APMA 3100 Probability

Project 3: Drone Delivery

This project may be done individually or in groups of two or three. Include group partners' names on your submission. You may discuss the project with your instructor, TAs, and classmates, but all work and coding must be done by you and your partner alone. The Honor Pledge should be written and signed by both members of the group.

The Problem

"Drones are set for a surge" (*The Wall Street Journal*, 19 March 2018). According to the Federal Aviation Administration (FAA) report, about 110,000 commercial drones operated in the U.S. continental airspace in 2017, but 450,000 are anticipated in 2022. The number of commercial drone pilots is expected to rise from about 70,000 to over 300,000 over the same period.

In this boom, imagine a venture that will replace the "newsboy" with the "newsdrone". Instead of walking, bicycling, or driving each morning from house to house to deliver the newspaper, a boy or girl will get a commercial-drone pilot license and will fly the drone delivering the newspapers. Given the latitude and longitude of the target drop point beside each house, the drone will fly at a specified altitude (which may be regulated by the FAA) and at a pre-calculated position and velocity will release the newspaper. The newspaper may miss the target point due to various random factors: fluctuation of flight velocity and release angle, variability of wind speed and direction, collisions with falling leaves, and so on. Because the drop error should not be frequently and excessively large, it must be controlled, at least partially. For this purpose a probability model is needed.

To model the distance X between the actual drop point, A, and the target point, T, the Cartesian coordinates are set with the origin at point T and the axes oriented East and North. When point A has the coordinates (Y_1, Y_2),

$$X = \sqrt{Y_1^2 + Y_2^2}$$

To model the randomness of point A, its coordinates Y_1 and Y_2 are assumed to be independent and identical Gaussian random variables, each having mean 0 and variance τ^2. (Experimental flights yielded $\tau = 57$ inches.) Then X has the ***Rayleigh*** distribution with the scale parameter $a = 1/\tau$, with probability density function (PDF)

$$f_X(x) = a^2 x e^{-\frac{1}{2}a^2x^2}, \quad x > 0$$

and cumulative distribution function (CDF)

$$F_X(x) = 1 - e^{-\frac{1}{2}a^2x^2}, \quad x > 0$$

and moments

$$\mu_X = \mathrm{E}[X] = \frac{1}{a}\sqrt{\frac{\pi}{2}} \qquad\qquad \sigma_X^2 = Var[X] = \frac{4-\pi}{2a^2}$$

What You Should Do:

Submit a document that contains the following three sections:

1. Model Analysis
    a. Graph f_X.
    b. Graph F_X.
    c. On one graph, create three circles centered at the origin (the origin will represent the target point T) that represent the regions that have a probability of 0.5, 0.7, and 0.9, respectively, of containing the actual drop point.

2. Law of Large Numbers
   Recall that the Weak Law of Large Numbers applied to the sample mean states that $P[|M_n - \mu_X| < c] \to 1$ as $n \to \infty$.

    a. Write simulation code that generates independent simulations of X.
        i. Re-use your linear congruential random number generator from the previous project, but change the parameter values to:

starting value (seed)	$x_0 = 1000$
multiplier	$a = 24\,693$
increment	$c = 3967$
modulus	$K = 2^{18}$

        These values yield a cycle of length $K = 2^{18}$ random numbers. The first three random numbers (rounded) are 0.2115, 0.4113, and 0.8275. Give the 51st, 52nd, and 53rd numbers in your report.

        ii. Write the code to simulate one value of X, but also write the code that simulates doing a sample of size n and calculating the sample mean M_n.

        iii. Include a copy of your code in your document.

    b. Write additional code that runs your simulation in part (a) above, and generates 110 independent values for M_n, for each of the following sample sizes:
    $n =$ 10, 30, 50, 100, 250, 500, 1000

    c. Create a scatterplot of all of your $110 \times 7 = 770$ values, using n as your horizontal axis and M_n as your vertical axis. Also, superimpose a horizontal line indicating the true mean μ_X.

    d. Analyze your scatterplot, commenting on how the Law of Large Numbers is being demonstrated.

    e. Suppose that the drone operator wants to perform a real-world experiment in which the drone will drop the morning newspaper n^* times, and the operator will then calculate the sample mean of the distance X. What value for n^* would you recommend they use?

