STRATHMORE UNIVERSITY
STRATHMORE INSTITUTE OF MATHEMATICAL SCIENCES
END OF SEM EXAMINATION
MASTER OF SCIENCE IN BIOMATHEMATICS
BMA 8104 STATISTICAL MODELLING WITH APPLICATION TO BIOLOGY

Date: 22nd April, 2022 **Time: 3 Hours**

Instruction: Answer Question one and any other two

Question One (20 Marks)

a. Clearly define and contrast the following terms as used in practice.

i. Bootstrap and Jackknife methods (2 marks)

ii. Clustering and Classification (2 marks)

b. Consider the following R code

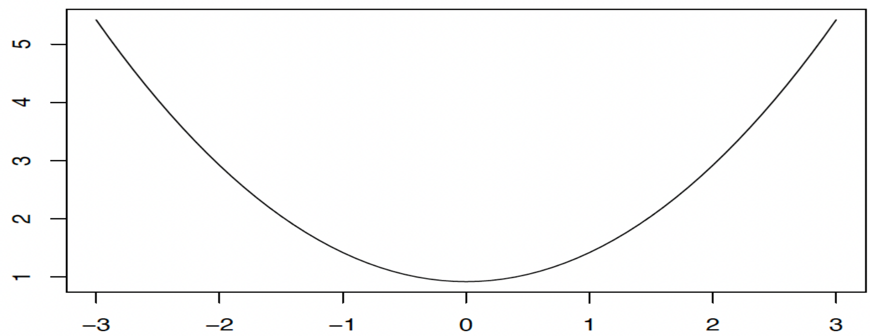
```
llik = function(x,par){
  m=par[1]
  s=par[2]
  n=length(x)
# log of the normal likelihood
# -n/2 * log(2*pi*s^2) + (-1/(2*s^2)) * sum((x-m)^2)
  ll = -(n/2)*(log(2*pi*s^2)) + (-1/(2*s^2)) * sum((x-m)^2)
# return the negative to maximize rather than minimize
  return(-ll)
}
```

i. What will this function return when x=5 and par = (3,10)? (2 Marks)

ii. Sketch the plot of the R command below: (3 Marks)

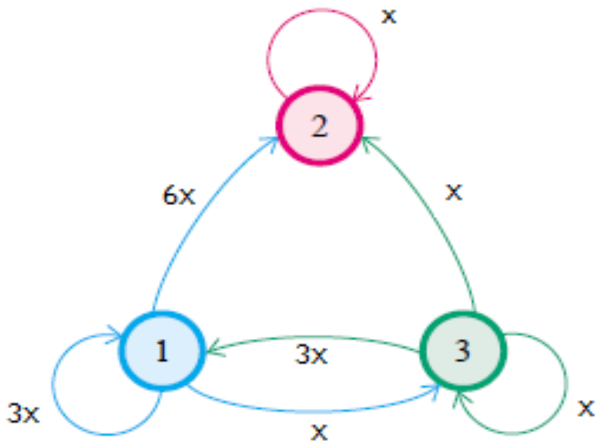
```
plot(seq(-3,3,.1),-1*sapply(seq(-3,3,.1),FUN=llik,par=c(0,1)),type='l',
     ylab='',xlab='')
```

iii. Write R code that will produce the graphic below: (3 Marks)



c. Find the maximum likelihood estimator (MLE) of θ: $X_i \sim \text{Binomial}(m,\theta)$, and we have observed $X_1, X_2, X_3, \ldots, X_n$. (5 marks)

d. Find the transition matrix from the transition diagram below (3 marks)



Question Two (20 marks)

Sociologists have long been interested in *social mobility* – the transition of individuals between social classes defined on the basis of income or occupation. Consider a society with three social classes. Each individual may belong to the lower class (state 1), the middle class (state 2), or the upper class (state 3). Thus, the social class occupied by an individual in generation t may be denoted by $s_t \in \{1,2,3\}$. suppose that *intergenerational mobility* is described by the transition matrix P

$$\begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.2 & 0.7 & 0.1 \\ 0.1 & 0.3 & 0.6 \end{pmatrix}$$

a. Determine the transition diagram from this transition matrix. (3 marks)
b. Find the transition probabilities after 3 years? (5 marks)
c. Write and run an **R code** to find the long term trend of the transition matrix. (9 marks)
d. State the Markov property and explain why it is important in MCMC methods? (3 marks)

Question THREE (20 marks)

Given a vector data that is drawn from a Poisson distribution with unknown μ.

a. Derive the Poisson likelihood for observation $y_1, y_2, \ldots, y_n$, and hence find the MLE of μ. (7 marks)

b. Write an R function to:

   i. Declare the Poisson log-likelihood function (6 marks)

   ii. Estimate the unknown Poisson parameter using the BFGS (Broyden, Fletcher, Goldfarb, and Shanno) algorithm. (3 marks)

c. Distinguish between Newton Raphson and Quasi Newton Raphson methods. (4 marks)

Question Four (20 marks)

Data of Clarke et al. (1959) reported excess of gastric ulcers in individuals with blood type O as follows: $n_A = 186$, $n_B = 38$, $n_{AB} = 36$, $n_O = 284$.

a. Write out the likelihood for these data. (7 marks)

b. What are complete data categories? (3 marks)

c. Express the complete data “counts” as a function of allele frequency estimates and the observed data. (5 marks)

d. Apply E-M algorithm to determine the genotype frequencies. (5 marks)

Question Five (20 marks)

a. Give a detailed definition of a hidden Markov model (HMM). Include in your answer a description of the assumptions made by this model. (4 marks)

b. Derive the Viterbi algorithm for computing, on the basis of some HMM, the most likely sequence of states to have produced a given sequence of observations. (6 marks)

c. An HMM has three states $\{s_1, s_2, s_3\}$ with the prior $\Pr(S_0 = s_1) = 0.3$, $\Pr(S_0 = s_2) = 0.3$, $\Pr(S_0 = s_3) = 0.4$. The transition model is

$$T = \begin{pmatrix} 0.2 & 0.3 & 0.5 \\ 0.1 & 0.4 & 0.5 \\ 0.6 & 0.1 & 0.3 \end{pmatrix}$$

where $T_{ij} = \Pr(S_{t+1} = s_j \mid S_t = s_i)$. Observations have values $\{e_1, e_2\}$, and the sensor model is $\Pr(E_t = e_1 \mid S_t = s_1) = 0.6$, $\Pr(E_t = e_1 \mid S_t = s_2) = 0.5$, $\Pr(E_t = e_1 \mid S_t = s_3) = 0.1$. You observe that at times 0, 1 and 2 the three corresponding observations are (e_1, e_2, e_1). Use the Viterbi algorithm to compute the most likely sequence (S_0, S_1, S_2) of states.

(10 marks)