Markov Spring 2026 Homework 1

Due Jan. 22 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.
- **Math.** Accompany your calculations with clear and complete sentences. Discuss your results.

Problems

1. **Airplane seat overselling.** An airplane has 98 passenger seats. Knowing that passengers miss their flight with probability 0.02, the airline overbooks the flight by selling 100 tickets (you can assume that passengers miss their flight independently of each other).

   (a) Find a simple approximation to the probability that there are not enough seats for the passengers that show up using the Poisson distribution.

   (b) Find the exact value for the probability that there are not enough seats for the passengers that show up, and compare this value with the one found in part (a).

   (c) Find the probability that the airline didn't have to reimburse anyone, given that the airplane took off with all its seats occupied. You can use the Poisson approximation.

2. **Joint distributions.** (25 points) Random variables X, Y have the joint PDF

$$f(x,y) = \begin{cases} Cxy & 0 \le y \le x, \quad 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

   (a) (5 points) Find C.

   (b) (5 points) Are X, Y independent?

   (c) (15 points) Let $Z = Y/X$. Find the PDF of Z, $f_Z(z)$. Make sure to specify the PDF for all values of z. (Hint: To find the PDF of Z, find the CDF first and then differentiate it.)

3. **Numerical evaluation of integrals via Monte Carlo simulation**

   Consider the definite integral

$$I = \int_1^\infty \frac{1}{1+x^6} dx \tag{1}$$

   (a) Use a substitution to convert the integral to a definite integral over the interval $[0,1]$ of a function $f(u)$ whose range is in $[0,1]$.

   (b) Estimate the integral by estimating the area under the function f. To do this, sample N points uniformly in the square $[0,1]^2$ and calculate the fraction of those that are below the function f. By doing this, you will obtain an estimate $E(N)$. Plot $E(N)$ as a function of N on a plot where the horizontal axis is logarithmic. Take $N = \text{floor}(10^x)$ with $x = 1, 1.1, 1.2, \ldots, 4.9, 5$. On the same plot, show a horizontal line with the value of the integral obtained by analytical methods or using numerical (not Monte Carlo) integration. You can use any software or numerical integration package you want, but make sure you mention it in your write-up.

   Note: Calculation of integrals using Monte Carlo simulation is too inefficient compared with other numerical methods for low dimensions (like the above integral, which is one-dimensional). However, for high-dimensional integrals, Monte Carlo simulation is often the only practical way to estimate definite integrals.

4. **Jill waiting.** Jill arrives at a restaurant at 6PM and waits for her three friends Moe, Agnes, and Dorothy. Each one of her friends arrives independently at a time uniformly distributed between 6PM and 7PM.

   (a) Find the PDF of the time T at which ALL three friends have arrived. (Hint: find the CDF of T and use independence of the arrivals).

   (b) Simulate this scenario $N = 10^5$ times and plot a histogram of T superimposed with the theoretical PDF from (b).

   Note: In order for the curves to match, you need to normalize the histogram properly. In matlab, you can use "histogram(your_array,'Normalization','PDF')".

5. **Moment generating function 1.** Suppose $X_1, X_2, X_3, \ldots$ are independent, identically distributed random variables with Moment Generating Function $\Phi_X(s)$, and that $N \sim \text{Poisson}(\lambda)$. Let

$$Y = \sum_{i=1}^{N} X_i.$$

   (a) Find $\Phi_Y(s)$ in terms of $\Phi_X(s)$ and λ.

   (b) Use $\Phi_Y(s)$ to find the mean and variance of Y.

6. **Finding PDFs.** Suppose that $X \sim \text{Exponential}(\lambda_1)$ and $Y \sim \text{Exponential}(\lambda_2)$, and X and Y are independent. Find the PDF of $Z = Y/X$ (you can find the CDF of Z first and then differentiate it).

7. **Central Limit Theorem - Convergence.** Let X_i be independent Poisson random variables with mean $\lambda = 1$ for $i = 1, 2, 3, \ldots$. The Central Limit Theorem implies that a properly normalized sum of n of these random variables converges to a standard Gaussian as $n \to \infty$. First, define the normalized variable

$$Y_n = \frac{\sum_{i=1}^{n} X_i - n}{\sqrt{n}}.$$

   (a) Show that the mean and variance of Y_n are 0 and 1, respectively, independently of n.

   (b) Calculate the MGF of Y_n, $\Phi_{Y_n}(s)$.

   (c) Prove that $\Phi_{Y_n}(s) \to e^{s^2/2}$ as $n \to \infty$ and identify the limit distribution. Hint: expand $\ln(\Phi_{Y_n}(s))$ in powers of $s/\sqrt{n}$ and compare with $\ln(e^{s^2/2})$.

   (d) The "Skewness" of a random variable Z with mean μ and standard deviation σ is defined as $E[(Z-\mu)^3]/\sigma$, and it measures the asymmetry of a distribution about its mean. Generally, positive skewness means the tail is longer to the right of the mean, and vice-versa. Expand $\Phi_{Y_n}(s)$ in series and use this to find $E(Y_n)$, $E(Y_n^2)$, and $E(Y_n^3)$ as a function of n. Calculate the skewness as a function of n, $S(n)$. Discuss and interpret your result in the context of the central limit theorem.

   You are not only allowed, but encouraged, to use symbolic math software like Mathematica to do symbolic calculations. For example, in Mathematica, you can use the 'Series[f[x],{x,0,3}]' command to expand a function $f(x)$ in series around $x = 0$ up to order 3.

   (e) **(For 5560/5100 students only.)** For a given n, generate 10^6 samples of Y_n and estimate $S(n)$ from this sample. Plot the numerical and theoretical values of $S(n)$ on the same log-log plot for $n = 2^k$, $k = 2, 3, 4, 5, 6, 7, 8, 9, 10$.

   Hint: the 10^6 samples can be generated in parallel. If you are using matlab, you can replace the "for" in your for loop with "parfor" and it will automatically recruit more processors to spread the work.

8. **(For 5560/5100 students only.)** **Uniqueness of MGF.** Let X and Y be discrete random variables with values in $0, 1, 2, \ldots, N$. Show that if $\phi_X(s) = \phi_Y(s)$ for all $s \in \mathbb{R}$, then $X \sim Y$.

END