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 $\mathsf{Beta}(\alpha, 1)$ distribution.

a. Use this relationship to sample 1000 variates from the $\mathsf{Beta}(\alpha, 1)$ 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?

Problem 7

The density of the (standard) folded normal distribution is given by

$$f(x) = \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right), \quad x > 0.$$

a. Implement the Accept-Reject algorithm to sample from the folded normal using proposals from the exponential distribution.
b. Once you have a sample from the folded normal distribution, you can transform them into a sample from the *standard normal* distribution by randomly selecting a sign (with equal probability). For example, if you sample 2.43 from a folded normal, you can transform it into either -2.43 or 2.43 by randomly choosing the sign. Using one of the methods discussed in class, show that this algorithm actually gives a random sample that matches the standard normal distribution.
c. Use the algorithm in b) to give an estimate of $Var(X)$ with a standard error approximately equal to 0.001.

Problem 8

We will generate random variates from a standard normal $N(0,1)$ using the double exponential distribution; its density is given by

$$g(x \mid \lambda) = \frac{1}{2\lambda} \exp\left(-\frac{1}{\lambda}|x|\right), \quad \lambda > 0.$$

a. Let $f(x)$ be the density of the normal distribution. Compute the ratio $f(x)/g(x \mid \lambda)$. Using calculus (or any other analytic method), find a uniform upper bound C for the ratio. (*Hint*: The upper bound C will be a function of λ, but not of x.)
b. Find the value $\hat{\lambda}$ that minimizes the upper bound C.
c. Implement the Accept-Reject algorithm for sampling from $N(0,1)$ using proposals from $g(x \mid \hat{\lambda})$. To generate samples from a double exponential distribution, you can use the function `rdoublex` from the `smoothmest` package.

Problem 9

Suppose we want to sample from a density $f(x)$ which we only know up to a constant:

$$f(x) \propto \exp\left(-\frac{1}{3}|x|^3\right), \qquad x \in \mathbb{R}.$$

In other words, we have $f(x) = M \exp\left(-\frac{1}{3}|x|^3\right)$ for an unknown constant $M > 0$.

a. Explain why we can still sample from this density using the Accept-Reject algorithm even if we don't know M.
b. Find a proposal density and implement the Accept-Reject algorithm to sample from f.
c. **Bonus** Explain how you could use this sample to estimate M.