The requirements are that n^* should be as small as possible (to minimize the cost of the experiment), but n^* should be large enough that it is essentially a proxy for ∞ in the Law of Large Numbers with $c = 10$ inches.

f. Estimate the actual probability $P[|M_{n^*} - \mu_X| < 10]$.

3. Central Limit Theorem

Define the *standardized sample mean*: $Z_n = \frac{M_n - \mu_X}{\sigma_X/\sqrt{n}}$.

Let F_n be the CDF of Z_n.

The Central Limit Theorem states that $F_n(z) \to \Phi(z)$ as $n \to \infty$.

In the following, we will pretend that we do not know the true underlying μ_X and σ_X. In a real physical experiment, the mean and standard deviation would not be known.

a. Write code that uses your simulation and generates $K = 550$ independent values for M_n, for each of the following sample sizes:

$$n = 5, 10, 15, 30$$

Denote these values as $m_5(1), m_5(2), \ldots, m_5(550)$, $m_{10}(1), m_{10}(2), \ldots, m_{10}(550)$, $m_{15}(1), m_{15}(2), \ldots, m_{15}(550)$, and $m_{30}(1), m_{30}(2), \ldots, m_{30}(550)$.

b. <u>For each of the four sample sizes n</u>, do the following:

i. Estimate the mean and variance of M_n by computing

$$\hat{\mu}_n = \frac{1}{K}\sum_{k=1}^{K} m_n(k) \qquad \hat{\sigma}_n^2 = \frac{1}{K}\sum_{k=1}^{K}\left(m_n(k)\right)^2 - \hat{\mu}_n^2$$

(These estimates will replace the true mean $\mu_n = \mu_X$ and standard deviation $\sigma_n = \sigma_X/\sqrt{n}$, which we are pretending we do not know.)

ii. Transform your sample mean values $\{m_n(k)\}$ into standardized sample mean values $\{z_n(k)\}$, where $z_n(k) = \frac{m_n(k) - \hat{\mu}_n}{\hat{\sigma}_n}$.

iii. Estimate from your $\{z_n(k)\}$ values the probabilities of the seven events

$$\hat{F}_n(z_j) = P[Z_n \le z_j]$$

where the $z_j's$ are $\{-1.4, -1.0, -0.5, 0, +0.5, +1.0, +1.4\}$.

iv. Evaluate how close your empirical CDF $\hat{F}_n$ is to the standard normal CDF Φ by computing the *maximum absolute difference*

$$MAD_n = \max_{1 \le j \le 7}\left|\hat{F}_n(z_j) - \Phi(z_j)\right|$$

v. Create a single graph that includes (a) the seven points $\left\{\left(z_j, \hat{F}_n(z_j)\right)\right\}$ and (b) the continuous graph of $\Phi(z)$. Then, highlight MAD_n as a vertical interval at the point where it occurs.

c. Note: You do not need to show all the calculations involved in b (i)-(v) in that order. For your report, just summarize it by including: (i) a table comparing the estimates $(\hat{\mu}_n, \hat{\sigma}_n)$ with the population values $(\mu_X, \sigma_X/\sqrt{n})$ for every n; (ii) a panel with the four graphs comparing the two CDFs; and (iii) a table reporting MAD_n for every n.

d. Analyze and comment on the behavior of the CDFs in the figure, judge the convergence $\hat{F}_n \rightarrow \Phi$ and $MAD_n \rightarrow 0$, and assess the extent to which this small simulation experiment has demonstrated to you the meaning of the Central Limit Theorem.