Consider the normal error simple linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

for $i \in \{1, 2, \ldots, n\}$ and ϵ_i are independently identically distributed as $N(0, \sigma^2)$ distribution.

(a) Show that the corresponding mean residual sum of squares (EMS) has the expectation $E(EMS) = \sigma^2$.

(b) Show that the corresponding mean regression sum of squares (RMS) has the expectation:

$$E(RMS) = \sigma^2 + \beta_1^2 \sum_{i=1}^{n} (X_1 - \bar{X})^2$$

(c) Prove that $F = RMS/EMS \sim F_{1,n-2}$ under $H_0 : \beta_1 = 0$.

5. (20 points) Consider the quadratic form $Y'AY$, where A is a $n \times n$ symmetric and idempotent square matrix and Y is a $n \times 1$ vector following normal distribution with zero mean and variance-covariance matrix $\sigma^2 I$, show that $Y'AY/\sigma^2$ follows a Chi-square distribution with degree of freedom equal to the rank of A.

6. (20 points) Consider the multiple linear regression model

$$Y = X\beta + \epsilon \text{ where } \epsilon \sim N(0, \sigma^2 I)$$

Assume that σ^2 is unknown. The unknown vector β has dimension $p \times 1$ with $p = k + 1$, and k is the number of covariates in the model. Consider following partitioning:

$$X = (X_1, X_2), \beta = (\beta_1', \beta_2')'$$

such that

$$X\beta = X_1\beta_1 + X_2\beta_2.$$

Consider the following models:

$$\text{Model 1} : Y = X_1\beta_1 + X_2\beta_2 + \epsilon$$

and

$$\text{Model 2} : Y = X_1\beta_1 + \epsilon.$$

Show that $\text{RegSS}_1 \geq \text{RegSS}_2$, where RegSS_i are the regression sum of squares for model i for $i \in \{1, 2\}$.