Y1.1 In a particular portfolio of 1,000 life assurance policyholders, deaths are assumed to occur independently with a probability of 0.05.

(i) Calculate the median number of deaths. [4]

(ii) Calculate the probability that the number of deaths, D, lies between 45 and 59 inclusive:

(a) exactly

(b) using a Poisson approximation

(c) using a normal approximation. [16]

[Total 20]

Y1.2 (i) (a) Calculate 1,000 simulated values from a $U(0,1)$ distribution using set.seed(13).

(b) Hence, determine 1,000 simulations from the distribution which has cumulative distribution function:

$$F(x) = 1 - \frac{1}{1+x}, \quad x > 0$$ [7]

(ii) Use your simulations from part (i)(b) to:

(a) plot a labelled graph of the empirical PDF of the simulations for the range $x \in (0,200)$

(b) calculate the empirical mean, standard deviation and coefficient of skewness and comment on the shape of the distribution. [13]

[Total 20]

Y1.3 A company that makes Gizmos™ is trying to ascertain the percentage of consumers who are aware of the existence of its product. A study is to be carried out in which a random sample of the population will be interviewed and asked whether or not they are aware of it.

(i) In a sample of 20 people, 10 had heard of Gizmos™. Determine the width of an exact 95% confidence interval for the underlying population proportion. [5]

(ii) Show exhaustively for a sample of size 20 that the greatest width of an exact binomial confidence interval occurs when half of the sample have heard of Gizmos™. [8]

[Total 13]

Y1.4 An insurer is measuring the inter-arrival times between notification of consecutive claims from a portfolio of policies with a low claim rate. The insurer believes that these inter-arrival times may have an exponential distribution with unknown parameter λ. A random sample gives the following time periods (in days) between consecutive claims:

14, 4, 3, 2, 3, 1, 5, 10, 4, 23

(i) Derive a 99% confidence interval for the exponential parameter λ using a non-parametric bootstrap and set.seed(17), based on a sample of 1,000 values. [10]

After extensive analysis it is decided that the inter-arrival times have an exponential distribution with parameter 0.145.

(ii) (a) Determine 1,000 simulated means from samples of size 10 from this exponential distribution using set.seed(19).

(b) Plot a histogram of the densities of these sample means.

(c) Use the results of part (ii)(a) to calculate the empirical probability that the sample mean is less than 5. [13]

A statistician points out that if $X \sim Exp(\lambda)$, then $\overline{X} \sim Gamma(n, n\lambda)$.

(iii) Plot the PDF of the appropriate gamma distribution on the histogram of part (ii)(b) and comment. [6]

(iv) Calculate the exact probability that the sample mean is less than 5 using this result and compare to part (ii)(c). [5]

(v) (a) Determine 1,000 simulated values from the appropriate gamma distribution using set.seed(21).

(b) Plot a Q-Q plot of the sample means from part (ii)(a) and the simulations from part (v)(a) and comment on the result. [13]

[Total 47]

END OF PAPER