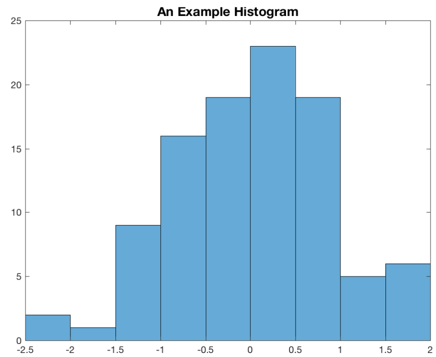
Lab 13: Histograms, Random Numbers & Monte Carlo Integration

Math 3341: Introduction to Scientific Computing Lab — Spring 2018

Histograms

A histogram is a graphical representation of the distribution of numerical data. Essentially, it is a type of bar graph.



To construct a histogram, the first step is to "bin" the range of values, that is, divide the entire range of values into a series of intervals, and then count how many values fall into each interval. The bins are usually specified as consecutive, non-overlapping intervals of a variable. The bins (intervals) must be adjacent, and are often (but are not required to be) of equal size.

If the bins are of equal size, a rectangle is erected over the bin with height proportional to the frequency: the number of cases in each bin. A histogram may also be normalized to display "relative" frequencies. It then shows the proportion of cases that fall into each of several categories, with the sum of the heights equaling 1.

A histograms is created in Matlab using `histogram(X)`, where `X` is an array of values. This command creates the plot itself, and so should be called in an appropriat figure environment. Titles, axes labels, etc can all be added in the usual way we add these options to the usual plots.

Random Number Generation

There are several ways to generate random numbers in Matlab. You can construct random numbers within a specified interval or as a normal distribution of values.

A normal distribution is specified by a given mean, μ, and a standard deviation σ. The graph of the standard normal distribution has $\mu = 0$ and $\sigma = 1$ and is pictured in Figure. When using the standard normal distribution 99.9% of your values will fall inside the interval $[-3, 3]$.

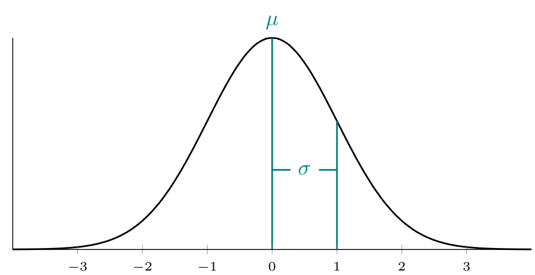


Figure 1: Standard Normal Distribution

The functions we will focus on in this lab are given in the table below.

Table 1

Command	**Description**
`rand(m,n)`	creates an m by n matrix of random values in $(0,1)$.In general, you can generate N random numbers in the interval (a,b) with the formula `r = a + (b-a).*rand(N,1)`
`randn(m,n)`	creates an m by n matrix of normally distributed random values using the standard normal distribution, i.e. with mean $\mu = 0$ and standard deviation $\sigma = 1$.
`normrnd(mu,sigma,[m,n])`	creates an m by n matrix of normally distributed random values with specified mean $\mu = 0$ and standard deviation σ.

Monte Carlo Integration

Monte Carlo integration is a technique for numerical integration using random numbers. While other algorithms usually evaluate the integrand at a regular grid, Monte Carlo randomly choose points at which the integrand is evaluated. This method is particularly useful for higher-dimensional integrals.

In class, you saw code that estimates the area of a circle. The general procedure is to generate a set of N random values in a given rectangle R, determine which of these values are inside of our circle and then calculate

$$I = \text{area}(R) \cdot \frac{S}{N}$$

where S is the sum of function values inside the circle. In other words, we are determining the fraction of the area of our rectangle R that the circle lives in. The code will look something like the following.

```
% area of unit circle using MC integration
N = 1000000; % number of random values
a = -1;      % endpoints of interval
b = 1;
xy = a + (b-a)*rand(2,N); % vector of random values in (a,b)
I = 0;
for j = 1:N
    I = I + inDisk( xy(1,j), xy(2,j) );
end
% this should be the area of the circle - the factor 4 is the
% area of the surrounding square
I = 4 * I / N;

fprintf('Area Unit Circle Using Monte Carlo\n')
fprintf('Area  = %f \n',I)
fprintf('Error = %f \n',abs(pi-I))

function [ val ] = inDisk( x, y )
    % determines if given point (x,y) is in the unit disk
    val = 0;
    if (x^2 + y^2) < 1
        val = 1;
    end
end
```

This procedure will be similar for calculating the volume of a sphere.

Lab Exercises

I. Histograms

1. Download the `lab13_histograms.m`. On an interval (a, b) of your choosing, create a vector of $N = 10000$ random values and then create a histogram of these values.
2. Create a vector of $N = 10000$ *normally distributed* random values. Create a histogram of these values.
3. Place these histograms in a single figure in which they appear side by side, with the same axes. What looks different about them? Don't forget to add labels to your axes and a title to each plot.

II. Monte Carlo Integration

1. Using the script file `MC_2D_Spring2018.m` that Dan provided you with as a starting point, write a script file that determines the value of the integral

$$\int_{-1}^{1}\int_{2x^2}^{1+x^2} (x+2y)\,dy\,dx$$

   using a Monte Carlo Integration Scheme with $N = 10000$.

2. The exact value of this integral is 32/15. In reporting your final results (using appropriate `fprintf` statements) state the result from Monte Carlo integration, the exact result, and the error in your calculation.

Submit your script files, plots, and output files in the provided report file.