Rodrigo Becerril Ferreyra
E E 381 Section 12
Lab 4
07 November 2020

Introduction

The purpose to this lab is to explore the Central Limit Theorem (CLT). Informally, the CLT states that the sum of random variables $X_1, X_2, \ldots, X_n$ approaches a normal (Gaussian) distribution as $n \to \infty$. That is, for

$$Z = \sum_{i=1}^{\infty} X_i,$$

$$P_Z(x) = \mathcal{N}(Z = x; \mu, \sigma; \infty) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

where $P_Z(x)$ is the probability density function (PDF) of Z, μ is the expected value of Z, and σ^2 is the variance of Z.

The notation I will use is the following: first is the distribution name (e.g. $\mathcal{N}$); next is the variable used to define the distribution (e.g. $X = x$); next are the parameters involved in the definition (e.g. μ and σ); lastly, the number of samples generated in creating the random variable (in other words, the length of the variable, or the amount of discrete values associated with the variable). This last value is necessary, because computers cannot hold an infinite amount of data.

In this lab, we used both uniform distributions and exponential distributions to test this theory. The PDF of the (continuous) uniform distribution is given by the following function:

$$\text{unif}(X = x; a, b; \infty) = \begin{cases} \frac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases}.$$

The mean is given by $\mu = (b+a)/2$ and the variance given by $\sigma^2 = (b-a)^2/12$.

The PDF of the exponential distribution is given by

$$\text{expon}(X = x; \lambda > 0; \infty) = \lambda e^{-\lambda x}$$

where $\mu = 1/\lambda$ and $\sigma^2 = 1/\lambda^2$. Note that in our lab, we used the parameter $\beta = 1/\lambda$; in this case, $\mu = \beta$ and $\sigma^2 = \beta^2$.

1 Problem 1

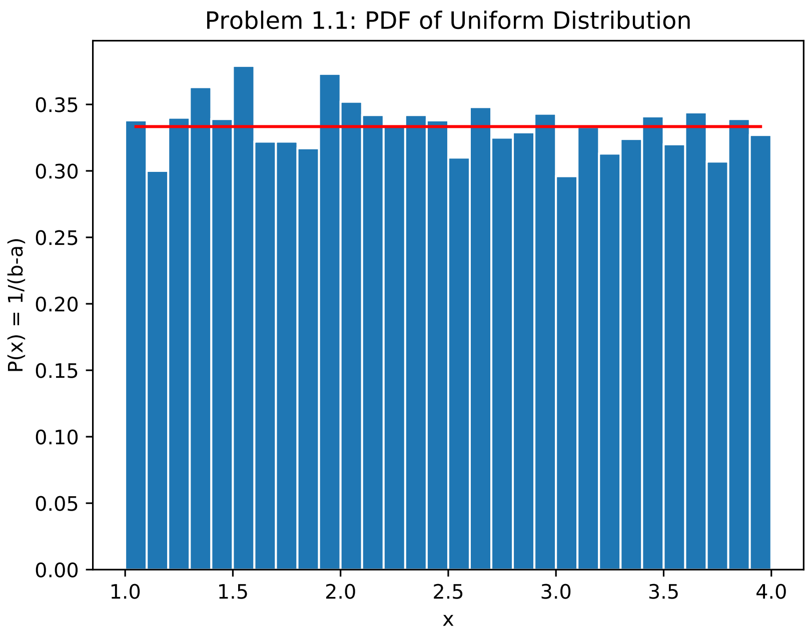
1.1 Question

In this problem, we were introduced to the `numpy` commands to generate three random variables, each representing one distribution. We also calculated their expected value and standard deviation, comparing them to the expected values.

For all PDF graphs, we generated 10 000 values for each random variable, then plotted it on a histogram. The blue bars represent the generated values, and the red line represents the theoretical PDF of that distribution. Results are posted below. Note that Problem 1 took about 0.69 s in total to finish.

