Expt 1: Statistics and Random Variables

Part I

Study the sample Scilab code '*gauss1.sce*' and plot the histogram of random numbers produced using a gaussian distribution. How will you modify the generated *r(n)* to get an arbitrary mean $\bar{x}$ and standard deviation σ?

1) Modify the sample code and calculate the average and standard deviation of *r(n)* using two '*for*' loops. (Set *normalization* to false and remove the gaussian comparison plot for simplicity). Try for N = 1000 and 100000 events. Use built-in help in Scilab or check the web. Show the results to your TA for credits. Check your answers using the built-in *mean*() and *stdev*() functions.

2) Using a '*for*' loop and '*if*' statements, count the fraction of events that lie outside range $x > 1\sigma$, $x > 2\sigma$ and $x > 3\sigma$ and compare with the expected numbers. What do you think are the errors on these actual measured counts? Try for N =1000 and 100000 events.

3) Note that in part 1, the calculation of standard deviation σ requires another separate '*for*' loop, after calculating the average in the first '*for*' loop. This problem can be avoided by finding sum of deviations around an arbitrary number k, which need not be the average $\bar{x}$, as:

$$\bar{x} = \Sigma\,(x_i - k)\,/N + k$$

$$\sigma_n^2 = \Sigma\,(x_i - k)^2/N - (\bar{x} - k)^2$$

Write a Scilab code to verify this. Take the arbitrary $k = 1$. Note when $k = \bar{x}$, these two expressions take the usual standard forms. Check your answers using the built-in *mean*() and *stdev*() functions.

4) The Scilab '*rand*' function can also produce a uniformly random distribution between 0 and 1. Show the histogram of this distribution and find and compare $\bar{x}$ and σ_n with the expected values. Set histogram *normalization* to true *%t* to see the pdf. From now on, for simplicity, just use the *mean*() and *stdev*() functions.

5) When you add two random variables $z = x + y$, show, as homework, that the pdf of the result z is given by the convolution: $p_Z(z) = \int p_X(x)p_Y(z-x)dx$. Check this by adding and histogramming two uniformly distributed random variables. Set *normalization* to true to see the pdf.

6) Central Limit Theorem: when you add and *average* a very large number of n random variables, each with a different type of 'bell-shaped' pdfs of arbitrary $\bar{x}$ and σ_n, the result tends to a gaussian distribution. Check this by producing and adding five different uniform and five different gaussian distributions. Note that if the pdfs are all gaussian with the same $\bar{x}$ and σ_n, the new $\sigma^2_{ave} = \sigma_n^2/n$ becomes smaller as n increases. Also the poisson and binomial distributions became gaussian for large N. For these reasons, the gaussian distribution is very common in nature.

Part II

1) Study the second sample code that generates a discrete random number using a poisson distribution. Find mean and standard deviation, and compare with the expected values. Find the fraction of events occur at the mean value λ and compare this with the expected value: $P(n) = \lambda^n e^{-\lambda}/n!$ (You can zoom the Scilab plot and enable the axes to see the values clearly.)

2) Show, as homework, that the sum of two poisson distributed random numbers is also poisson. Check this by adding two random numbers. (Hint: you can use the binomial expansion theorem to prove this.)

3) Check that when the mean is very large, say $\lambda > 20$, the poisson distribution becomes nearly gaussian. Show this as homework.

4) Generate a binomial distributed random number and find its mean and standard deviation. Compare with the expected values from: $P(n) = {}^{N}C_{n}\, p^{n}(1-p)^{N-n}$. Take $p = 0.2$ and $N = 10$ trials.

5) When you add two binomial random numbers, is the result binomial distributed? Check this for $p_1 = 0.2$, $p_2 = 0.8$ and $N = 10$ trials.

6) Note that when the trials N is very large and Np is finite, binomial distribution becomes nearly gaussian. Check this for the case $p=0.05$ and $N = 200$ trials.

(Try to do as many parts as possible during the lab hour, and the rest can be completed at home. Show the results to your TA for credits.)