**Lecture 12**

**[From Ref. #4]**

**Bernoulli's theorem**

[Stated without proof]

Let A denote an event whose probability of occurrence in a single trial is *p*. Let *k* denote the number of occurrences of A in *n* independent trials. Then, for any chosen degree of accuracy e, we have:

Prob[ |*k*/*n* - *p*| > e ] < *p*(1–*p*)/*n*e2

That is, for any chosen degree of accuracy e, we can make the LHS probability arbitrarily low by increasing the number of trials *n*. If we choose smaller e, we need larger *n*.

Note that LHS is the probability of “error" =|*k*/*n* - *p*| exceeding e.

This theorem forms the basis for interpreting *probability* as *relative frequency* (in a “large enough” number of trials). The property described in this theorem can also be dubbed ***the law of large numbers***.

\*\*\*

Let X and Y be two random variables. It can be shown, with the usual notation, and without much difficulty, that:

ma+bX = a + bmX Here a+bX is another RV, a function of X.

s2a+bX = b2 s2X

mX+Y = mX + mY

s 2X+Y = s2X + s2Y

\*\*\*

Let random variable X denote the outcome of an experiment, and let m and s2 be its mean and variance, respectively.

Now let us assume that the experiment is repeated *n* times. We denote the outcomes of these trial as x1, x2 .... xn respectively.

In effect, we have taken *n* *samples* of X. Then we can also denote the RV associated with the kth sample as Xk , k = 1, 2 ... n.

These Xk are independent, identically distributed (i.i.d.) random variables.

We can define two other random variables Xmean and Xvar as, respectively:

Xmean = (S1..n Xk)/*n*

and

Xvar = [S1..n (Xk – Xmean)2 ]/(*n*-1)

These two **random variables** are known as, respectively, *sample mean* and *sample variance*.

**IMPORTANT NOTE**: The sample mean we calculate numerically after *n* actual trials is only *an instance* of the random variable Xmean.

Then it can be shown that:

Exp[ Xmean ] = m

Var[ Xmean ] = s2/*n*

and

Exp[ Xvar ] = s2

So values of Xmean and Xvar serve as *estimators* of m and s2 respectively.

**CENTRAL LIMIT THEOREM**

[slightly simplified version]

Theorem: Let X1 , X2 ... Xn ... be a sequence of independent, identically distributed RVs with E(Xi) = m and Var(Xi) = s2 , for i = 1, 2 ...



Let Sn be defined as Sn = X1 + X2 + ... + Xn.

Then, under certain very general conditions, as *n* tends to infinity, Sn follows a normal distribution with mean = *n*m and variance *n*s2.

Corollary: Let another RV Xmean defined as Xmean = Sn/n = [X1 + X2 + ... + Xn]/n. Then, as *n* tends to infinity, Xmean tends to N( m, s/**sqrt(n)** ).

Recall that Xmean is the sample mean, and *n* is the sample size.

NOTE: NOTHING WHATSOEVER IS SAID ABOUT THE DISTRIBUTION OF X!

Simple example:

You are given a fair die. Roll it once and note the number which comes up. Roll it a second time and note the number. Roll it a third time and note the number ... Roll it the *n*th time and note the number. You have *n* samples.

Calculate the average of the above *n* numbers. Call it Xmean, which is of course a random variable.

As we increase *n*, Xmean comes closer and closer to N( 3.5, **sqrt(2.917/*n*)** ).

How?

Roll of a single fair die has mean 3.5 and variance 2.917.

To cross-check:

[1+2+3+4+5+6]/6 = 3.5.

[(2.5^2+1.5^2+0.5^2)\*2]/6 = 2.917.

Meaning: In practical terms, sample mean becomes “more and more reliable” as sample size increases 🡪 The “bell shape" of Xmean grows more and more “peaked”.

But note that this particular result here assumes that “the population is homogeneous”; or, in other words, the random variables X1 , X2 ... Xn ... are *independent and identically distributed*.

How does this happen?

Marbles are being dropped from a height, through a vertical tube, on a flat surface. The position of each initial impact is marked with a dot. Clearly, after *n* such drops, an equal number of dots will be marked on the surface.

How will these dots be distributed around their central point? [The central point will be aligned with the axis of the tube.]

Examples from Ref. #02 [slightly modified] :

1. The lifetime of a brand of electric bulbs is an RV with mean 1200 hours and standard deviation 250 hours. Using CLT, find the probability that the average lifetime of a batch of 100 bulbs is less than 1150 hours.

Let RV X represent the lifetime of a given bulb, with mean 1200 hours and standard deviation 250 hours [distribution unknown]. Let Xmean represent the RV which is the mean lifetime of the given batch of 100 bulbs.

From the corollary to CLT given above, we know that the distribution of Xmean is N( 1200, 250/sqrt(100) ) = N( 1200, **25** ). Note the standard deviation!

We need Prob[ Xmean < 1150 ].

We can standardize the distribution of Xmean by subtracting mean 1200 from it and then scaling the result by a factor equal to 1/[std dev of Xmean].

Then the required probability is given by Prob[ z < -2 ]. **How?**

From the table of standard normal z, we know that:

Prob[ z > 2 ] = 1 - Prob[ z < 2 ] = 1 – 0.9772 = 0.0228

From the symmetry of standard normal distribution about zero, we then know that Prob[ z < -2 ] = Prob[ Xmean < 1150 ] = 0.0228. This is the required answer.

Suppose you are taking delivery of these bulbs in huge quantity. You test a randomized sample of 100 bulbs, and find that their mean lifetime is 1125 hours. You are faced with a question: Should we accept the delivery?

What will you recommend?

2. X1 , X2 ... Xn ... represent hourly traffic at a port, which is assumed to be Poisson distributed with parameter l = 2. Samples of 72 traffic readings are taken, and their average Xmean calculated. Find Prob[ 1.5 < Xmean < 2.5 ].

Xi are Poisson distributed RVs with mean = variance = l = 2.

By CLT, Xmean has normal distribution with mean 2 and standard deviation sqrt(2)/sqrt(72) = 1/6. Note that 0.5 = 3\*[std dev of Xmean]

Therefore, in terms of the standard normal z = N( 0, 1 ), the specified bounds on Xmean correspond to -3 < z < 3 [in units of s.]

Tables of standard normal distribution are used to find the answer. Do that using Excel as exercise.

Question: When is sample size *n* “big enough"? What happens if we choose it to be too small or too big?