MURANG'A UNIVERSITY OF TECHNOLOGY

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF APPLIED SCIENCES

UNIVERSITY ORDINARY EXAMINATION

2017/2018 ACADEMIC YEAR

EXAMINATION FOR MASTER OF SCIENCE IN STATISTICS

AMS 606: THEORY OF NON-PARAMETRIC STATISTICS

DURATION: 3 HOURS

DATE: 13TH AUGUST, 2018

TIME: 9.00 A.M. – 12.00 NOON

Instructions to Candidates:

1. Answer **Any Four** questions.
2. Mobile phones are not allowed in the examination room.
3. You are not allowed to write on this examination question paper.

QUESTION ONE

a. Discuss the 5 properties of kernel functions (10 marks)

b. Using the orthogonal series estimator $f(x) = \sum_{j=0}^{\infty} \theta_j \psi_j(x)$, with trigonometric basis function $\psi_j(x)$ defined by:

$$\psi_j(x) = \begin{cases} 1, & j = 0 \\ \sqrt{2} \cos \pi(j+1)x, & j \text{ is odd} \\ \sqrt{2} \sin \pi jx, & j \text{ is even} \end{cases}$$

and the constants $\theta_j = \frac{1}{n} \sum_{i=1}^{n} \psi_j(x)$. Calculate $\hat{f}(2)$ for the data set $x = 1, 3, 4, 5$ assuming the smoothing parameter $m = 2$ and $j = 0, 1, 2$. (15 marks)

QUESTION TWO

a. An engineer in a car manufacturing plant experiments on new spark plugs to observe the duration until burning out. Some spark plugs cause engine vibration and are replaced before completely burning out. The following times (in days) until burning out (no asterisk) or replacement before burning out (asterisk) were observed.

17, 13, 15*, 7*, 21, 18*, 5, 18, 6*, 22, 19*, 15, 4, 11, 14*, 18, 10, 10, 8*, 17

   i. Define $n, m, t_j, d_j, c_j, \text{and } n_j$ for these data, assuming that censoring occurs just after the failures were observed (6 marks)

   ii. Calculate the Kaplan-Meier estimate of $F(t)$ (10 marks)

   iii. Using Greenwood's formula, estimate $\text{var}\left(\tilde{F}(16)\right)$ (9 marks)

QUESTION THREE

a. An investigation is carried out into the lifestyle of male accountants. A group of 10,000 accountants is selected at random on 1 January 2001. Each member of the sample group supplies detailed personal information as at 1 January 2001 including name, address, date of birth and marital status. The same information is collected as at each 1 January in the years 2002, 2003, 2004 and 2005. The investigation closes in 2005.

A PhD student wishes to use the data from this investigation for her thesis on the mortality of married men. Describe the ways in which the available data for this

investigation are censored   (8 marks)

b. Selecting appropriate priors, initial values and assuming n=30, write a WinBUGS
    code (on your booklets) to fit a quadratic spline to a regression model with an
    intercept and one independent variable   (17 marks)

QUESTION FOUR

a. Briefly discuss the following density estimation techniques
    i.   Penalized likelihood Estimators   (3 marks)
    ii. K-Nearest Neighbors   (3 marks)
    iii. Kernel Density estimation   (3 marks)
    iv. Local polynomial regression   (3 marks)

b. By defining what Bezier curves are, derive the quadratic Bernstain basis polynomials.

                                                                        (8 marks)

c. Let $\{X_i, Y_i\}_{i=1}^{n} = \{(2,6),(8,13),(4,2),(3,1),(6,5)\}$, compute the K-NN estimate $M_k(x)$ for
    $x = 4 \, and \, k = 3$ (5 marks)

QUESTION FIVE

a. Define the following terms:
    i.   Homoscedastic errors   (2 marks)
    ii. Heteroscedastic errors   (2 marks)
    iii. Type II censoring   (3 marks)
    iv. Fixed design regression   (3 marks)

b. Show that the Kaplan-Meier and the Nelson Aalen estimates are the same   (6 marks)

c. Assuming that the errors in a regression problem increase by an observed variable m_i,
    describe how you would perform and ordinary least squares estimation for parameter
    estimation.   (9 marks)