LECTURE 02

SAMPLING DISTRIBUTIONS AND CENTRAL LIMIT THEOREM

Outline of today's lecture:

I. Uniform Distribution Functions in Excel

The Uniform distribution assigns an equal probability to all events. In Bayesian probability theory the uniform distribution represents *ignorance* in the sense that the all events are assumed to be equally probable.

For example, if a fair die is thrown, the probability of obtaining any one of the six possible outcomes is 1/6. Since all outcomes are equally probable, the distribution is uniform.

Continuous Uniform Distribution The form of the continuous uniform distribution function is given by:

$$f(X)=\frac{1}{b-a} \ \ if \ \ a \le X \le b \ \ and \, 0 \ elsewhere$$

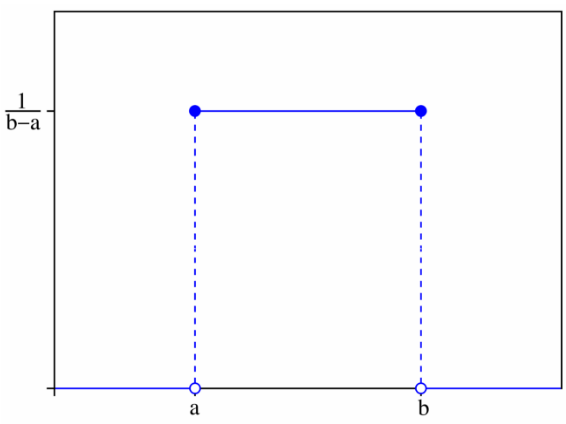
where a is the minimum value of X and b is the maximum value.

The expected value and variance of the continuous uniform distribution are:

$$E(X)=\mu=\frac{a+b}{2}, \quad Var(X)=\sigma^2=\frac{(b-a)^2}{12}$$

The shape of the uniform density function is:

Probability Density Function



Continuous Uniform Distribution

Discrete Uniform Distribution The outcome space of X is the set of all integers x such that $a \le X \le b$. The form of the discrete uniform distribution function is given by:

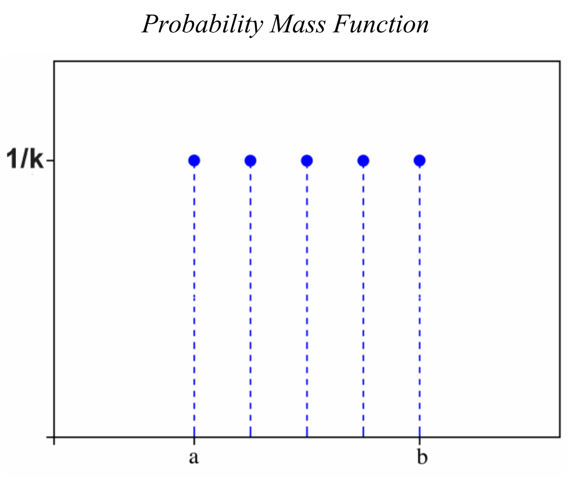
$$f(X)=\frac{1}{k} \quad \text{if } a \le X \le b \text{ and } 0 \text{ elsewhere}$$

where $k = b - a + 1$ for all x in the outcome space of X. This means that, k is the number of possible outcomes.

The expected value (mean) and variance of the *discrete uniform distribution* are:

$$E(X) = \mu = \frac{k+1}{2}, \quad Var(X) = \sigma^2 = \frac{k^2 - 1}{12}$$

The shape of the uniform density function is:



- Note that, $k = b - a + 1$.
- Hence, if b=6 and a=1, k=6
  - An example of such a case can be *tossing a fair die*: each face may happen with equal probability of *k=1/6*, where a=1 and b=6.

Discrete Uniform Distribution

A. RAND Function (Continuous Uniform Distribution)

You can use the *RAND()* worksheet function to generate random numbers from a *continuous uniform distribution*.

=RAND()

The *RAND* function returns numbers from the interval *[0,1)*, and if you need to generate numbers from another interval, you should use the following formula:

=RAND() * (b−a) + a

- This will return random numbers from the *interval [a,b)*. In other words, greater than or equal to *a*, and less than *b*.

Examples

- To generate a random number between 0 and 1, use *=RAND()*
- To generate a random number between 0 and 100, use *=RAND() *100*
- To generate a random number between 100 and 200, use *=RAND() *(200-100)+100*

B. RANDBETWEEN Function (Discrete Uniform Distribution)

Excel 2007 The RANDBETWEEN function can be used to generate discrete distribution.[1]

=RANDBETWEEN(1,6)

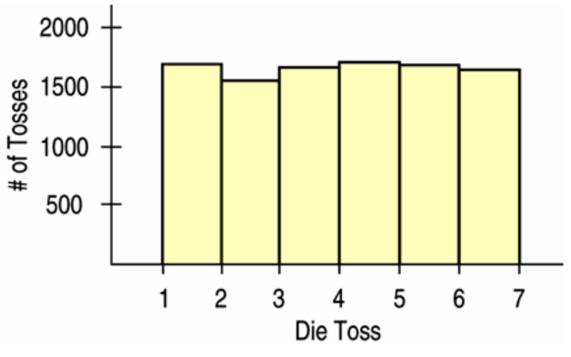
returns a *discrete uniform distribution* of integers between *1* and *6*.

II. Simulating the Sampling Distribution of a Mean

A simulation model is a computer model that imitates a real-life situation.

