

The Inverse CDF Method

Question 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 called `inv.cdf` 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 for F_X^{-1} in part (b) to observations generated randomly from a $U(0, 1)$ distribution.
- (d) Execute the following code to produce a histogram of 1000 realisations from X , on a density scale, and to superimpose the PDF of X :

```
1 X = inv.cdf(1000)
2 hist(X, probability = TRUE)
3 curve((2/x)^3, from = 0, to = 100, col = "red", add = TRUE)
```

Question 2

Annual maximum sea levels (in feet), observed at locations on the US Gulf coast, are assumed to follow a *Gumbel* distribution with cumulative distribution function

$$F_X(x) = \exp \left\{ -\exp \left(-\left[\frac{x - \mu}{\sigma} \right] \right) \right\}, \quad \mu \in \mathbb{R}, \quad \sigma > 0.$$

- (a) Show that the inverse of the CDF is given by

$$F_X^{-1}(x) = \mu - \sigma \log(-\log x).$$

- (b) Write an R function called `gumbel.sim` that takes three input variables: `n`, the number of realisations to generate; `mu`, the value of the Gumbel location parameter; and `sigma`, the value of the Gumbel scale parameter. Within this function, you should implement the inverse CDF method to generate values from X by applying your answer for F_X^{-1} in part (a) to observations generated randomly from a $U(0, 1)$ distribution.

- (c) From annual maximum sea levels observed at New Orleans, maximum likelihood estimates for the Gumbel parameters are obtained as $\hat{\mu} = 8.36$ and $\hat{\sigma} = 1.20$. With these estimates, use your function in part (b) to simulate 10,000 years of sea level maxima for this location. Store your realisations, and produce numerical/graphical summaries of your realisations.
- (d) Use your simulated series in part (c) to estimate the probability that sea levels at New Orleans will exceed 16 feet in any given year.

Question 3

The examples used so far, in this week's lecture and in this practical, have focused on simulating realisations from continuous probability distributions. Suppose we wanted to simulate realisations from a discrete probability distribution. The method works in the same way, except we need to account for the fact that the corresponding CDF will be a stepped function, rather than a continuous smooth curve.

- (a) As an example, suppose we wanted to simulate realisations from a random variable with the following probability mass function:

x	1	2	3	4
$\Pr(X = x)$	0.2	0.15	0.25	0.4

- (i) First, find the cumulative distribution function (CDF):

x	1	2	3	4
$\Pr(X = x)$	0.2	0.15	0.25	0.4
$\Pr(X \leq x)$	0.2	0.35	0.6	1.0

- (ii) Now, use R to generate 5 values from $U \sim U(0, 1)$ (hint: use the `runif` command). Use these to obtain 5 corresponding realisations from X :
- If $u < 0.2$, set $X = 1$;
 - If $0.2 \leq u < 0.35$, set $X = 2$;
 - If $0.35 \leq u < 0.6$, set $X = 3$;
 - If $u \geq 0.6$, set $X = 4$.

- (b) Can you write a function to simulate n realisations from X in part (a)?
- (c) For the uniform random numbers:

$$u_1 = 0.253, u_2 = 0.588, u_3 = 0.789,$$

simulate three random numbers from a Poisson distribution with rate $\lambda = 2$. Note: R is not needed to answer this question!