Gan
Phys 416

LAB 3

The goal of this lab is to familiarize the student with the Central Limit Theorem, an amazing result from probability theory that explains why the Gaussian distribution (aka "Bell Shaped Curve" or Normal distribution) applies to areas as far ranging as economics and physics. Below are two statements of the Central Limit Theorem (C.L.T.).

I) "If an overall random variable is the sum of many random variables, each having its own arbitrary distribution law, but all of them being small, then the distribution of the overall random variable is Gaussian".

II) Let Y_1, Y_2,...Y_n be an infinite sequence of independent random variables each with the same probability distribution. Suppose that the mean (μ) and variance (σ^2) of this distribution are both finite. Then for any numbers a and b:

$$\lim_{n\to\infty} P\left[a < \frac{Y_1 + Y_2 + \cdots Y_n - n\mu}{\sigma\sqrt{n}} < b\right] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-(1/2)y^2}\, dy$$

Thus the C.L.T. tells us that under a wide range of circumstances the probability distribution that describes the sum of random variables tends towards a Gaussian distribution as the number of terms in the sum $\to \infty$.

Some things to note about the C.L.T. and the above statements:

a) A random *variable* is not the same as a random *number*! Devore in "Probability and Statistics for Engineering and the Sciences" defines a random variable as (page 81):

"A random variable is any rule that associates a number with each outcome in S".

b) If y is described by a Gaussian distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$ then the probability that $a < y < b$ is:

$$P(a < y < b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-(1/2)y^2}\, dy$$

c) The C.L.T. is still true even if the Y_i 's are from different probability distributions! All that is required for the C.L.T. to hold is that the distribution(s) have a finite mean(s) and variance(s) and that no one term in the sum dominates the sum. This is more general than definition II).

1) In this exercise, we will see that the landing location of stainless steel balls is governed by a Gaussian distribution. The apparatus (Fig. 1) used consists of a grid of evenly spaced pins sandwiched between two plates. When a ball is dropped into the grid, it can scatter to the left or right on the first pin it encountered. The scattered ball then hits the next pin and re-scatters either to the left or right. The process continues until it reaches the bottom of the grid where there are evenly spaced slots to receive the ball. The array of slots acts like a “histogram machine” with each slot representing a “bin” in a histogram. The apparatus is slightly tilted so that the balls in a slot will accommodate from the bottom for easy counting.