1.2 Result 1

Below is the graph of the PDF of a uniform distribution, and the table of experimental and expected values. The distribution that was modeled is $\text{unif}(X = x; 1, 4; 10\,000)$.



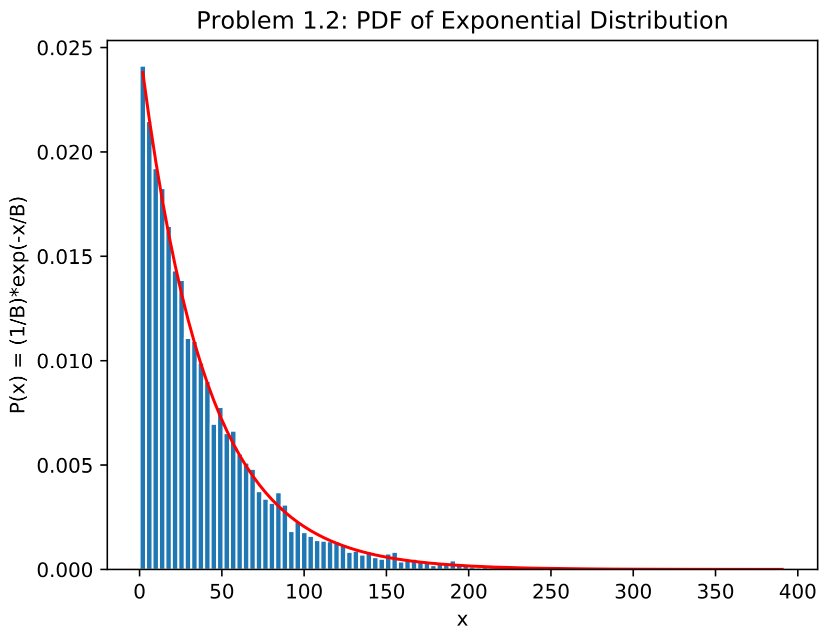
Statistics for a Uniform Distribution			
Expected Value		Standard Deviation	
Theoretical Calculation	Experimental Measurement	Theoretical Calculation	Experimental Measurement
2.5	2.506	$\sqrt{3}/2$	0.866

Figure 1: Uniform Distribution Plot and Statistics

1.3 Result 2

Below is the graph of the PDF of an exponential distribution, and the table of experimental and expected values. The distribution that was modeled is

$\text{expon}(X = x; 1/40; 10\,000)$.

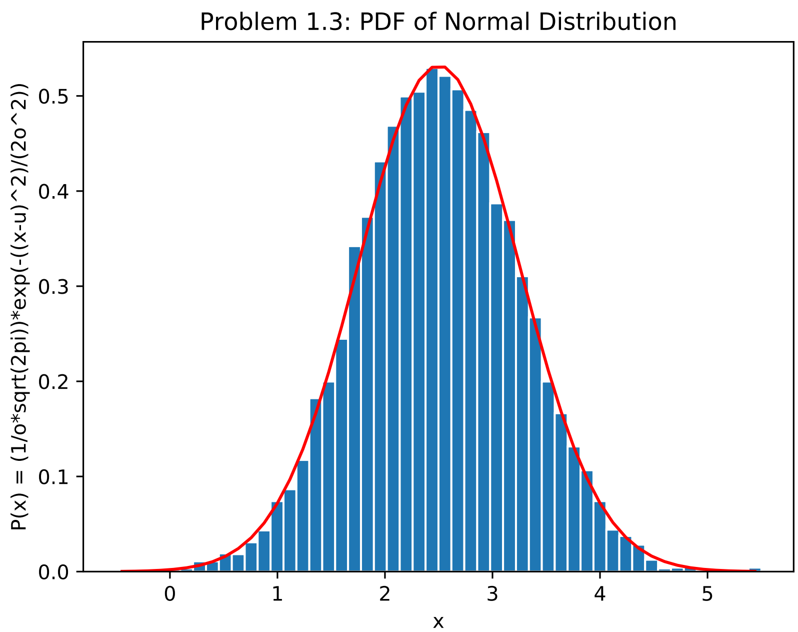


Statistics for an Exponential Distribution			
Expected Value		Standard Deviation	
Theoretical Calculation	Experimental Measurement	Theoretical Calculation	Experimental Measurement
40	39.91	40	40.00

Figure 2: exponential Distribution Plot and Statistics

1.4 Result 3

Below is the graph of the PDF of a normal distribution, and the table of experimental and expected values. The distribution that was modeled is $\mathcal{N}(X = x; 2.5, 0.75; 10\,000)$.



