CALIFORNIA STATE UNIVERSITY, LONG BEACH

EE 381 - Probability and Statistics with Applications to Computing

Central Limit Theorem

Simulate RVs with Exponential Distribution

0. Introduction and Background Material

0.1. Simulate a R.V. with Uniform Probability Distribution

The Python function "`numpy.random.uniform(a,b,n)`" will generate n random numbers with uniform probability distribution in the open interval $[a,b)$. The PDF of a random variable uniformly distributed in $[a,b)$ is defined as following:

$$f(x)=\begin{cases}\dfrac{1}{(b-a)}, & a\le x\le b\\ 0, & \text{otherwise}\end{cases} \quad ; \quad \text{and} \quad P(X\le x)=F(x)=\begin{cases}0, & x<a\\ \dfrac{(x-a)}{(b-a)}, & a\le x<b\\ 1, & x\ge b\end{cases}$$

It is noted that the mean and variance of a uniformly distributed random variable X are given by:

$$\mathrm{E}(X)=\mu_X=\frac{a+b}{2} \quad ; \quad \mathrm{Var}(X)=\sigma_X^2=\frac{(b-a)^2}{12}$$

0.2. Simulate a R.V. with Exponential Probability Distribution

The Python function "`numpy.random.exponential(a,n)`" will generate n random numbers with exponential probability distribution.

The PDF of a random variable exponentially distributed is defined as following:

$$f_T(t\,;\beta) = \begin{cases} \frac{1}{\beta}\exp(-\frac{1}{\beta}t), & t \ge 0 \\ 0, & t < 0 \end{cases}$$

From the above definition, the CDF of T is found as:

$$P(T \le t) = F(t) = \begin{cases} 0, & t < 0 \\ 1 - \exp(-\frac{1}{\beta}t), & t \ge 0 \end{cases}$$

It is noted that the mean and variance of the exponentially distributed random variable T are given by:

$$\mathrm{E}(T) = \mu_T = \beta \quad ; \quad \mathrm{Var}(T) = \sigma_T^2 = \beta^2$$

0.3. Central Limit Theorem

If $X_1, X_2, \ldots X_n$ are independent random variables having the same probability distribution with mean μ and standard deviation σ, consider the sum $S_n = X_1 + X_2 + \ldots X_n$.
This sum S_n is a random variable with mean $\mu_{S_n} = n\mu$ and standard deviation $\sigma_{S_n} = \sigma\sqrt{n}$.
The Central Limit Theorem states that as $n \to \infty$ the probability distribution of the R.V. S_n will approach a normal distribution with mean μ_{S_n} and standard deviation σ_{S_n}, *regardless of the original distribution* of the R.V. $X_1, X_2, \ldots X_n$. The PDF of the normally distributed R.V. S_n is given by: $f(s_n) = \frac{1}{\sigma_{S_n}\sqrt{2\pi}}\exp\{-\frac{(x-\mu_{S_n})^2}{2\sigma_{S_n}^2}\}$

PROBLEMS

1. The Central Limit Theorem

Central Limit Theorem.
Consider a collection of books, each of which has thickness W. The thickness W is a RV, uniformly distributed between 1 and 3 cm. The mean and standard deviation of the thickness will be: $\mu_w = \frac{1+3}{2} = 2;\ \sigma^2{}_w = \frac{(3-1)^2}{12} = 0.33;\quad \sigma_w = 0.57$.

These books are piled in stacks of $n = 1, 5, 10,$ or 15 books. The width S_n of a stack of n books is a RV (the sum of the widths of the n books). This RV has a mean $\mu_{S_n} = n\mu_w$ and a standard deviation of $\sigma_{S_n} = \sigma_w \sqrt{n}$.