We can use simulation to get a sense as to what the sampling distribution of the sample mean might look like.

Now, let's go through a simulation of 10,000 tosses of a die. A histogram of the results is as follows (*Mean - The "Average" of One Die*):

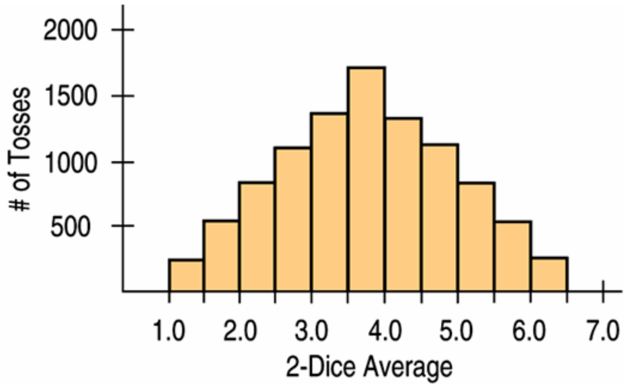


[1] **In Excel 2003** The *INT* function (integer function) used in conjunction with the RAND function generates discrete distribution:

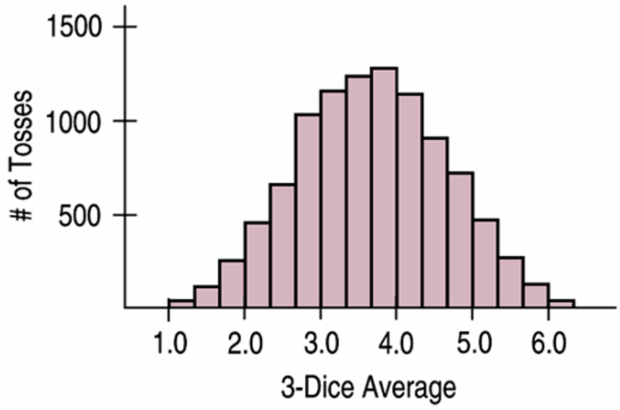
*=INT(1+6*RAND())*

returns a discrete uniform distribution of integers between 1 and 6.

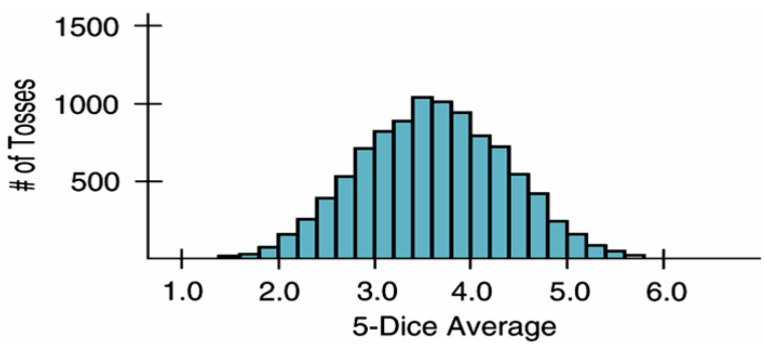
Looking at the average of two dice after a simulation of 10,000 tosses (*Means – Averaging Two Dice*):



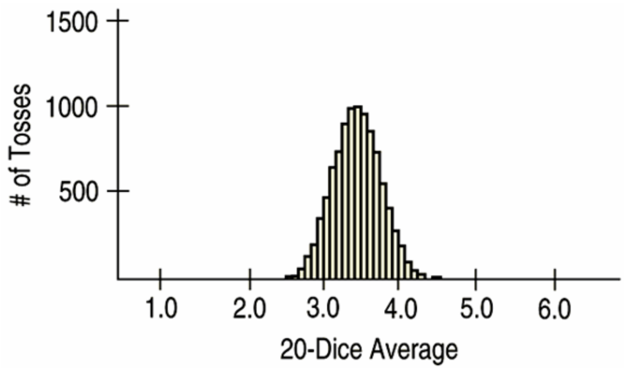
The average of three dice after a simulation of 10,000 tosses looks like (*Means – Averaging Three Dice*):



The average of 5 dice after a simulation of 10,000 tosses looks like:



The average of 20 dice after a simulation of 10,000 tosses looks like:



What the Simulations Show?

- As the sample size (number of dice) gets larger, each sample average (sample mean) is more likely to be closer to the population mean.
    - So, we see the shape continuing to tighten around 3.5

- And, it probably does not shock you that the sampling distribution of a mean becomes Normal.

III. The Central Limit Theorem (CLT)

Hence, we can summarize the main finding of this simulation exercise as follows:

- The sampling distribution of any mean becomes more nearly Normal as the sample size grows.

    - All we need is for the observations to be independent and collected with randomization.

    - We don't even care about the shape of the population distribution!

In fact, this is a fundamental theorem of statistics and it is called the *Central Limit Theorem (CLT)*.

- The CLT is surprising:

    - Not only does the histogram of the sample means get closer and closer to the Normal model as the sample size grows, but this is true regardless of the shape of the population distribution.

- The CLT works better (and faster) the closer the population model is to a Normal itself.

- It also works better for larger samples.
    - Particularly, samples of size larger than 30, $n \geq 30$

The Central Limit Theorem (CLT) The mean of a random sample has a *sampling distribution* whose shape can be approximated by a Normal distribution. The larger the sample, the better the approximation will be.