Ph.D. Qualifying Exam A

Probability Distributions and Regression Analysis

January 10, 2023

1. Let $Y_1, \ldots, Y_5$ be a random sample of size 5 from a standard normal distribution and $\bar{Y} = \sum_{i=1}^{5} Y_i/5$. Let Y_6 be another observation independent of $(Y_1, \ldots, Y_5)$ also from the standard normal distribution. Prove that $a\bar{Y} + bY_6$ is independent of the sample variance, S^2 based on $Y_1, \ldots, Y_5$, where a and b are any two fixed real numbers.

2. Let X_n and Y_n be random variables such that X_n and Y_n are independent for each $n \geq 1$ and their moment generating functions (m.g.f.'s) exist on the real line. If X_n converges to X in distribution and Y_n converges to Y in distribution, where X and Y are random variables, derive the limiting distribution of $aX_n + bY_n$, where a and b are any two fixed real numbers.

3. Let $(X_1, X_2, \ldots, X_{k-1})$ have a multinomial distribution with parameters $(n, p_1, p_2, \ldots, p_{k-1})$, where $0 < p_i < 1, i = 1, \ldots, k$, $p_1 + p_2 + \ldots + p_k = 1$, and $X_k = n - (X_1 + X_2 + \ldots + X_{k-1})$. (a) Find the m.g.f. of $(X_2, \ldots, X_{k-2})$. (b) What is the joint probability mass function (p.m.f.) of $(X_2, X_3, \ldots, X_{k-2})$? (c) Determine the conditional joint p.m.f. of (X_1, X_{k-1}) given that $X_2 = x_2, \ldots, X_{k-2} = x_{k-2}$.

4. Consider the normal error linear regression model with two explanatory variables:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i$$

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:

$$E(RMS) = \sigma^2 + \frac{1}{2}\Big[\beta_1^2\sum_{i=1}^{n}(X_{i1}-\bar{X}_1)^2 + \beta_2^2\sum_{i=1}^{n}(X_{i2}-\bar{X}_2)^2$$

$$+2\beta_1\beta_2\sum_{i=1}^{n}(X_{i1}-\bar{X}_1)(X_{i2}-\bar{X}_2)\Big].$$

(c) Prove that $F = RMS/EMS \sim F_{2,n-3}$ under $H_0 : \beta_1 = \beta_2 = 0$.

5. Consider the linear model $Y = X\beta + \epsilon$, where ϵ follows a multivariate normal distribution with 0 mean and diagonal variance-covariance matrix $\Sigma = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$. The vector of weighted-least-squares (WLS) estimates $\hat{\beta}_w$ can be written as:

$$\hat{\beta}_w = (X'WX)^{-1}X'WY,$$

where Y is the vector of observed values and X is the usual model matrix (including a column of 1's) and W is a diagonal matrix with diagonal elements $w_1, w_2, \ldots, w_n$.

(a) Suppose there is only one predictor variable X, derive the explicit formulas for $\hat{\beta}_{w0}$ and $\hat{\beta}_{w1}$.

(b) If $w_i = 1/\sigma_i^2$ and σ_i^2 are proportional to X_i. Simplify your expressions for $\hat{\beta}_{w0}$ and $\hat{\beta}_{w1}$ as much as you can.

6. 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 partionings:

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

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 $R_2^2 \leq R_1^2$, where R^2 is the coefficient of determination.