APMA 3100
Spring 2022

Project 3

Due: Friday, April 29

1. OBJECTIVE

The limit theorems of probability — ***Laws of Large Numbers*** (Weak and Strong) and ***Central Limit Theorem*** — are the backbones of many probability models and statistical procedures. Essentially, these theorems extend the concepts of a sequence and of a limit from the domain of real numbers to the domain of random variables.

Our grand objective is to observe first-hand, through *Monte-Carlo simulation* experiments, the various convergence properties of the sequences of random variables. The background material for Monte-Carlo simulation is contained in Sections 1–4 of Project 2. The mathematical model for the experiments comes from a control problem like this one.

2. CONTROL PROBLEM

2.1 Newsdrone Venture

"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.

2.2 Model of Drop Error

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),

$$X = \sqrt{Y_1^2 + Y_2^2}. \tag{1}$$

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 $a = 1/\tau$, the probability density function (PDF)

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

the cumulative distribution function (CDF)

$$F_X\,(x) = 1 - e^{-\frac{1}{2}a^2 x^2}, \quad x > 0; \tag{3}$$

and the moments

$$\mu_X = E[X] = \frac{1}{a}\sqrt{\frac{\pi}{2}}, \tag{4}$$

$$\sigma_X^2 = Var[X] = \frac{4-\pi}{2\,a^2}\ . \tag{5}$$

3. MODEL ANALYSIS

OBJECTIVE 1: Gain an understanding of the drop error model. Towards this end, perform these tasks.

1. **Graph** function f_X.

2. **Graph** function F_X.

3. **Graph** three circles, each centered at point T and having radius x_p such that $P[X \leq x_p] = p$, where $p = 0.5, 0.7, 0.9$.

4. **Explain** the meaning of the circles to a homeowner who subscribes to the newspaper that the drone will deliver. (Imagine your explanation will be included in a flyer advertising the "newsdrone" service.)

4. LAWS OF LARGE NUMBERS

4.1 Background

Let X be a random variable with finite mean μ_X. Consider the problem of estimating μ_X. Let: $n -$ sample size, $M_n -$ *estimator* of μ_X, $m_n -$ *estimate* of μ_X. (M_n is a random variable, m_n is a realization of $M_n -$ a number.)

OBJECTIVE 2: Demonstrate empirically the convergence of the sample mean M_n to the population mean μ_X (in short, $M_n \to \mu_X$), when the sample size n increases without bound (in short, $n \to \infty$).

Weak Law of Large Numbers: $M_n \to \mu_X$ *in probability*; i.e., for any constant $c > 0$,

$$\lim_{n\to\infty} P[|M_n - \mu_X| < c] = 1.$$

It reads: the *sequence of probabilities* [of the event: the distance between the sample mean M_n and the population mean μ_X is smaller than an arbitrary positive constant c] *converges to one*, as $n \to \infty$.

Strong Law of Large Numbers: $M_n \to \mu_X$ *with probability one*; i.e.,

$$P\left[\lim_{n\to\infty} M_n = \mu_X\right] = 1.$$

It reads: *probability* [of the sequence of sample means M_n converging to the population mean μ_X, as $n \to \infty$] *equals one*. The statement popular among probabilists is: "M_n converges to μ_X almost surely". The adverb "almost" concedes that $n = \infty$ can never be experienced.

4.2 Experiment

1. **Design a Monte-Carlo simulation algorithm** which generates independent realizations of X. Within this algorithm, re-use the linear congruential random number generator from your Project 2, but change the parameter values to:

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

They yield the cycle of length $K = 2^{18}$. The first three random numbers are $0.2115, 0.4113, 0.8275$. Show numbers u_{51}, u_{52}, u_{53} in your report. The way of implementing the algorithm on a computer is up to you. [Report the computer language used.]

2. **Simulate** the outcomes of many deliveries of the newspaper via a drone by generating independent realizations of X, as many as needed for Step 3.

3. **Calculate** 110 independent estimates m_n of the sample mean M_n for each sample size n:

$$n = 10, 30, 50, 100, 250, 500, 1000.$$

4. **Make a graph**, with n as the abscissa and m_n as the ordinate, showing all $110 \times 7 = 770$ estimates; include a horizontal line having μ_X as its ordinate.

5. **Interpret** the behavior of the graph using language of the Laws of Large Numbers, as appropriate. Which Law does this graph demonstrate?

6. **Recommend** the sample size n^* to be used in a real-life experiment wherein a drone will drop the morning newspaper n^* times, and then the operator will calculate the sample mean of the distance X. The operator has two requirements: (i) n^* should be as small as possible (to minimize the cost of the experiment), and (ii) n^* should be a proxy for ∞ in the Weak Law of Large Numbers with $c = 10$ inches. Justify your recommendation.

7. **Provide** an estimate of probability p in the statement

$$P[|M_{n^*} - \mu_X| < c] \approx p.$$

This statement reflects the fact that $n^* < \infty$ and that, therefore, the Weak Law of Large Numbers holds only "approximately", with some $p < 1$. *Hint*: Consider not only c and n, but also the number of realizations m_n you have.

