

Practical 1: Thursday 8th November, 4–6pm

1. Write an R function to generate samples from a 3-component normal mixture distribution with parameters

$$\begin{array}{lll} \mu_1 = 2, & \sigma_1 = 1, & w_1 = 0.3, \\ \mu_2 = 5, & \sigma_2 = 2, & w_2 = 0.4, \\ \mu_3 = 7, & \sigma_3 = 0.5, & w_3 = 0.3. \end{array}$$

- (a) Simulate a sample of size 1000 from this distribution and plot a histogram.
- (b) Compute the sample variance for your sample. Compare it to the variance of the random variable

$$T = w_1 Z_1 + w_2 Z_2 + w_3 Z_3$$

where $Z_i \sim N(\mu_i, \sigma_i^2)$ and the Z_i are independent. (Calculate the variance of T “by hand”.) Should these answers be similar?

2. Write an R function that uses the inverse CDF method to sample from the distribution with PDF

$$f_X(x) = \begin{cases} \left(\frac{2}{x}\right)^3 & \text{when } x > 2, \\ 0 & \text{otherwise.} \end{cases}$$

Plot a histogram of a sample of size 100.

3. Suppose the letters of the word AARDVARK are randomly permuted. What is the approximate probability that the permuted word contains no occurrences of the same letter next to itself i.e. no “AA” and no “RR”? (Modify the R function for the MISSISSIPPI example in lectures.)

Practical 2: Thursday 15th November, 4–6pm

4. In a hurricane-prone region, wind speed excesses (X metres per second) over some high threshold are modelled with a *Generalised Pareto Distribution*, with probability density function (PDF)

$$f_X(x; \sigma) = \sigma^{-1} e^{-x/\sigma}, \quad x \geq 0,$$

and cumulative distribution function (CDF)

$$F_X(x; \sigma) = 1 - e^{-x/\sigma}, \quad x \geq 0.$$

Hourly maximum wind speed excesses during a storm, for a 24 hour period, resulted in a sample mean of 11.6 ms^{-1} , with lower and upper quartiles of 3.4 ms^{-1} and 15.9 ms^{-1} respectively. From these wind speed excesses, the maximum likelihood estimate for σ can be shown to be $\hat{\sigma} = 12$.

- (a) Find $F_X^{-1}(x)$, and write an R function to return values from this inverse.
 - (b) Use F^{-1} to obtain a sample of size 1000 from the Generalised Pareto Distribution, and store this sample in the vector **X**.
 - (c) Produce a histogram of your sample in **X**, using the argument **probability = TRUE**, and overlay the PDF as provided at the start of this question. Make sure you give your histogram a suitable title, axis labels etc., and experiment with different colours and/or line types, as demonstrated in lectures. [Hint: You may need to experiment with the **curve** command for overlaying the PDF.]
 - (d) Produce summary statistics for your sample of realisations from this model. How does your simulated sample compare to the real life data?
 - (e) Use R to directly generate 1000 realisations from an exponential distribution with rate $1/\hat{\sigma}$, and explore this simulated series. How does this compare to your series stored in **X**?
5. Suppose $X \sim U(1, 5)$, and that we consider two transformations of X : $Y_1 = \sqrt{X}$ and $Y_2 = \log X$.
- (a) Use R to simulate 1000 values on each of Y_1 and Y_2 , and store the results in the vectors **Y1** and **Y2** respectively.
 - (b) Derive the probability density function (PDF) for each of Y_1 and Y_2 .
 - (c) On a 1×2 panel of plots in R (i.e. 1 row, 2 columns), produce a histogram of your samples from Y_1 and Y_2 . On each histogram, overlay the PDF you derived in part (b) to confirm your simulations. Make sure you give your histograms suitable titles, axis labels etc., and experiment with different colours and/or line types, as demonstrated in lectures. [Hint: You will need to use the **par(mfrow = c(1, 2))** command to partition the plotting space, and you may need to experiment with the **curve** command for overlaying your derived densities.]
6. Suppose $Z \sim N(0, 1)$, and that we consider the transformation $Y = Z^2$.
- (a) Use R to simulate 1000 values on Y , and store the results in the vector **Y**.

- (b) Produce a histogram of your sample from Y , and comment on the shape of the distribution. Make sure you use the argument `probability = TRUE` when constructing your histogram.
- (c) The *chi-squared distribution on 1 degree of freedom* has PDF

$$f_X(x) = \frac{1}{2^{1/2}\Gamma(1/2)} x^{-1/2} e^{-x/2}, \quad x \geq 0.$$

Superimpose the PDF of the chi-squared distribution onto your histogram of realised values of Y , and comment.

[Hint: the Gamma function $\Gamma(\cdot)$ can be evaluated using the built-in R command `gamma`]

- (d) Use R to find the probability $\Pr(Y > 3.84)$ by considering the proportion of your simulated values in Y that exceed 3.84.
- (e) R has built-in functions for random simulation, and computing cumulative probabilities, from the chi-squared distribution. Find out what these are so you can simulate 1000 realisations from the chi-squared distribution directly, and so you can find a more accurate answer to the probability calculation in part (d).

[Hint: recall the prefixes `r` and `p` for random generation, and obtaining cumulative probabilities, from standard models such as the Normal distribution; i.e. `rnorm`, `pnorm`. What do you *think* you might use for the chi-squared distribution?]