Markov Spring 2025 Homework 2

Due Jan. 30 on Gradescope

Plotting and reporting.

- **Figures.** Include clean, labeled plots (axes, units, parameters) with consistent limits across panels. No photos of screens are accepted. Calculate key results such as empirical means, variances, and probabilities with 3–4 significant digits.
- **Code.** You don't need to include code in your written homework. Instead, create a github account if you don't already have one, create a public repository labeled "Markov2026", and upload your code there, ready to run. Label the code as "HW1_problem4.m", for example. Include the name of your account and repository in the homework. You can use any programming language you want (excel is strongly discouraged, though).
- **Math.** Accompany your calculations with clear and complete sentences. Discuss your results.

Problems

1. **Power-law sampling.** Suppose X has a power-law distribution with exponent γ and minimum value x_0,

$$f(x) = \begin{cases} Cx^{-\gamma}, & x \geq x_0, \\ 0, & x < x_0, \end{cases} \tag{1}$$

where $x_0 > 0$ and $\gamma > 3$.

   (a) Find C.

   (b) Find the cumulative distribution function of X, $F(x)$, and its inverse, $F^{-1}(x)$.

   (c) Give a two-line pseudo-code to generate a sample of X using a random variable $U \sim U(0,1)$.

   (d) Implement a code to generate samples from X for $\gamma = 4$ and $x_0 = 10$. Generate $N = 100, 1000$, and 10000 samples from X and plot histograms of the samples together with the theoretical distribution. Normalize your histograms so that they have area 1, and set the horizontal plot range to be $[0, 60]$.

2. **Gamma sampling.** Suppose $X \sim \text{Gamma}(2,1)$, i.e., it has a PDF given by

$$f(x) = \begin{cases} xe^{-x}, & x \geq 0, \\ 0, & x < 0. \end{cases} \tag{2}$$

In this problem you will implement and compare three different ways of sampling from this distribution.

   (a) In order to use Inverse Sampling, calculate the CDF of X, $F(x)$.

   (b) Since one can't explicitly solve $F(x) = u$ for x to use inverse sampling, do this numerically: generate 10^6 independent samples of a uniform random variable $U \sim U(0,1)$. For each one, solve $F(x) = u$ for x using a root finder (like Newton's method). Make sure that your solver is finding the positive roots. Report the runtime of your code using a timer like tic/toc in matlab or similar.

   (c) Use the acceptance-rejection algorithm with an exponential proposal, $g(x) = \frac{1}{2}e^{-x/2}$. State the optimal value of c and the acceptance rate. Generate 10^6 *accepted* samples and report the runtime of your code.

   (d) Finally, recognizing that a Gamma(2,1) variable is the sum of two independent exponentials with parameter 1, generate 10^6 samples of X by generating each sample from a sum of two independent exponentials. Report the runtime of your code.

(e) Show a histogram of your samples (from either of the three methods) together with the theoretical PDF. Normalize your histogram so that it has area 1.

(f) Rank the three methods in order of fastest to slowest.

3. **Mixture Sampling → Polynomial Sampling.** Consider a PDF which is a normalized, weighted average of other PDF's (a "mixture"),

$$f(x) = \sum_{j=1}^{N} a_j f_j(x), \tag{3}$$

where $\int_0^1 f_j(x)dx = 1$ and $\sum_{j=1}^N a_j = 1$. Suppose that the CDF of X is $F(x)$. Suppose that first we sample a discrete random variable J with values $1, 2, \ldots, N$ and PMF $P(J = j) = a_j$. Then, if $J = j$, we sample Y from a random variable X_j with PDF $f_j(x)$.

(a) Show that $Y \sim X$. Hint: condition on j to show that Y and X have the same CDF.

(b) Find the CDF $F_j(x)$ for

$$f_j(x) = (j+1)x^j, \qquad 0 < x < 1$$

and write down pseudo-code to sample from this distribution.

(c) Combining parts (a) and (b), write a pseudo-code to sample from a random variable Z with PDF

$$f(z) = z + 3z^5, \qquad 0 < z < 1. \tag{4}$$

4. **Gaussian Sampling.** Based on what we have studied so far, write down a pseudo-code to simulate a standard Normal random variable (there are more efficient algorithms, but use either acceptance-rejection or inverse transform sampling for this problem).

5. **Discrete Random Variable.** Consider a random variable Z that can take the values 1, 2, and 3 with probabilities 0.01, 0.9, and 0.09, respectively.

(a) Write the cumulative thresholds and a pseudo-code to simulate Z using one uniform random variable $U \sim U(0, 1)$.

(b) If you may relabel {1, 2, 3} before sampling, which order minimizes the expected comparisons (inequalities that need to be checked)? If generating M samples using your pseudo-code, estimate the expected number of comparisons when using the original and the reordered versions (don't count an "else" as a comparison).

6. **[Required for APPM 5560/STAT 5100 Students Only.]** Let Y be a random variable with Moment Generating Function $\phi_Y(t) = (1 - 2t)^{-3}$, for $t < 1/2$. Based on this, respond:

(a) What is the expected value of Y?

(b) What is the variance of Y?

(c) It can be shown (take this for granted) that if $Z \sim Normal(0, 1)$ then Z^2 has a moment generating function $\phi_{Z^2}(t) = (1-2t)^{-1/2}$, for $t < 1/2$. Based on this explain — but do not write an algorithm — how you would simulate Y if you were able to simulate any number of i.i.d. standard normals.

END