

Project 2

Random Numbers, Statistical Distributions and Histograms

In this project, we are going to generate and plot some random numbers using two different statistical distributions, the Gaussian (or Normal or Bell Curve) distribution, and the Poisson distribution.

1. Gaussian Statistics

We will generate these random numbers (“random deviates”) using the built-in Uniform Random Number generator that is present in all programming languages. This URN generator, often called *ran()* or *rand()* or *random()* or something similar, produces a *pseudo-random* number between $[0,1]$.

These numbers are pseudo-random because they are generated with a spatially very high-dimensional (but deterministic) algorithm, but they pass most tests for randomness due to the high dimensionality of the space of numbers.

Or in other words, if you produce numbers in a space of 2^{32} dimensions, then project the numbers down onto the one-dimensional number line, they will appear for the most part random.

First consider the problem of Gaussian random numbers. The basic Gaussian Probability Density Function $p_G(x)$ is given analytically by the equation:

$$p_G(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

This statistical distribution is one of the most common. It is used to describe student grades, height of individuals in a population, IQ scores, accident rates for insurance, and a large number of other uses.

There are various ways to produce Gaussian random numbers from the URNs. One easy way is to use the Central Limit Theorem:

- First generate a series of URNs, but on the interval $X \in [-1,+1]$ (i.e., use a simple linear transformation to go from $[0,1]$ to $[-1,+1]$)
- Then add, say, N of them together: $x = \frac{1}{\sqrt{N}} \sum_{i=1}^N X_i$
- Then the result will be random numbers x having mean $\mu = 0$ and standard deviation $\sigma = 1$

More efficient and standard ways are to use the Box-Muller transformation:

http://en.wikipedia.org/wiki/Box%E2%80%93Muller_transform

2. Poisson Statistics.

To generate Poisson random deviates, we need to do something just a bit more complicated. The Poisson distribution describes counting events, where the events are identical, independent, and uncorrelated. The basic parameter is λ , the rate at which the events (such as rate of neutron decay, or cars arriving at Nuggets per hour) occur. The Poisson Probability Density function $p_{Pn}(n)$ for the number n of events occurring in an interval is given by:

$$p_{Pn}(n) = \frac{\lambda^n}{n!} e^{-\lambda}$$

In addition, the Poisson Probability Density function $p_{PI}(t)$ for some interval t between events is given by:

$$p_{PI}(t) = \lambda e^{-\lambda t}$$

The Cumulative Poisson probability $P_{PI}(t)$ for inter-event intervals less than t is given by:

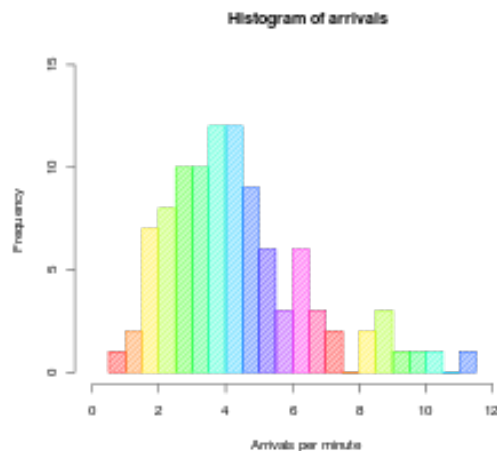
$$P_{PI}(t) = 1 - e^{-\lambda t}$$

There are several algorithms that can be used to produce Poisson random deviates. They are summarized here (for the Wikipedia entry, scroll down about half way to find the pseudo code that is used). http://en.wikipedia.org/wiki/Poisson_distribution

It can be shown that the Poisson mean $\mu = \lambda$ and the Poisson standard deviation $\sigma = \sqrt{\lambda}$

3. Histograms

A histogram is basically a discrete-variable representation of a probability density function. An example is shown in the figure below from (<http://en.wikipedia.org/wiki/Histogram>) :



A histogram displays the number of times a given random deviate occurs in a population as a function the value of the random deviate.

Or, to put it another way, suppose we have a population of $N=1,000,000$ numbers x . We want to display the number of values of x that occur between x_i and $x_i + \Delta$. Thus a histogram is a plot of $n_i = n(x_i < x \leq x_i + \Delta)$ as function of x_i .

Note that if we normalize the values of n_i so that $\sum_1^N n_i = 1$, then what we obtain is a discrete form of the probability density function for the process.

4. Assignment.

- a) Write a computer code that will generate 100,000 random deviates for both Gaussian and Poisson statistics. Plot the resulting histograms.
- b) Normalize your histograms so that $\sum_1^N n_i = 1$, and plot the best fitting Gaussian and Poisson probability density functions on top of the histograms.