Statistics for a Normal Distribution			
Expected Value		Standard Deviation	
Theoretical Calculation	Experimental Measurement	Theoretical Calculation	Experimental Measurement
2.5	2.491	0.75	0.749

Figure 3: Uniform Distribution Plot and Statistics

2 Problem 2

2.1 Question

In this problem, we are tasked with setting up and witnessing the CLT in action. To do this, we were tasked with creating four random variables

$$S_n = \sum_{i=1}^{10\,000} \mathrm{unif}(x; 1, 4; n)$$

for $n \in \{1, 5, 10, 15\}$. In other words, S_n is the sum of $10\,000$ n uniformily-distributed random variables. In this example, $\mathrm{unif}(x; 1, 4; n)$ represents the thickness of n books where each book can be anywhere within the interval $[1\,\mathrm{cm}, 4\,\mathrm{cm})$. The purpose of this is to witness S_n converging to a normal distribution as n gets larger. A table of the expected values and standard deviations of S_n is given below (note that these values are calculated values; the expected values are displayed on the raw output, available at the end of this lab report). Note that Problem 2 took about 1.22 s to finish.

Number of books n	Mean thickness of a stack of n books	Standard deviation of a stack of n books
$n = 1$	$\mu_1 = 2.499$	$\sigma_1 = 0.8633$
$n = 5$	$\mu_5 = 12.53$	$\sigma_5 = 1.935$
$n = 10$	$\mu_{10} = 25.02$	$\sigma_{10} = 2.750$
$n = 15$	$\mu_{15} = 37.48$	$\sigma_{15} = 3.372$

Figure 4: Mean and STD of S_n.

2.2 Result 1

The following is the PDF of S_1. Note that this plot does not follow the red line (a normal distribution) at all. This is because $n = 1$ and therefore S_1 follows a uniform distribution; there is nothing to sum, because each uniform distribution only contains one value. CLT only starts to come out when $n > 1$; even the example with $n = 2$ given in the lab manual is starting to look like a normal distribution. This is because summing 10 000 uniform distributions is a lot.

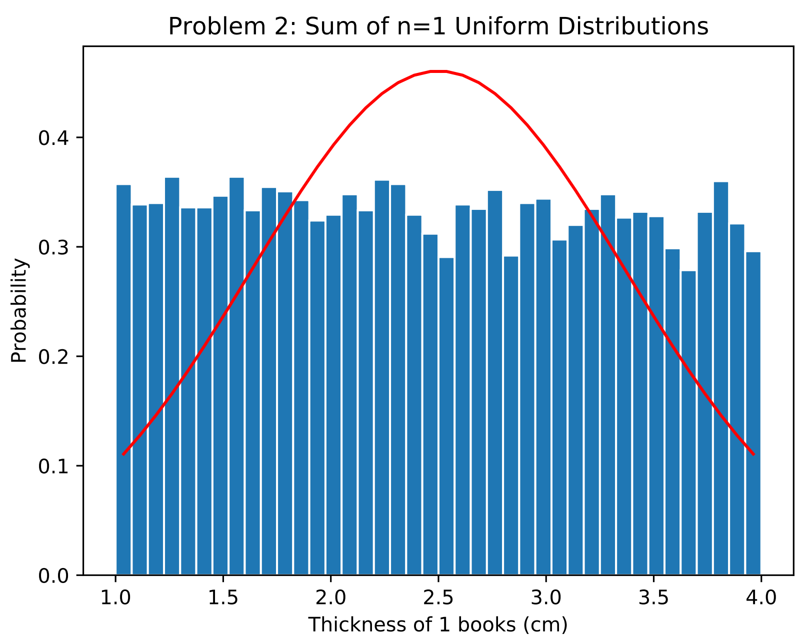


Figure 5: PDF of S_1.

