**Ph.D. Qualifying Exam A**

**Probability Distributions and Regression Analysis**

**January 13, 2021**

1. Let $\{\mathbf{X_n}\}$ be a sequence of independent identically distributed random vectors with common mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$, which is positive definite. Assume that the common moment generating function $M(\mathbf{t})$ of $\mathbf{X_n}$ exists in an open neighborhood of $\mathbf{0}$. State and prove the Multivariate Central Limit Theorem.

2. Let $T$ be Geometric with parameter $p_1$ in $(0, 1)$ and $U$ be an independent Geometric with parameter $p_2$ in $(0, 1)$. (a) Compute $P(T < U)$. (b) Derive the conditional probability mass function of $T$ given the event $\{T < U\}$.

3. Let $Y_1, ..., Y_n$ be a random sample of size $n = 6$ from a normal population with mean 0 and variance 1 and let $\bar{Y}$ denote the sample mean based on this sample of size $n = 6$. Let $Y_7$ be another independent observation from the same population. (a) Prove the independence of $\bar{Y}$ and $\sum_{i=1}^{6}(Y_i - \bar{Y})^2$. (b) What is the distribution of

$$\frac{6\bar{Y}}{\sqrt{\sum_{i=1}^{6}(Y_i - \bar{Y})^2 + Y_7^2}},$$

and why?

4. For a normal error simple linear regression model:

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

where $\epsilon_i$ are i.i.d. $N(0, \sigma^2)$ random errors for $i = 1, 2, \ldots, n$. Prove the following fact that the ANOVA F test and the two-sided t test for $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$ are equivalent.

5. For a normal error general linear regression model,

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{ip-1} + \epsilon_i,$$

where $\epsilon_i$ are i.i.d. $N(0, \sigma^2)$ random errors for $i = 1, 2, \ldots, n$, show that the matrices associated with quadratic forms of the regression sum of squares and that of the residual sum of squares are both symmetric and idempotent.

6. Consider the linear model with vector of observations $Y' = (Y_1, \ldots, Y_n)$,

$$E(Y_i) = \beta_0 + x_{i1}\beta_1 + \ldots + x_{ip-1}\beta_{p-1}, i = 1, \ldots, n,$$

satisfying the assumptions of normality, independence and homoscedasticity, with design matrix $X$ of dimension $n \times p$ and rank $p, p < n$. Consider $r_i = e_i / \sqrt{SSE(1 - h_{ii})/(n-p)}$, where $e_i$ is the ith residual, $h_{ii}$ is the $i^{th}$ diagonal element of the hat matrix $H = X(X^T X)^{-1} X^T$ and SSE is the sum of squared errors. Here, '$T$' denotes the transpose of a matrix. Show that $r_i^2$ has $(n-p)F_{1,n-p-1}/[(n-p-1)+F_{1,n-p-1}]$ distribution. Use the fact that $SSE - \{e_i^2/(1-h_{ii})\}$ is independent of $e_i^2$.