# Assignment 1

**STA410, Winter 2026**

**Assigned on January 16, 2026.**
**Submit by 11:59 pm on January 29, 2026 on Quercus.**

This assignment covers some topics related to sampling. The goal of this assignment is to ensure that you understand the concepts covered in class and that you can integrate them together. If you get stuck, you may refer to books or the internet, or you may talk with your peers. All solutions you submit should be your own work. If you use any outside resources (including generative AI) to solve the problems, you must cite them in your solutions. For questions involving implementation, include appropriate plots, code and a written answer in your submisssion with a description of the results.

1. **(Monte Carlo) [30pts]** Random walks have found application in countless fields, ranging from ecology and economics to physics, chemistry, and biology. Here we will consider the example of polarization mode dispersion (PMD) in optical fiber communication systems. A single mode optical fiber actually supports two orthogonally polarized electromagnetic waves. Under ideal conditions—i.e., when the fiber core is perfectly circular—the two modes propagate identically. In reality, however, imperfections break the circular symmetry and cause the two modes to propagate with different speeds. In general, PMD is a form of dispersion wherein the two different polarization states of light in a waveguide travel at different speeds due to random imperfections and asymmetries of the optical fiber, causing random splitting of optical signals, which in turn induces transmission errors.

   (a) Warm-up: As a starting point, we consider a standard random walk in one dimension. The walker starts out at the origin and can take a fixed number of steps N = 100. Each step is of length 1 and is randomly chosen to point either left or right with equal probability. That is,

   $$S_N = \sum_{j=1}^{N} X_j,$$

   where $X_1, X_2, \cdots, X_N$ are i.i.d. Bernoulli random variables with $\mathbb{P}(X_j = 1) = \mathbb{P}(X_j = -1) = 0.5$ for each $j$. Write a code to simulate the N-step random walk. Make a visualization of realizations of $S_n = \sum_{j=1}^{n} X_j$ as a function of $n$ for $n = 1, \dots, N$.

   (b) Use a simple Monte Carlo method with $10^5$ trials to compute the probability $\mathbb{P}(S_N > 10)$. Compute an approximate 95% confidence interval for the desired probability.

   (c) For comparison, derive analytical expressions for the desired probabilities and verify that the confidence interval contains the true value over many replicates of your experiment.

   Next, we consider a three-dimensional Gaussian random walk:

   $$S_N = \sum_{j=1}^{N} X_j,$$

   where each $X_j$ is a three-dimensional random vector composed of independent standard normals, i.e., $X_j = (x_1, x_2, x_3)$ and $x_i \sim N(0, 1)$ for $i = 1, 2, 3$. This is a simple model for PMD in optical fibers. Suppose we are interested in the probability that the total distance traveled in $N = 100$ steps is larger than a certain value $L$: $\mathbb{P}(\|S_N\| > L)$, where $\|S\| = \sqrt{s_1^2 + s_2^2 + s_3^2}$ for $S = (s_1, s_2, s_3)$.

   (d) Write a code to simulate the N-step random walk in 3-dimensions.

(e) Use a simple Monte Carlo method with $10^5$ trials to compute the probability $\mathbb{P}(\|S_N\| > 10)$. What happens if you use Monte Carlo to compute the probability $\mathbb{P}(\|S_N\| > 50)$?

2. **(Another CDF-based Sampling Method) [10pts]** Let $p$ be a distribution with CDF $F_p \colon \mathbb{R} \to [0,1]$ and let $q$ be another distribution with CDF $F_q \colon \mathbb{R} \to [0,1]$. Define the function

$$T(x) = F_p^{-1} \circ F_q(x),$$

which is a well-defined map $T : \mathbb{R} \to \mathbb{R}$.

(a) Show that the map satisfies $T(X) \sim p$ if $X \sim q$. (*Hint: Think of the map in two steps and use the properties of the CDF and inverse CDF.*)

(b) **For STA2102 students** (Optional for STA410): A location-scale family is distributions is parameterized by a location parameter and a scale parameter (e.g., mean and variance for the normal distribution). Let $p$ and $q$ be two distributions in the same location-scale family of distributions. Derive the form of $T$ that maps samples from one distribution to another in the same family. (*Hint: If $X \sim p$ is a member of the family, $Y = aX + b$ is also a member of the family and has distribution $q(y) = \frac{1}{b}p(\frac{y-a}{b})$.*)

3. **(Rejection Sampling) [15pts]** Suppose we want to generate a random variable $X$ from the tail of a standard normal distribution, that is, a normal distribution conditioned to be greater than some constant $b$. The density in question is

$$p(x) = \frac{\exp(-x^2/2)}{\sqrt{2\pi}(1 - \Phi(b))}, \quad x \geq b$$

with $p(x) = 0$ for $x < b$ where $\Phi(x)$ is the standard normal CDF. Consider rejection sampling using the shifted exponential proposal density

$$q(x) = b\exp(-b(x - b)), \quad x \geq b.$$

(a) Define $Y$ to be an exponential random variable with mean 1 and $U$ to be a uniform random variable on $[0,1]$ independent of $Y$. Show that the rejection sampling scheme defines the accepted samples to be $X = b + Y/b$ if

$$-2\log(U) \geq \frac{Y^2}{b^2}$$

(*Hint: Note that $b + Y/b$ has density $q$.*)

