RENR 680 (Advanced Biometrics) – Midterm take home

1. Which of the following functions are possible probability functions? Justify your answer (3 marks).

$$(1)\quad f(x) = \begin{cases} \frac{1}{4} & \text{for } x = 1, 2, 3, 4. \\ 0 & \text{elsewhere} \end{cases}$$

$$(2)\quad f(x) = \begin{cases} \left(\frac{1}{4}\right)^x & \text{for } x = 1, 2, 3, 4, \ldots.. \\ 0 & \text{elsewhere} \end{cases}$$

$$(3)\quad f(x) = \begin{cases} -0.1 & \text{for } x \le 1 \\ 0.4 & \text{for } 1 < x \le 10 \\ 0.7 & \text{for } 10 < x \end{cases}$$

2. Given a probability function:

$$f(x) = \begin{cases} 1/3 & \text{for } x = 0 \\ 1/3 & \text{for } x = 1 \\ 1/6 & \text{for } x = 2 \\ 1/6 & \text{for } x = 3 \\ 0 & \text{elsewhere} \end{cases}$$

find (4 marks):

(1) $E(x)$ (2) $\text{Var}(x)$

(3) $E(X^2 + 2X)$ (4) The median

3. An experiment is designed to investigate the effect of N fertilization on tree growth (2 marks).

(1) What are the null hypothesis H_0 and the alternative hypothesis H_a?

(2) By accident, the *N* fertilizer was contaminated by a Phosphate. What kind of statistical error the experiment may commit? Why?

4. In analyzing data for your thesis, you often have to check whether or not your data follow a normal distribution. What types of test you may use to check the normality of your data. Write down the *R* commands you may use to implement the test. (Note: several types of tests are available. This question only refers to the nonparametric test(s) we have learned.) (2 marks)

5. Given a random sample: $X_1, X_2, \ldots, X_n$, please (8 marks)

(1) Describe the steps of using the bootstrap method to construct a 95% confidence interval for $s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

(2) Write an R program that implements the bootstrapping.

6. A random sample of 30 lodgepole pines was categorized by size class and mountain pine beetle attack. Is there a relationship between the preference of mountain pine beetle attacks and tree size? Confirm your manually-calculated test using **R**. (8 marks)

Data:

tree.size	**attack.status**
small	attacked
small	healthy
small	healthy
small	healthy
small	healthy
small	healthy
small	healthy
small	healthy
intermediate	healthy
intermediate	healthy
intermediate	healthy
intermediate	attacked
intermediate	attacked
intermediate	healthy
intermediate	healthy
intermediate	healthy
intermediate	attacked
intermediate	healthy
large	healthy
large	attacked

large	attacked
large	attacked
large	healthy
large	attacked
large	attacked
large	healthy
large	attacked
large	attacked
large	attacked
large	healthy

7. Tree size (diameter at breast height, dbh, in cm) for 14 black spruces from an upland site (denoted by *x*) and 19 spruces from a nearby poorly-drained site (denoted by *y*) were measured. Is there any difference in dbh between the spruces from the two sites? Parametric or nonparametric test may be used to answer this question. Choose an appropriate test and justify your choice. Confirm your test using *R*. (8 Marks)

x: 11.9 21.7 10.2 17.0 11.4 29.4 18.7 23.9 8.2 11.4 10.2 23.6 13.4 17.7
y: 8.6 13.3 8.1 7.3 14.2 10.0 10.7 2.8 8.1 9.6 20.3 12.2 10.0 15.9 4.8 3.3 6.9 6.8 10.0

Hint: You may need these formulae:

$$T_1 = \frac{T - n\frac{N+1}{2}}{\sqrt{\frac{nm}{N(N-1)}\sum_{i=1}^{N} R_i^2 - \frac{nm(N+1)^2}{4(N-1)}}} \sim N(0,1)$$

$$T = \sum_{i=1}^{n} R(x_i)$$

8. Modeling traffic accidents in Edmonton. In a summer month, the number of traffic accidents on each of the 30 days was recorded as follow: 0 4 4 2 2 2 1 2 4 2 3 5 4 0 0 3 6 2 3 2 3 3 3 2 2 1 0 2 1 5. (10 Marks)

   (1) Choose a probability distribution to model the daily traffic accident in Edmonton.

   (2) Estimate the parameter of the model using the method of moment. Hint: you may need the formula: $\sum_{x=0}^{\infty} x\frac{\lambda^x}{x!} = \sum_{x=1}^{\infty} \frac{\lambda \bullet \lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda}$.

   (3) Interpret the parameter of the model.

(4) Based on the model you parameterize, what is the probability there is at least one accident in a summer day in Edmonton?

9. Consider Mendelian inheritance of color genotypes: $Aa \times aa$, where allele A is dominant (controlling color trait) and a is recessive.
Offspring are: ½ *Aa* + ½ *aa*
Color: White Red

In an experiment, we observed 16 white plants and 12 red plants. We want to know whether this result is expected from the simple Mendelian inheritance law (4marks).

(1) Write down the probability model you may use to describe the genotype distribution.
(2) Test the Mendelian inheritance law.

10. Weibull distribution is traditionally used to describe the distribution of tree size. It has the form:

$$f(x) = \beta(1-\alpha)x^{-\alpha}\exp(-\beta x^{1-\alpha}), \text{ where } x \text{ is size of tree.}$$

(1) Fit this model to the dbh distribution of western hemlock trees (data in **hl.dat**), i.e., estimate α and β of the Weibull distribution. (Either moment method or maximum likelihood method can be used.)
(2) Test if the Weibull distribution is an adequate model for the dbh distribution of the hemlock trees.