2.3 Result 2

The following is a PDF of S_5. This random variable, and the two that follow it, are more reminiscent of a normally-distributed variable than S_1.

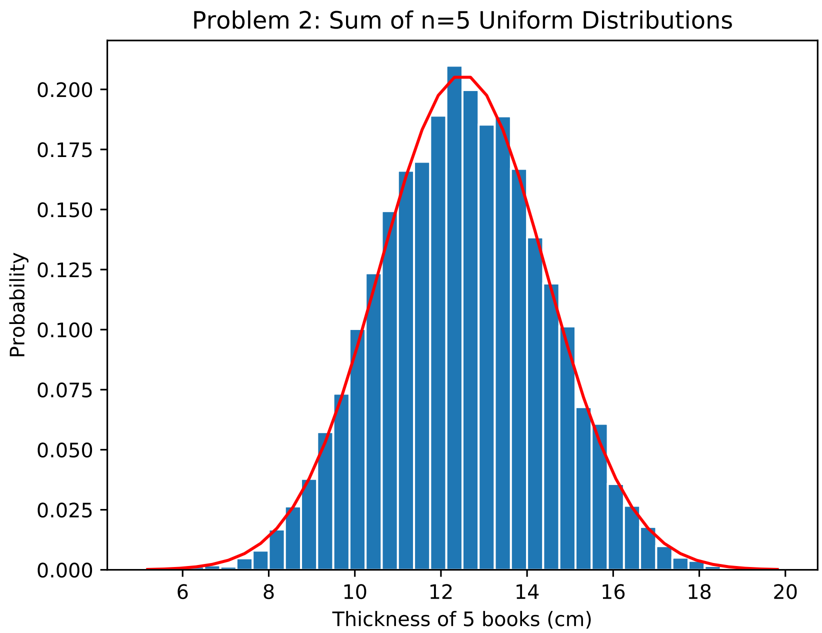


Figure 6: PDF of S_5.

2.4 Result 3

The following is a PDF of S_{10}.

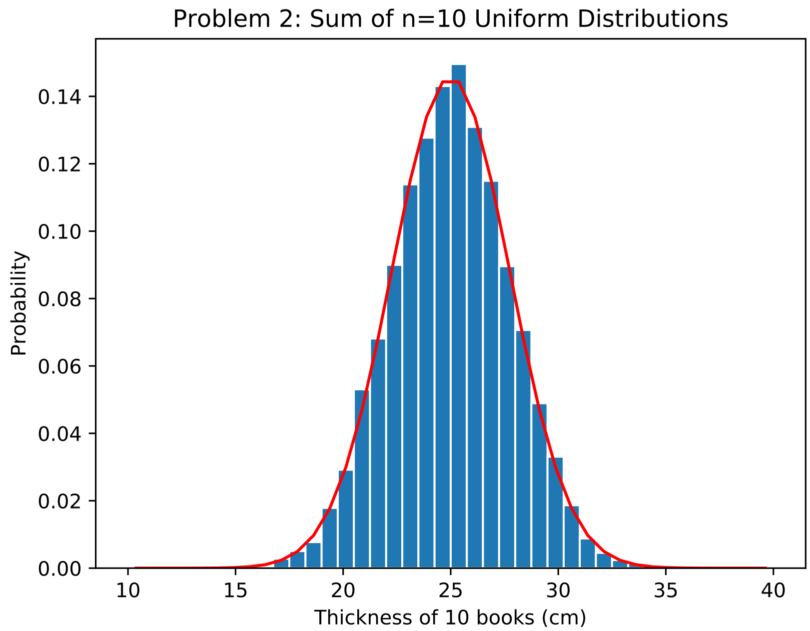


Figure 7: PDF of S_{10}.

