

The Inverse CDF Method

1. Aim: Write an R function that uses the inverse CDF method to generate a sample from the distribution with probability density function

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

- (a) Note that you are given the PDF f_X , not the CDF F_X . Thus, you should obtain the CDF by integrating, i.e. find

$$\int_2^x \left(\frac{2}{t}\right)^3 dt.$$

- (b) Find the inverse of the CDF, that is, F_X^{-1} .
- (c) Write an R function that takes a single input variable: N , the number of realisations to generate. Within this function, you should implement the inverse CDF method to generate values from X by applying your answer to F_X^{-1} in part (b) to observations generated randomly from a $U(0, 1)$ distribution.
- (d) Plot a histogram of a sample of 1000 realisations from X .

2. It is suggested that the survival times of patients (X years) receiving a new cancer drug will follow a Weibull distribution with CDF

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

- (a) Write an R function that uses the inverse CDF method to generate a sample of size N from the Weibull distribution. Your function should have *three* input variables: N , λ and κ .
- (b) Past studies from a similar drug indicate that $\lambda = 2$ and $\kappa = 1.5$. Use these values, and your function in part (a), to simulate 10,000 realisations from X , and plot a histogram of these values.
- (c) Use your simulated values in part (b) to estimate
 - (i) the mean survival time of patients receiving this new cancer drug;
 - (ii) the probability that a randomly chosen patient receiving this new drug will survive beyond five years.
- (d) Find out if there is a built-in function in R to generate samples from a Weibull distribution. If there is, use this to generate a sample from a Weibull distribution with $\lambda = 2$ and $\kappa = 1.5$, and re-produce your answers to parts (b) and (c).

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

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

and 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) Write an R function to return realisations from the Generalised Pareto Distribution.
- (b) Use your function in part (a) 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.
- (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 ?