5. CENTRAL LIMIT THEOREM

5.1 Background

Let X be a random variable with finite mean μ_X and finite variance σ_X^2. Let: $\{X_1, \dots, X_n\}$ — random sample of X, M_n − sample mean. Define the *standardized sample mean*:

$$Z_n = \frac{M_n - \mu_X}{\sigma_X/\sqrt{n}} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu_X}{\sigma_X/\sqrt{n}} = \frac{\sum_{i=1}^{n} X_i - n\,\mu_X}{\sqrt{n}\,\sigma_X}.$$

Let: F_n − CDF of random variable Z_n (which is unknown yet), Φ − CDF of the standard normal random variable Z.

Central Limit Theorem: $Z_n \to Z$ *in distribution*; i.e., for every $z \in (-\infty, \infty)$,

$$\lim_{n\to\infty} F_n(z) = \Phi(z).$$

It reads: the sequence of CDFs of the standardized sample mean M_n (or equivalently, of the standardized sum $X_1 + \dots + X_n$) converges to the standard normal CDF, as n increases without bound, regardless of the form of the CDF of X (a Rayleigh CDF in the newsdrone problem).

OBJECTIVE 3: Demonstrate empirically the convergence $Z_n \to Z$ in distribution. To wit, when an empirical CDF $\hat{F}_n$ of Z_n is constructed for each sample size n, and the sequence of functions $\hat{F}_n$ is examined, we hope to notice that the distance between $\hat{F}_n$ and Φ decreases, as n increases.

5.2 Experiment

1. **Prepare samples** for the analyses as follows. For each sample size n,

$$n = 3, 9, 27, 81,$$

calculate $K = 550$ independent estimates m_n of the sample mean M_n. With k serving as the index of these estimates, form a sample of M_n:

$$\{m_n(k) : k = 1, \ldots, K\}.$$

2. **Perform analyses**: for each n, execute Steps 2.1–2.5.

   2.1 **Calculate** the estimates of the mean and the variance of M_n:

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

   2.2 **Transform** a given sample of M_n into a sample of the standardized random variable Z_n:

$$\{z_n(k) : k = 1, \ldots, K\},$$

   where

$$z_n(k) = \frac{m_n(k) - \hat{\mu}_n}{\hat{\sigma}_n}.$$

   [In this standardization, the estimates $\hat{\mu}_n$ and $\hat{\sigma}_n$ replace the population mean μ_X and the population standard deviation $\sigma_X/\sqrt{n}$ because, in a physical experiment, the realizations of X would come from field measurements wherein the population mean and variance would be unknown.]

   2.3 **Estimate** from the sample of Z_n the probabilities of seven events:

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

   for $j = 1, \ldots, 7$, where the set $\{z_1, \ldots, z_7\}$ is

$$\{-1.4, -1.0, -0.5, 0, 0.5, 1.0, 1.4\}.$$

   2.4 **Evaluate** the goodness-of-fit of the standard normal CDF Φ to the empirical CDF $\hat{F}_n$ in terms of the *maximum absolute difference*:

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

2.5 **Draw** a figure showing (i) the seven points $\{(z_j,\ \hat{F}_n\ (z_j)) : j = 1, \dots, 7\}$, which constitute the *empirical* CDF of Z_n, (ii) the standard normal CDF Φ, graphed as a continuous function on the domain $(-2.5, 2.5)$, (iii) the MAD_n, as a highlighted interval of probability (ordinate) at point z_j at which it occurs (abscissa).

3. **Summarize** the results on two pages: (i) a panel with 4 figures comparing the two CDFs; (ii) a table comparing the estimates $(\hat{\mu}_n, \hat{\sigma}_n)$ with the population values $(\mu_X, \sigma_X/\sqrt{n})$ for every n, and a table reporting the absolute difference $|\hat{F}_n(z_j) - \Phi(z_j)|$ for every j and n, and MAD_n for every n.

4. **Conclude** the experiment. (i) Describe the behavior of the CDFs in the figure and of the numbers in the table. (ii) Note any patterns; judge the convergence $\hat{F}_n \to \Phi$ and $MAD_n \to 0$; then draw conclusions. (iii) Assess the extent to which this small simulation experiment has demonstrated to you the meaning of CLT.

5. **Reflect**. If you were to repeat this experiment, would you recommend changing any of its specifications?

6. REPORT

Instructions:

1. Organize the report into 3 sections, parallel to Sections 3, 4.2, 5.2 above; organize each section parallel to the steps of the analysis or experiment.
2. Explain your work, summarize the results, answer questions.
3. Do not submit computer code or raw data but archive them.
4. Draw figures professionally: to scale, with labels on axes, and captions.
5. Submit your final report through Gradescope.

Help and Honor Pledge

6. You may discuss the project with the instructors, teaching assistants, and classmates, but all work must be your own. The Honor Pledge must be printed on a cover page and signed by every member of the team.