Question 2 (*Chapter 6 Point estimation*)

(a) The following simulated data set constitutes a random sample from a Pareto distribution with parameters $K = 1000$ and unknown θ. The data set is not included in the SSC data subdirectory, and if you wish to use SSC to answer this part of the question, then you will need to key the data in to your worksheet. However, any calculator will also be suitable.

1032	1130	1286	1290	1218	1606	1252	1398	1165	1080
1974	1584	2677	2034	1567	1152	1042	1315	1181	1034
1188	1204	1007	1073	1033	1112	1052	1475	1005	1296

Find the moment estimate and the maximum likelihood estimate for θ, based on these data. Discuss which of the two estimates is to be preferred, and why. (Use the estimating formulas $\widehat{\theta}_{MM}$ and $\widehat{\theta}_{ML}$ given on page 253 of the course text.) [6]

(b) According to genetic theory, plants grown as a consequence of a particular cross may be of three types A, B or C, with corresponding probabilities

$$P(A) = \theta, \quad P(B) = 2\theta, \quad P(C) = 1 - 3\theta,$$

where $0 < \theta < 1/3$. In a particular experiment where there were 82 offspring plants, 18 were of type A, 30 were of type B and the remaining 34 were of type C.

Write down the likelihood of θ given these experimental results, and hence find the corresponding maximum likelihood estimate of θ. (You will need to use the routine `maxpt()` here.) [8]

(c) The logseries distribution is an integer-valued discrete probability distribution defined over the positive integers $1, 2, \ldots$, with probability mass function

$$p(x) = \frac{(1 - e^{-\theta})^x}{\theta x}, \quad x = 1, 2, 3, \ldots,$$

where θ is an indexing parameter, $\theta > 0$. The model has many applications, particularly in botany and biology where it is used for species counts. The mean of the distribution is $\mu = (e^{\theta} - 1)/\theta$ and increases with increasing θ.

The distribution of the number of eggs found in the nests of egg-laying species has often usefully been modelled by the logseries distribution, with the parameter θ observed to alter substantially between different species.

(i) For a species of bird the indexing parameter is known to be about $\theta = 0.4$. What proportion of nests contains exactly one egg? What proportion contains more than two eggs? (You can use the SSC functions `logss()` and `Logss()` here, if you so wish.) What is the mean nest size? [6]

(ii) Data collected on 20 nests for a species with unknown θ resulted in the sample

1	1	1	1	1	1	1	3	1	1
1	8	9	3	3	1	1	6	4	8.

Key these data in to your SSC worksheet, and calculate the sample mean. Verify that the moment estimate of θ is given by $\widehat{\theta}_{MM} = 1.7970$. (That is, check that the value of $(e^{1.7970} - 1)/1.7970$ is close to the sample mean.) [4]

(iii) Show how you could use the routine `zero()` to discover the moment estimate if you had not been told what it was. [6]

(iv) Calculate the maximum likelihood estimate of θ based on these data using the `maxpt()` routine in SSC. [4]

In fact, it is possible to show using differentiation that for the logseries distribution with parameter θ, the moment estimator of θ and the maximum likelihood estimator of θ are exactly the same thing. Any differences