Notice: Do four of the following five problems. Place an X on the line opposite the number of the problem that you are NOT submitting for grading. Please do not write your name anywhere on this exam. You will be identified only by your student number, given below and on each page submitted for grading. Show all relevant work.

1. ____
2. ____
3. ____
4. ____
5. ____
Total ____

Student Number ________________________________

1. A piece of equipment functions according to the following probabilities. When checked at the end of a day’s operation, the equipment will be found to be “UP” with probability $p = 0.7$ or found to be “DOWN” with probability $q = 0.3$. In addition, the UP or DOWN status of the equipment at the end of the day is independent of its performance on other days. If the equipment is DOWN following an UP day, it is repaired and put back into service. If it is DOWN two days in a row, it is discarded. Let the random variable, $X$, denote the number of days until the equipment is discarded. For example, one of the several ways that $X$ can equal 7 is $(duuddd)$. Another way is $(uduuddd)$.

   (a) Let $Y$ be the number of days until the first “DOWN” day. Include the “DOWN” day in the value of $Y$. What is $E(Y)$?
   (b) Find $E(X)$. Hint: Condition on $Y$.
   (c) Find $P(X = 5)$ and $P(X = 6)$. (Aside: Note the difficulty in trying to extrapolate the probability mass function in order to find the expected value of $X$. The utility of conditioning is illustrated in this problem.)

2. For a random sample $X_1, X_2, \ldots, X_n$ from a continuous distribution with density function $f(x)$ and cumulative distribution function $F(x)$, let $U = \min(X_1, X_2, \ldots, X_n)$ and let $V = \max(X_1, X_2, \ldots, X_n)$.

   (a) Find the density function for $U$.
   (b) Find the density function for $V$.
   (c) Find the joint probability density function for $U$ and $V$. State where it is 0 and where it is what you claim it is.
   (d) Let $R = V - U$. Find the density for $R$. State where it is 0 and where it is what you claim it is. Your final answer will be in terms of an integral. In general, the density for $R$ can not be simplified any more than that.
3. Suppose $Y_1, Y_2, \ldots, Y_n$ are independent, identically distributed random variables with the common probability mass function $P(Y_i = k) = p(1-p)^{k-1}$; that is, having a geometric distribution with parameter $p$. Here, $k$ can take values ranging over the positive integers.

(a) Find the maximum likelihood estimator for $p$. Show all of your calculus steps.

(b) Let $W = \sum_{j=1}^{n} Y_j$ denote the sum of the observations. Show $W$ is sufficient for the estimation of $p$. Justify your answer.

(c) Find the minimum variance unbiased estimator for $p$. Justify your answer. Hint: You might want to begin by computing $E(1/(W - 1))$. To do this, it will be helpful to know that $W$ has a negative binomial distribution,

i.e. $P(W = k) = \binom{k-1}{n-1} p^n(1-p)^{k-n}$ where $k = n, n+1, n+2, \ldots$

4. The equation

$$\sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 + \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$$

is frequently referred to as the fundamental equation of regression analysis. It can be equivalently written as $\text{SST} = \text{SSE} + \text{SSR}$, where $\text{SST}$ is the total sum of squares, $\text{SSE}$ is the residual (or error) sum of squares and $\text{SSR}$ is the sum of squares due to the regression.

(a) Verify that this equation holds in the two parameter model given by

$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ with $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ and $i = 1, \ldots, n$.

(b) Compute $E(\text{SSR})$ under the assumption that $X_1, X_2, \ldots, X_n$ are known exactly.

Hint: It might be helpful to first show that $\text{SSR} = \hat{\beta}_1^2 \sum_{i=1}^{n} (X_i - \bar{X})^2$.

(c) Assuming that you know that $\text{SSE}$ and $\text{SSR}$ are independent (they are!) describe how you would test the null hypothesis that $\beta_1 = 0$ versus the alternative hypothesis that $\beta_1 \neq 0$. Your answer should include the test statistic and a description of the rejection region.
5. Let $X_1, X_2, \ldots, X_m$ be a random sample from a population with mean $\mu_x$ and variance $\sigma_x^2$. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a population with mean $\mu_y$ and variance $\sigma_y^2$. Assume the $X$’s are independent of the $Y$’s. In addition, we use the usual notation:

\[ S_x^2 = \frac{1}{m-1} \sum_{j=1}^{m} (X_j - \bar{X})^2 \quad \text{and} \quad S_y^2 = \frac{1}{n-1} \sum_{j=1}^{n} (Y_j - \bar{Y})^2. \]

In your answers to the following questions, be sure to include values for the relevant parameters such as degrees of freedom, means, variances, etc.

(a) For large values of $m$ and $n$, what is the approximate distribution of

\[ \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{S_x^2}{m} + \frac{S_y^2}{n}}} \]

(b) If the $X$’s are normally distributed, if the $Y$’s are normally distributed, if

\[ S^2 = \frac{(m-1)S_x^2 + (n-1)S_y^2}{m+n-2} \]

and if $\sigma_x^2 = \sigma_y^2 = \sigma^2$, what is the exact distribution of $(m+n-2)S^2/\sigma^2$?

(c) In the setting of part (b), what is the exact distribution of

\[ \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{S\sqrt{\frac{1}{m} + \frac{1}{n}}} \]?

(d) In the setting of part (b) with $m$ and $n$ large and with $\mu_x = \mu_y$, what is the approximate distribution of $-2\ln(\lambda)$ where

\[ \lambda = \frac{[\sum_{j=1}^{m}(X_j - W)^2 + \sum_{j=1}^{n}(Y_j - W)^2] / (m+n)]^{(m+n)/2}}{[\sum_{j=1}^{m}X_j^2 / m + \sum_{j=1}^{n}Y_j^2 / n]^{(m+n)/2}} \]

Here $W = \frac{m\bar{X} + n\bar{Y}}{m + n}$ is the grand average of all of the $X$’s and $Y$’s.

(e) In the setting of part (b), give the exact distribution of

\[ \frac{[\bar{X} - \bar{Y} - (\mu_x - \mu_y)]^2}{S^2\left(\frac{1}{m} + \frac{1}{n}\right)} \]