M.Sc. (Data Science and Analytics) Year I
STA 8405- PROBABILITY AND STOCHASTIC PROCESSES.
Assignment III: Poisson Distribution and Hidden Markov Chains

Instruction: Submit the assignment by CoB 7th August, next week. Use RMarkdown to do your work. Answer all questions. Use R in all your computation

1. Let X be a random variable which is distributed as a mixture of two distributions with expectations μ_1, μ_2, and variances σ_1^2 and σ_2^2 , respectively, where the mixing parameters are $\delta 1$ and $\delta 2$ with $\delta 1 + \delta 2 = 1$.
    a. Show that $Var(X) = \delta 1\ \sigma_1^2 + \delta 2\ \sigma_2^2 + \delta 1 \delta 2 (\mu 1 - \mu 2)\ 2$.
    b. Show that a mixture of two Poisson distributions, $Po(\lambda 1)$ and $Po(\lambda 2)$, with $\lambda 1 \neq \lambda 2$, is overdispersed, that is $Var(X) > E(X)$.
2. Write a set of R functions that generates and executes the scripts:
   dpoismix(x,lambda,delta),
   ppoismix(q,lambda,delta),
   qpoismix(p,lambda,delta),
   rpoismix(n,lambda,delta),
   You may use any of the available R functions, such as dpois() and ppois() to construct your functions. The tricky one to do is qpoismix(p,lambda,delta). This should compute the quantile, defined as the smallest non-negative integer x which is such that $F(x) \geq p$. For experienced R users: Write qpoismix() so that it works when p is a vector. (b) Use graphics to check and illustrate your functions. In particular verify that the random samples generated using rpoismix() have the required properties.
3. Describe how to use the following R commands:
    - %*% (matrix multiplication),
    - t() (transpose a matrix),
    - solve() (solve a system of linear equations, or invert a matrix),
    - diag() (extract or replace the diagonal of a matrix, or construct a diagonal matrix),

   Then, write a R function statdist(gamma) that computes the stationary distribution, δ, of a stationary m-state Markov chain with transition probability matrix gamma.
4. Find out how to use the following R commands:
    - for() (used for looping),
    - sample() (a very useful function for drawing random samples).

    a) Then, write a R function genPoisHMM(n,gamma,lambda) that generates a series of length n from a stationary m-state Poisson HMM with transition probability matrix gamma and Poisson parameters lambda. Regard the following notes and specifications. The function should determine the number of states, m, e.g. by using m <- length(lambda).
    b) To generate the first observation, you will need to compute the stationary distribution, δ. You can use the function statdist() to do this (see Problem 2.5).
    c) Try to avoid using if() statements; rather use the function sample() in this application.

d) Test your function by generating a long sequence of observations (say n = 1000) and then check whether the sample mean, variance, histogram, etc. correspond to what you should be getting