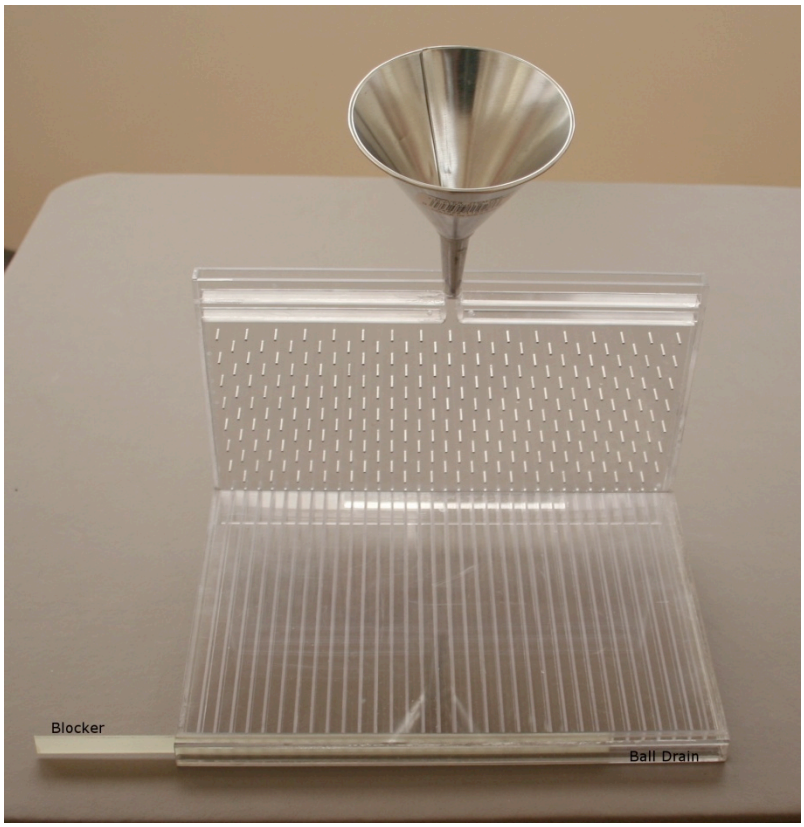

Fig. 1: A picture of the apparatus for demonstrating the C.L.T.

In this experiment, the final location where a ball landed is determined by the number of right and left scatterings. There are 14 rows of staggered pins. This is the n in the C.L.T. equation. Each scattering contributes a deviation, Y_n, from the center horizontal location where the ball is released. The deviation can have positive or negative sign and its value depends on the particular angle of the scattering. The final location is then the sum of all Y_n's. C.L.T. tells us that the location should be distributed approximately like a Gaussian probability distribution.

The procedure for the experiment is as follow:

- Make sure the blocker is inserted all the way in so that no ball will escape from the slots.
- Gather ~70 ml of SS balls in a beaker. This corresponds to ~300 balls, enough to almost fill up the central slot occasionally.
- Hold the funnel with the base in the notch at the top of the apparatus and pour the balls slowly into the funnel. Pouring too fast will cause the balls to jam inside the funnel. This will happen more often than you can imagine.
- Count the number of balls in each slot to produce a histogram.
- Pull out the blocker and hold the ball drain over a beaker to collect the balls. You might need to lightly shake the apparatus to collect all the balls as they often jam up.

We should compare the histogram to a Gaussian distribution for a more qualitative understanding of C.L.T. First, calculate the statistical uncertainty on the number of measurements in each bin and plot the error on the data point in the histogram. Then calculate your average location (μ) and variance (σ^2) of your measurement. Superimpose a Gaussian p.d.f. on your histogram (this can be

done with Kaleidagraph) and comment on how Gaussian-like your measurements are. For a Gaussian probability function describing N measurements with mean μ and variance σ^2, the number of measurements in a bin (ΔN_i) of bin size Δx (= 1) can be approximated by:

$$\Delta N_i = \frac{N\Delta x}{\sigma\sqrt{2\pi}} e^{\frac{-(x_i-\mu)^2}{2\sigma^2}}$$

where we take x_i to be the value at the center of the i^{th} bin.

2) In this exercise we will use the computer and definition II) to illustrate the C.L.T. This exercise uses the properties of the random number generator (RAN). The random number generator gives us numbers uniformly distributed in the interval [0, 1]. This uniform distribution (p(x)) can be described by:

p(x) = 1 for 0 < x < 1
p(x) = 0 for all other x.

a) Prove using the integral definitions of the mean and variance that the uniform distribution has $\mu = 1/2$ and $\sigma^2 = 1/12$.

b) According to definition II) if we add together 12 ($Y_1 + Y_2 + \cdots Y_{12}$) numbers taken from our random number generator RAN then:

$$P\left[a < \frac{Y_1 + Y_2 + \cdots Y_{12} - 6}{1} < b\right] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-(1/2)y^2}\, dy$$

This says that just by adding 12 random numbers (each between 0 and 1) together and subtracting off 6 we will get something that very closely approximates a Gaussian distribution for the sum ($\equiv Z = Y_1 + Y_2 + \cdots Y_{12} - 6$) with $\mu = 0$ variance $\sigma^2 = 1$! Write a program to see if this is true. Generate 10^6 values of Z and make a histogram of your results. I suggest using x bins of 0.5 unit, e.g. $Z < -5.5$, $-5.5 \le Z < -5.0$...$Z > 5.5$. Superimpose a Gaussian p.d.f. with $\mu = 0$ and $\sigma^2 = 1$ on your histogram and comment on how well your histogram reproduces a Gaussian distribution.

NOTE: save this program, we will use it again in LAB 5.