

Instructions

There are 3 questions on this test. The test lasts for one hour. The test is Open Book.

Write your solutions in a Microsoft WORD document. Make sure you include your name and student number in the document.

Once your solutions document is complete, save your document as a PDF in WORD by clicking on **File**→**Save As...** Once you have selected a folder in which to save your file, save it with **File name** “rtest”, and select “PDF” from the “Save as type” drop-down box.

Upload your PDF solutions file to NESS. Go to <https://ness.ncl.ac.uk>. On the MAS2602 NESS page, click **Submit** under the **Courserwork** option down the left-hand-side. Then select **R Class Test**. Enter your email address, and then in the **File to be submitted** box, browse – and upload – your saved solutions document.

Question 1

- (a) Hourly wind speed maxima at a location in the Peak District, $X \text{ ms}^{-1}$, are modelled with a Rayleigh distribution with probability density function (PDF)

$$f_X(x) = xe^{-x^2/2}, \quad x > 0.$$

Which of the following options A–E represents the cumulative distribution function (CDF) for the Rayleigh distribution?

- A. $F_X(x) = x^2 e^{-x^2/2}, \quad x > 0;$
- B. $F_X(x) = 1 - e^{-x^2/2}, \quad x > 0;$
- C. $F_X(x) = 1 - e^{-x^2}, \quad x > 0;$
- D. $F_X(x) = xe^{-x^2}, \quad x > 0;$
- E. $F_X(x) = 1 - 2x^2 e^{e^{-2x}}, \quad x > 0.$

In your solutions document, simply state the answer next to the question number - e.g. A, B, C, D or E.

- (b) An engineer wishes to simulate a series of logged hourly wind speed maxima from the Rayleigh distribution. She *could* use the inverse CDF method to do this, making use of the answer to part (a). However, she realises that a Rayleigh distribution is just a chi-squared distribution on 2 degrees of freedom. Write some R code to simulate 1000 logged wind speeds, that is 1000 realisations on $Y = \log X$.
- (i) Produce a histogram of your 1000 realisations of Y , and **include this histogram in your solutions document.**
 - (ii) Estimate the interquartile range for Y , and **include this value in your solutions document.**

Question 2

Suppose the function

$$f_X(x) = w_1 f_1(x) + w_2 f_2(x)$$

is the probability density function (PDF) of an exponential mixture distribution, where f_1 and f_2 are the PDFs for $Z_1 \sim \text{Exp}(\lambda_1)$ and $Z_2 \sim \text{Exp}(\lambda_2)$ respectively.

- (a) Write an R function called **expmix** to generate samples from this mixture distribution. The function should take as inputs the number of realisations to generate (n), as well as the marginal rate parameters (λ_1, λ_2) and the weights (w_1 and w_2); for example:

```
1 expmix = function(n, lambda1, w1, lambda2, w2){  
2   ...  
3   ... }
```

Include this function in your solutions document

- (b) Use your function in part (a) to generate 500 realisations from an exponential mixture with $\lambda_1 = 0.01$, $\lambda_2 = 2$, $w_1 = 0.95$ and $w_2 = 0.05$. Plot a histogram of your realisations from this mixture distribution, coloured **purple**, with appropriately labelled axes and a suitable title. **Include this histogram in your solutions document.**

Question 3

Continuous random variables (X, Y) are jointly distributed. Marginally, X is distributed such that $X \sim N(100, 10)$. Given $X = x$, Y also has a normal distribution with mean $\log |X|$ and the same variance as X .

- (a) Write a function in R that simulates pairs of realisations (x, y) from this joint distribution. **Include this function in your solutions document.**
- (b) Produce a scatterplot of 1000 pairs from the joint distribution of (X, Y) , making sure your axes are labelled appropriately. **Include this plot in your solutions document.**
- (c) Estimate the correlation between X and Y , and **include this estimate in your solutions document.**

The End