Practice problems–Generating Random Variates

Problem 1

Use the inverse-CDF transform to generate samples from the geometric distribution with parameter p and support on the positive integers. In other words, X is the number of Bernoulli(p) trials until a first success. You can choose p and the number of variates.

Compute the relative frequencies and compare them to the theoretical values coming from the probability mass function. Are they similar enough? (*Hint*: Use the Central Limit Theorem to justify this.)

Problem 2

Using the definition of a geometric distribution, we could also draw random variates using the following approach:

- Draw Bernoulli variates sequentially until you obtain a first success.
- If it took k trials to reach the success, then your random variate is k.
- Repeat the steps above for each new variate.

Explain why the approach above is **inferior** to the inverse-CDF approach in Problem 1.

Problem 3

Consider the following density function:

$$f(x) = \frac{3}{2}(1 - x^2), \qquad x \in (0, 1).$$

a. Prove that this is indeed a density, i.e. its integral over the support is equal to 1.
b. Find the CDF of this distribution.
c. Use the inverse-CDF transform to sample from this distribution. *Hint*: The quantile function doesn't have a closed form solution.

Problem 4

Given a uniform random variable X on $(0, 1)$ and a positive real number $\alpha > 0$, it follows that $X^{1/\alpha}$ follows a Beta($\alpha, 1$) distribution.

a. Use this relationship to sample 1000 variates from the Beta($\alpha, 1$) distribution for a value α of your choice.
b. Construct a QQ-plot to assess the validity of your implementation.

Problem 5

The $\mathsf{Pareto}(a, b)$ distribution has CDF

$$F(x) = 1 - \left(\frac{b}{x}\right)^a, \qquad x \geq b > 0, \quad a > 0.$$

Derive the quantile function, and use the inverse-CDF transform to generate a random sample from the $\mathsf{Pareto}(2, 2)$ distribution.

Problem 6

In probability theory, a **copula** is a multivariate distribution such that each marginal is uniform on $(0, 1)$. Through Sklar's theorem, we can model *any* multivariate distribution as a combination of marginal distributions and a copula.

In this exercise, we will use a Gaussian copula to combine two exponential distributions into a bivariate distribution.

a. Read the help page for the function `rmvnorm` in the package `mvtnorm`. Use this function to generate 1000 samples from a bivariate normal with means 0, variances 1, and correlation $\rho = 0.5$. Produce a scatter plot of your sample (i.e. the output will have two columns, so plot column 1 against column 2).

b. Let Φ be the CDF of a standard normal distribution (cf. `pnorm`). Then let F_1^{-1} be the quantile function of an exponential with $\lambda = 1$, and let F_2^{-1} be the quantile function of an exponential with $\lambda = 2$. Transform the bivariate normal generated in part a) using the following transformation:

$$g(x, y) = \left(F_1^{-1}(\Phi(x)), F_2^{-1}(\Phi(y))\right).$$

c. Using QQ-plots, check that after transformation, the first column follows an exponential distribution with $\lambda = 1$ and the second column follows an exponential distribution with $\lambda = 2$.

d. Compute the correlation between the two columns. Is it close to $\rho = 0.5$?

e. Repeat this simulation for different values of ρ. Create a scatter plot with ρ on the x-axis and the correlation of the transformed variables on the y-axis. Do you see any relationship?