Perform the following simulation experiments, and plot the results.

a) Make $n = 1$ and run $N = 10,000$ experiments, simulating the RV $S = W_1$.
b) After the N experiments are completed, create and plot a probability histogram of the RV S
c) On the same figure, plot the normal distribution probability function and compare the probability histogram with the plot of $f(x)$

$$f(x) = \frac{1}{\sigma_S \sqrt{2\pi}} \exp\{-\frac{(x-\mu_S)^2}{2\sigma_S^2}\}$$

d) Make $n = 5$ and repeat steps (a)-(c)
e) Make $n = 10$ and repeat steps (a)-(c)
f) Make $n = 15$ and repeat steps (a)-(c)

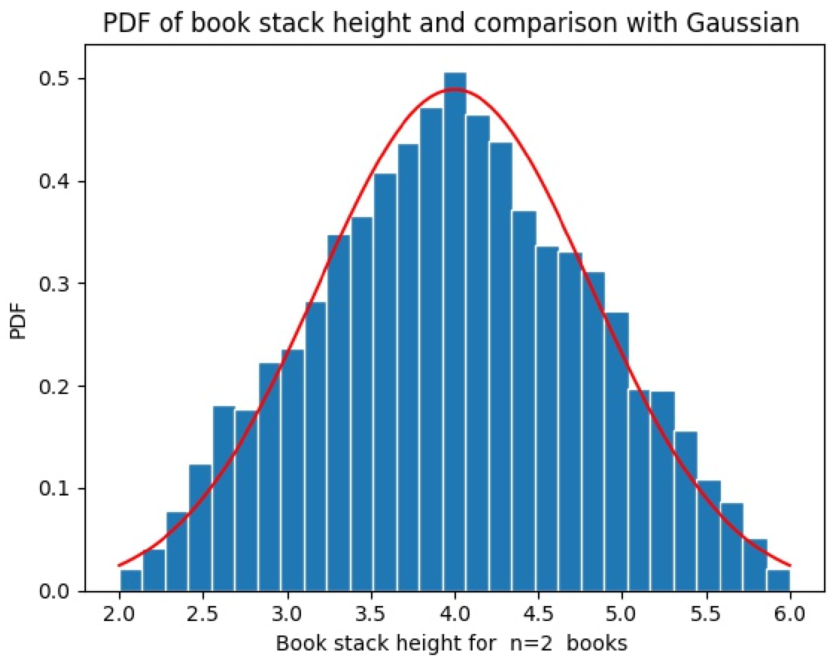
SUBMIT a report following the guidelines as described in the syllabus.
The report should include, among the other requirements:

- The four histograms for $n = \{1, 5, 10, 15\}$ and the overlapping normal probability distribution plots.
- The Python code, included in an Appendix.
- Make sure that the graphs are **properly labeled**.

An example of the graph for $n = 2$ is shown below.
The code below provides a suggestion on how to generate a bar graph for a continuous random variable X, representing the total bookwidth, when $n = 2$. Note that X has already been calculated.
The code shows the bar graph plotting only. It does not show the calculations for X s and it does not show the plotting of the Gaussian function.

Note that the value of "barwidth" is adjusted as the number of bins changes, to provide a clear and understandable bar graph.
Also note that the "density=True" parameter is needed to ensure that the total area of the bargraph is equal to 1.0.

```
    # X is the array with the values of the RV to be plotted
    a=1; b=3;           # a=min bookwidth ; b=max bookwidth
    nbooks=2; nbins=30; # Number of books ; Number of bins
    edgecolor='w';      # Color separating bars in the bargraph
    #
    # Create bins and histogram
    bins=[float(x) for x in linspace(nbooks*a, nbooks*b,nbins+1)]
    h1, bin_edges = histogram(X,bins,density=True)
    # Define points on the horizontal axis
    be1=bin_edges[0:size(bin_edges)-1]
    be2=bin_edges[1:size(bin_edges)]
    b1=(be1+be2)/2
    barwidth=b1[1]-b1[0] # Width of bars in the bargraph
    close('all')
#
    fig1=plt.figure(1)
    plt.bar(b1,h1, width=barwidth, edgecolor=edgecolor)
```



2. Exponentially Distributed Random Variables

Exponentially Distributed RVs

The goal is to simulate an exponentially distributed R.V. (T), given by the following PDF:

$$f_T(t) = \begin{cases} 2\exp(-2t), & t \geq 0 \\ 0, & t < 0 \end{cases}$$

1. Perform $N = 10,000$ experiments and generate the probability histogram of the random variable T. Plot the histogram of the RV T.