(b) Show the probability of acceptance is given by

$$\frac{\sqrt{2\pi}b(1 - \Phi(b))}{\exp(-b^2/2)}.$$

What happens to this probability for large values of $b$? (*Hint: You need to evaluate $M = \max_x p(x)/q(x)$.*)

4. **(Importance Sampling) [15pts]** We will now implement an importance sampling estimator to estimate rare event probabilities for the 3-dimensional random walk experiment in Problem 1.

(a) Write an explicit formula for the importance sampling approximation of

$$I = \int \mathbb{1}(\|u\| \geq L)p(u)du,$$

in the case when $p$ is the distribution for $S_N$ for Problem 1.

(b) Use importance sampling with $10^5$ trials (repeats of the experiment) to compute the probability $\mathbb{P}(\|S_N\| > 55)$. Think intuitively about how to construct a good biasing distribution in this case!

5. **(MCMC Sampling) [30pts]** In this problem we will invent an MCMC algorithm for a simple inference problem involving Gaussians. Specifically, we will infer the correlation between two Gaussian random variables. Consider the model $(y, z) \sim \mathcal{N}(\mu, \Sigma(\theta))$, with

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \Sigma(\theta) = \begin{bmatrix} 1 & \theta \\ \theta & 1 \end{bmatrix},$$

where $\theta$ is the unknown parameter. Draw $N = 1000$ i.i.d. samples from the distribution $(y, z) \sim \mathcal{N}(\mu, \Sigma(0)) = \mathcal{N}(0, I_2)$; henceforth we refer to this as the data. You will develop a Metropolis-Hastings MCMC algorithm to find the posterior distribution of $\theta$, given the data; you already know that the data was generated using $\theta = 0$ which provides intuition as you develop the algorithm. There are multiple aspects to developing this algorithm: finding the likelihood and posterior, specifying a proposal distribution, and determining the acceptance function. In the next few parts, you will be stepped through developing each ingredient.

(a) Show that the likelihood for the data $\mathbb{P}(\{y^{(i)}, z^{(i)}\}_{i=1}^N \mid \theta)$ is given by

$$\prod_{i=1}^N \mathbb{P}(y^{(i)}, z^{(i)} \mid \theta) = \prod_{i=1}^N \frac{1}{2\pi\sqrt{1-\theta^2}} \times \exp\left\{-\frac{1}{2(1-\theta^2)}\left[(y^{(i)})^2 - 2\theta y^{(i)} z^{(i)} + (z^{(i)})^2\right]\right\}$$

(b) Consider a prior distribution for the parameter $\theta$ to be

$$\mathbb{P}(\theta) = \frac{1}{\pi(1-\theta^2)^{\frac{1}{2}}}.$$

Using Bayes theorem, write (up to normalization) a formula for the posterior distribution

$$\mathbb{P}(\theta \mid \{y^{(i)}, z^{(i)}\}_{i=1}^N).$$

(c) Consider the proposal distribution

$$\theta^{(n+1)} \sim \text{Uniform}\left(\theta^{(n)} - 0.1, \theta^{(n)} + 0.1\right).$$

This proposal distribution is symmetric with respect to $\theta^{(n)}$, meaning that there is equal probability of moving in either direction of $\theta^{(n)}$. The Metropolis-Hastings algorithms with these types of proposal distributions are often referred to as *random walk Metropolis algorithms.* Using this proposal distribution, find the acceptance probability function. Starting from $\theta^{(0)} = 0.1$ and after a burn-in time of $10^4$ samples, execute the Markov chain to generate $10^3$ samples. Keep a running sample mean and variance in the burn-in period. Plot the sample mean and variance as a function of $N$. Discuss your findings.

*Observation 1:* The running sample mean and variance are often used as a diagnosis for the convergence of the Markov chain.

*Observation 2:* Note that an on-line method for computing the running sample mean and running sample variance given by:

$$m^{(n+1)} = \frac{nm^{(n)} + \theta^{(n+1)}}{n+1}, \quad c^{(n+1)} = \frac{(n-1)c^{(n)} + (\theta^{(n+1)} - m^{(n+1)})^2}{n}.$$

(d) Repeat the previous experiment but with step-size of the proposal distribution from 0.1 to 0.4. That is, consider the proposal $\theta^{(n+1)} \sim \text{Uniform}\left(\theta^{(n)} - 0.4, \theta^{(n)} + 0.4\right)$. What do you observe about the convergence of the MCMC algorithm?