2.5 Result 4

The following is a PDF of S_{15}.

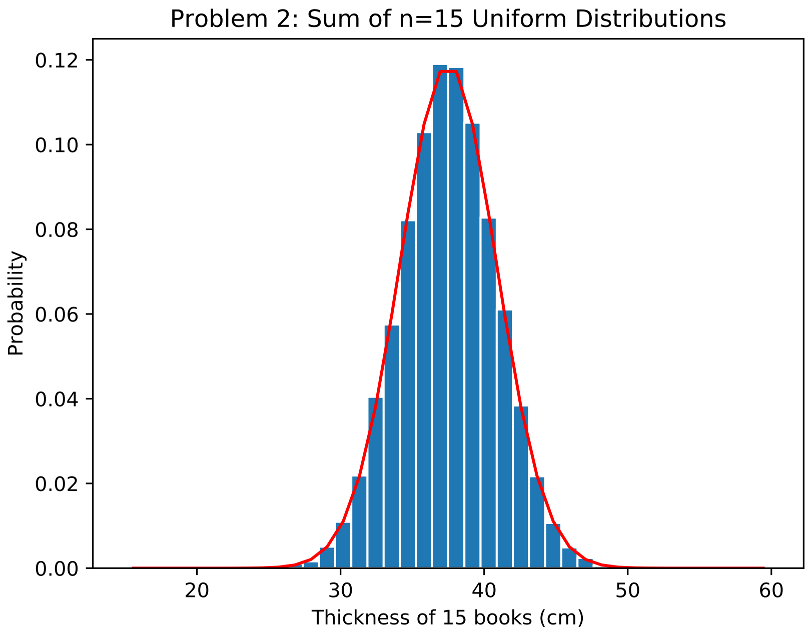
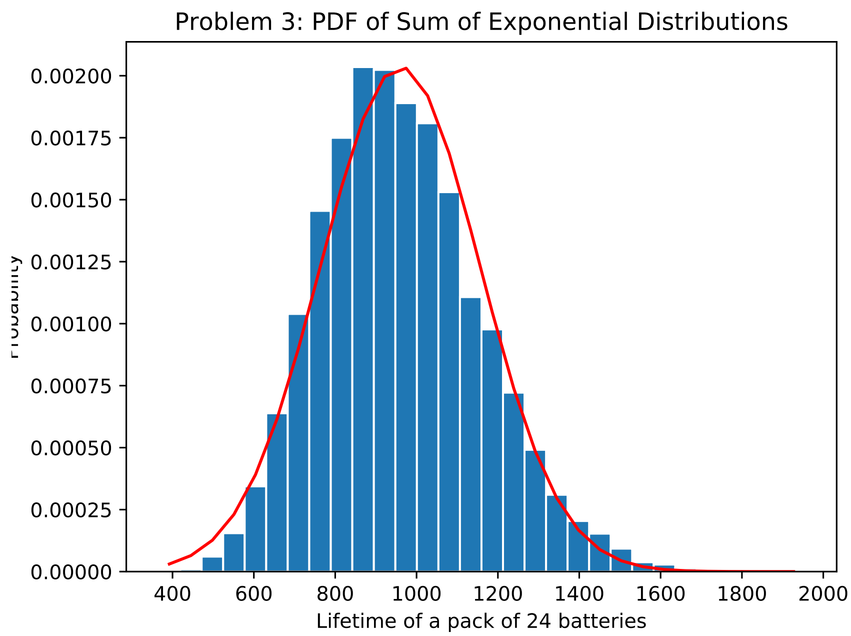


Figure 8: PDF of S_{15}.

