CALIFORNIA STATE UNIVERSITY, LONG BEACH EE

381 - Probability, Statistics, and Stochastic Modeling

Projects

Central Limit Theorem

Simulate RVs with Arbitrary Distributions

0. Introduction and Background Material

0.1. 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: $f(s_n) = \frac{1}{\sigma_{S_n}\sqrt{2\pi}} \exp\{-\frac{(x-\mu_{S_n})^2}{2\sigma_{S_n}^2}\}$

0.2. The Uniform Probability Distribution

The standard MATLAB function "`rand`" will generate random variables with uniform probability distribution in the interval $[0,1]$, which is defined as following:

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

It can also be seen that: $$P(X \le x) = F(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \le x < 1 \\ 1, & x \ge 1 \end{cases}$$

A uniform distribution in the general interval $[a,b]$ is defined as:

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

0.3. Simulate a R.V. with Arbitrary Probability Distribution

As mentioned above, the standard MATLAB function "`rand`" will generate random variables with uniform probability distribution in the interval $[0,1]$. In many situations, however, you need to generate RVs with different distributions, such as the Poisson distribution. This section will provide you with the tools to generate RVs with arbitrary probability distributions using the uniform RV generate by the function "`rand`" .

To see how this can be done consider the following:

- Let X be a continuous random variable with an arbitrary cumulative distribution function $F(X)$. Assume for our purposes that $F(X)$ is strictly increasing, so that the inverse function $F^{-1}(y)$ exists for all y.
- Let U be a new random variable, defined as: $U = F(X)$. Obviously the range of U will be in the interval $[0,1]$, since U is equal to a CDF.
- Then: $P\{U < u\} = P\{F(X) < u\} = P\{X < F^{-1}(u)\} = F(F^{-1}(u)) = u$
- The relation $P\{U < u\} = u\,; \quad 0 \le u \le 1$ implies that the random variable U is uniformly distributed in $[0,1]$
- It follows that a random variable X with a given CDF $F(X)$ can be generated from a variable U uniformly distributed in $[0,1]$, by the transformation: $X = F^{-1}(U)$

Example: Simulate a RV (X) with a uniform PDF in the interval $[a,b]$.

Start from $U = F(X) \Rightarrow X = F^{-1}(U)$

The CDF of the random variable X is given by:

$$u = F(x) = \begin{cases} 0, & x < a \\ \dfrac{(x-a)}{(b-a)}, & a \le x < b \\ 1, & x \ge b \end{cases} \Rightarrow x = F^{-1}(u) = a + (b-a)u, \quad 0 \le u < 1$$

The following code will create a vector of RV (X) with a uniform PDF in the interval $[a,b]$.

```
N=100000;
u=rand(N,1);
a=1; b=3;
x=a+(b-a)*u;
```

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

Generate an exponentially distributed random variable T, with probability distribution given by the exponential function:

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

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

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

Based on the results of the previous section, T can be generated from the standard uniform distribution U as:

$$T = F^{-1}(U) = (-1/\alpha)\ln(1-U)$$

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

$$\mathrm{E}(T) = \mu_T = \frac{1}{\alpha} \quad ; \quad \mathrm{Var}(T) = \sigma_T^2 = \frac{1}{\alpha^2}$$

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 a minimum of a and a maximum of b cm. Use the values of a and b that were provide to you, and calculate the mean and standard deviation of the thickness. Use the following table to report the results. Points will be taken off if you do not use the table to report .

Mean thickness of a single book (cm)	Standard deviation of thickness (cm)
$\mu_w =$	$\sigma_w =$

The books are piled in stacks of $n = 1, 5, 10, \text{or } 15$ 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 $\sigma_{S_n} = \sigma_w\sqrt{n}$.
Calculate the mean and standard deviation of the stacked books, for the different values of $n = 1, 5, 10, \text{ or } 15$. Use the following table to report the results. Points will be taken off if you do not use the table to report.

Number of books n	Mean thickness of a stack of n books (cm)	Standard deviation of the thickness for n books
n=1	$\mu_w =$	$\sigma_w =$
n=5	$\mu_w =$	$\sigma_w =$
n=10	$\mu_w =$	$\sigma_w =$
n=15	$\mu_w =$	$\sigma_w =$

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

An example code of creating the PDF graph for $n = 2$ is shown below.

Example code for $n = 2$ and $a = 1$, $b = 3$

```
N=1e4;
a=1; b=3;
mu_w=(a+b)/2; sig_w=sqrt((b-a)^2/12);
X=zeros(N,1);

% CREATE EXPERIMENTS -----------------------------------------
n=2;
mu=n*mu_w;
sig=n^0.5*sig_w;
for k=1:N
        x=(b-a)*rand(1,n)+a;
        w=sum(x);
        X(k)=w;
end


% PLOTTING ------------------------------------

% Create plot of PDF
nbins=15;
del=(max(X)-min(X))/(nbins-1);
bins=min(X):del:max(X);
[h, xout]=hist(X,bins);
% Create the PDF
pdf=h/N/del;  % Divide by del to make Total_Area=1

figure(1);  bar(xout,pdf); grid;
hold;

% Overlay plot of Gaussian distribution for comparison
% Define inline function for graphing the normal distribution
gf=inline('(1/(sig*sqrt(2*pi)))*exp(-(z-mu).^2/(2*sig^2))');

% Overlay plot of Gaussian
z=min(X):0.1:max(X);
figure(1); plot(z,gf(mu, sig, z), 'r','LineWidth', 3)

% Create the CDF -- This will be needed for PROBLEM 3
%
cdf=cumsum(pdf*del);  % Multiply by del for the CDF to approach 1.0

figure(2);  bar(xout,cdf); grid;

% -----------------------------------------------------------------
```

2. Exponentially Distributed Random Variables

Exponentially Distributed RVs

The goal is to simulate an exponentially distributed RV (T) based on the discussion of a previous section. The generated RV T should be exponentially distributed. The probability density of the RV T is given by the following PDF:

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

In order to generate the exponentially distributed RV T, you must generate a random variable U first, where U uniformly distributed in $[0,1]$, i.e. $U = \text{rand}();$ Then you can create the RV T by using the formula

$$T = (-1/\alpha)\ln(1-U) \quad ; \quad \alpha = 2$$

This is considered one experiment.

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

2. On the same graph, plot the function $f(t) = \begin{cases} 2\exp(-2t), & t \geq 0 \\ 0, & t < 0 \end{cases}$ and compare to the experimentally generated histogram.

3. **SUBMIT** a Word file that contains:

- the histogram of the RV T;
- the graph of the function $f(t)$ overlaying the histogram on the same plot;
- the MATLAB code.

4. 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 β days, which has been provided to you. Under these conditions, the PDF of the battery lifetime is given by:

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

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 mean lifetime equal to β. 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 "`cumsum`" function of MATLAB 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 2.5 years	
2. Probability that the carton will last between 1.50 and 2.0years	