2. On the same graph, plot the function $g(x) = \begin{cases} 2\exp(-2x), & x \geq 0 \\ 0, & x < 0 \end{cases}$

   and compare to the experimentally generated histogram.

SUBMIT a report following the guidelines as described in the syllabus. The report should include, among the other requirements:

- the histogram of the RV T ;
- the graph of the function $g(x)$ overlaying the histogram on the same plot;
- the Python code.

3. Make sure that the graph is **properly labeled**.

3. Distribution of the Sum of RVs

This problem involves a battery-operated critical medical monitor. The lifetime (T) of the battery is a random variable with an exponentially distributed lifetime. A battery lasts an average of $\tau = 45$ days. Under these conditions, the PDF of the battery lifetime is given by:

$$f_T(t\,;\beta) = \begin{cases} \frac{1}{\beta}\exp(-\frac{1}{\beta}t), & t \ge 0 \\ 0, & t < 0 \end{cases} \qquad \text{where } \beta = 45$$

As mentioned before, the mean and variance of the random variable T are:

$$\mu_T = \beta \quad ; \quad \sigma_T^2 = \beta^2$$

When a battery fails it is replaced immediately by a new one. Batteries are purchased in a carton of 24. The objective is to simulate the RV representing the lifetime of a carton of 24 batteries, and create a histogram. To do this, follow the steps below.

a) Create a vector of 24 elements that represents a carton. Each one of the 24 elements in the vector is an exponentially distributed random variable (T) as shown above, with $\beta = 45$. Use the same procedure as in the previous problem to generate the exponentially distributed random variable T.

b) The sum of the elements of this vector is a random variable (C), representing the life of the carton, i.e.

$$C = T_1 + T_2 + \cdots T_{24}$$

where each T_j , $j = 1, 2, \cdots 24$ is an exponentially distributed R.V. Create the R.V. C, i.e. simulate one carton of batteries. This is considered one experiment.

c) Repeat this experiment for a total of N=10,000 times, i.e. for N cartons. Use the values from the N=10,000 experiments to create the experimental PDF of the lifetime of a carton, $f(c)$.

d) According to the Central Limit Theorem the PDF for one carton of 24 batteries can be approximated by a normal distribution with mean and standard deviation given by:

$$\mu_C = 24\mu_T = 24\beta \quad ; \quad \sigma_C = \sigma_T\sqrt{24} = \beta\sqrt{24}$$

Plot the graph of a normal distribution with

mean = μ_C and (standard deviation) = σ_C,

over plot of the experimental PDF on the same figure, and compare the results.

e) Create and plot the CDF of the lifetime of a carton, $F(c)$. To do this use the Python "`numpy.cumsum`" function on the values you calculated for the experimental PDF.

Answer the following questions:

1. Find the probability that the carton will last longer than three years, i.e. $P(S > 3\times 365) = 1 - P(S \le 3\times 365) = 1 - F(1095)$. Use the graph of the CDF $F(t)$ to estimate this probability.

2. Find the probability that the carton will last between 2.0 and 2.5 years (i.e. between 730 and 912 days): $P(730 \le S \le 912) = F(912) - F(730)$. Use the graph of the CDF $F(t)$ to estimate this probability.

3. **SUBMIT a report following the guidelines as described in the syllabus.** The report should include, among the other requirements:

- The numerical answers using the table below. Note: You will need to replicate the table, in order to provide the answer in your report. Points will be taken off if you do not use the table.
- The PDF plot of the lifetime of one carton and the corresponding normal distribution on the same figure.
- The CFD plot of the lifetime of one carton
- Make sure that the graphs are **properly labeled**.
- The code in an Appendix.

QUESTION	ANS.
1. Probability that the carton will last longer than three years	
2. Probability that the carton will last between 2.0 and 2.5 years	