3 Problem 3

3.1 Question

This problem puts together all previous ideas. In this scenario, 24 batteries come in a carton; we are tasked with creating the random variable $C = \sum_{i=1}^{10\,000} \text{expon}(x, 1/40, 24)$ and figuring out the PDF and CDF of this random variable. The two plots are as follows:



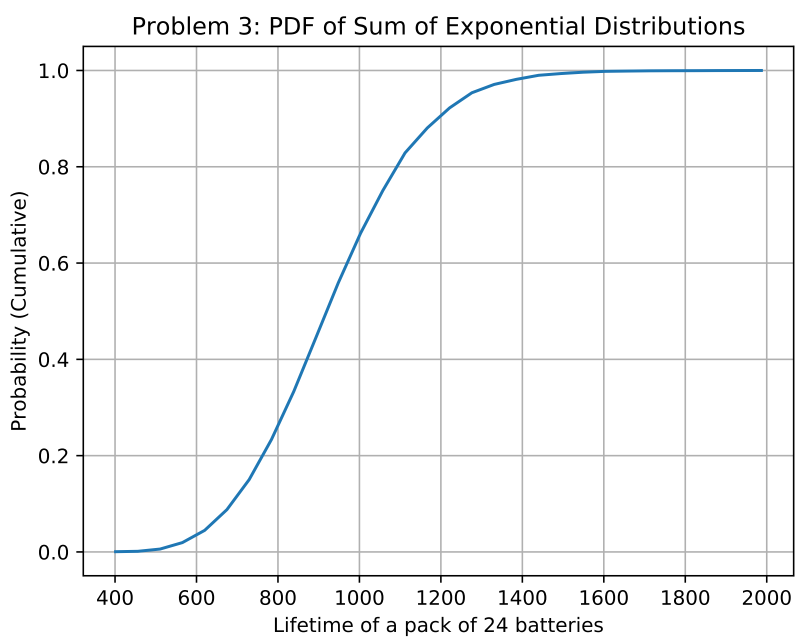


Figure 9: PDF and CDF of C.

Again, this random variable is akin to a normally-distributed random variable by CLT.

Using the plot of the CDF (because it is a vector graphic, feel free to zoom in if viewing the PDF), we can determine the following values:

QUESTION	**ANS.**
1. Prob. that the carton will last longer than three years	0.25
2. Prob. that the carton will last between 2.0 and 2.5 years	0.275

These values are obtained by the formulas $1-F(1095)$ and $F(912)-F(730)$, respectively; these are given in the lab manual and are visual estimates only.

4 Raw Output

```
Problem 1.1 starting...
The calculated expected value and standard deviation of X are
2.486907991195168 and 0.8665243032026828, respectively.

The theoretical expected value and standard deviation of X are
2.5 and 0.8660254037844386, respectively.

Problem 1.1 finished! Time taken: 0.21909165382385254

Problem 1.2 starting...
The calculated expected value and standard deviation of X are
40.06058878724873 and 40.38989783324536, respectively.

The theoretical expected value and standard deviation of X are
40 and 40, respectively.

Problem 1.2 finished! Time taken: 0.24942564964294434

Problem 1.3 starting...
The calculated expected value and standard deviation of X are
2.49401572198525 and 0.7556440339013419, respectively.

The theoretical expected value and standard deviation of X are
2.5 and 0.75, respectively.

Problem 1.3 finished! Time taken: 0.21802902221679688

Problem 2 starting...
Expected: n=1, mu=2.5, sigma=0.8660254037844386
Calculated: n=1, mu=2.4991530528563906, sigma=0.8632875171860275
Expected: n=5, mu=12.5, sigma=1.9364916731037085
Calculated: n=5, mu=12.525401937023174, sigma=1.9350492553264014
Expected: n=10, mu=25.0, sigma=2.7386127875258306
Calculated: n=10, mu=25.024355089434703, sigma=2.749866786667331
Expected: n=15, mu=37.5, sigma=3.3541019662496843
Calculated: n=15, mu=37.48169930332168, sigma=3.3972006925475697
Problem 2 finished! Time taken: 1.2205824851989746

Problem 3 starting...
Problem 3 finished! Time taken: 0.6729896068572998
```