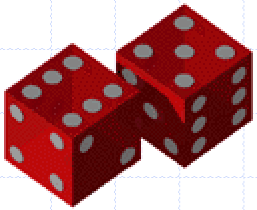
LEC2A:
Review of Probability & Statistics

1. Random events or experiments
2. Sample space (s)
3. The relative frequency & probability
4. Random number and its Probability distribution



Reading Assignment: The lecture ppt.

1. Random events or experiments

There are usually two types of **events** or **experiments:**

Type 1: Events or experiments for which the outcome is **certain or 100 % sure.** We call such events or experiments **deterministic events or deterministic experiments,** e.g.,

- You will fall on the ground if you jump;
- You will not pass the class if you don't do exam.

Type 2: Events or experiments for which the outcome is **uncertain** or **not sure.** We call such events or experiments **random events or random experiments,** e.g.,

- You will get a head if you flip a fair coin;
- You will get "3" if you roll a die;
- The amount of rain that we get next week.

A Random Event or Experiment

is such an event or experiment for which the outcome can’t be predicted with certainty and **can only be predicted with probability**.

When we are dealing with a random event, we may want to know:

- What is the probability of flipping a fair coin to get a head?
- What is the probability that I pass this class?

In order to answer these questions or find out the probabilities, we need to know the **sample space** or all possibilities.

2. SAMPLE SPACE (S)

Let’s look at some examples:

Example 1: Flipping a fair coin and observing if we get a head or tail.

Before we do the experiment or flip a coin, we don’t know if we’ll get a head or tail. However, we know we’ll get either a head or tail, i.e., we know all the possible outcomes or **Sample Space (S)** or **the collection of every possible outcome.**

The sample space for the random experiment of flipping a fair coin is

S = {Head, Tail} = {H, T} = {1,0}

Examples of Sample Space

Example 2: The sample space for rolling a die

What is the sample space (S)?

S = {1,2,3,4,5,6} (discrete)

Example 3: The sample space for the weight (w) of a bird captured by a biologist:

S = {0 < w < 10} (continuous)

Example 4: The sample space for drilling for oil or gas

S = {Success, Failure}

Now, we know what is a sample space. The next question is to how to calculate probability.

3. The relative frequency and probability

Example 1: The probability of obtaining the head in one toss. Your guess for obtaining the head (or tail) in one toss is 0.5. Why? Common sense? Some one told you? Or …

Can you prove it?

Here may be one way to prove it. Let's use

H for the event that we got a head;

T for the event that we got a tail;

and we do the random experiment 1, 2, 5, 10, 100, …, times and record the result.

Relative Frequency

Who did	n	n_H	$f_n(H)$
We	1		
We	2		
We	5		
We	10		
Me and my son	100	46	0.4600
Me and my son	1000	488	0.4880
Pofeng	4040	2048	0.5069
K. Peterson	12000	6019	0.5016
K. Peterson	24000	12012	**0.5005**

We then can calculate the **relative frequency for H** by

$$f_n(H) = \frac{n_H}{n}$$

where

n = the number of tossing or experiment;

n_H = the number of times that we got the head;

We can do this experiment many more times. Here are some of the experimental results:

Probability

It is seen that as n increases, the relative frequency approaches a constant (0.5 in this case), i.e.,

$$\lim_{n\to\infty} f_n(H) = 0.5$$

We define the probability of event H as

$$P(H) = \lim_{n\to\infty} f_n(H)$$

You may say: I don’t have to do the experiment. I just know it is 0.5.

But how do you know?

Here is how. **We account the number of all the possible outcomes (n) and we then assume that each outcome has an equal chance to occur. The probability of any one of these events to occur is 1/n**.

Examples

Example 1: The probability of obtaining the head in one toss.

$$S=\{H, T\}$$

Since n=2,

$$P(H) = 1/2 = 0.5 \quad \text{and} \quad P(T) = 1/2 = 0.5$$

Example 2: What is the probability that you get “3” when you roll a die?

You probably know the answer. But before we give the answer, let’s use X for the random number you may get by rolling a die and we know X = x and x = 1,2,3,4,5, or 6. We call X a random variable because the value of X varies randomly with the experiment. We can’t determine the value of X before we roll the die and we can only get its probability.

In this example, n = 6 and thus

$$P(X=3) = 1/6$$

Classical Probability

The probability of an event occurring is the number in the event divided by the number in the sample space. Again, this is only true when the events are **equally** likely. This kind of probability is called classical probability.

Probability Rules

There are two rules which are very important.

Rule 1: All probabilities are between 0 and 1 inclusive.

$$0 \le P(E) \le 1$$

Rule 2: The sum of all the probabilities in the sample space is 1.

Other rules which are also important.

- The probability of an event which cannot occur is 0.
- The probability of any event which is not in the sample space is zero.
- The probability of an event which must occur is 1.
- The probability of an event not occurring is one minus the probability of it occurring, i.e., $P(E') = 1 - P(E)$

4. Random number and its probability distribution

- **Discrete Random Variables**
- **Continuous Random variables**
- **Probability Distribution**
- **Normal Distribution**
- **Are Data Normally Distributed?**
- **Log Normal Distribution**

A Random Variable (X)

is a variable **whose value varies randomly** or a variable that takes a value in the sample space based on certain probability.

Discrete Random Variable

Example: Flipping a fair coin and observing if we get a head or tail.

P(X=1) = 0.5 and **P(X=0) = 0.5**

Note: We use 1 for H and 0 for T.

S = {1, 0}

Continuous Random Variable

Example: The weight (w) of a bird captured by a biologist, P(w):

$$S = \{0 < w < 10\}$$ (continuous)

How can we know or calculate the probability that the weight of a bird captured by a biologist is equal to 5 lb, i.e.,

$$P(W = 5) = ?$$

We need a probability distribution to answer above question.

Probability Distribution

A listing of all the values the random variable can assume with their corresponding probabilities makes a **probability distribution.**

Example: In an introductory statistics class of 50 students where are

Freshman:	11
Sophomore:	19
Junior:	14
Senior:	6

The probability distribution of X is given by

$P(X=1) = 11/50$

$P(X=2) = 19/50$

$P(X=3) = 14/50$

$P(X=4) = 6/50$

One student is selected at random. Let the random variable X equal 1, 2, 3, or 4 if a freshman, sophomore, junior, or senior is selected, respectively.

More example

Example: Flipping a fair coin and observing if we get a head or tail.

P(X=1) = 0.5

P(X=0) = 0.5

Example: Rolling of a single fair die.

x	1	2	3	4	5	6	**sum**
p(x)	1/6	1/6	1/6	1/6	1/6	1/6	**6/6=1**

More example…

- **Rolling two dice and count the sum.**

  The sums are {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}. However, each of these aren't equally likely. The only way to get a sum 2 is to roll a 1 on both dice, but you can get a sum of 4 by rolling a 1&3, 2&2, or 3&1. The following table illustrates a better **sample space** for the sum obtain when rolling two dice.

Sample space ?

More example…

Rolling two dice and count the sum.

The sums are {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}. However, each of these aren't equally likely. The only way to get a sum 2 is to roll a 1 on both dice, but you can get a sum of 4 by rolling a 1&3, 2&2, or 3&1. The following table illustrates a better **sample space** for the sum obtain when rolling two dice.

First			**Second Die**			
Die	**1**	**2**	**3**	**4**	**5**	**6**
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Probability Distribution of Rolling Two Dice

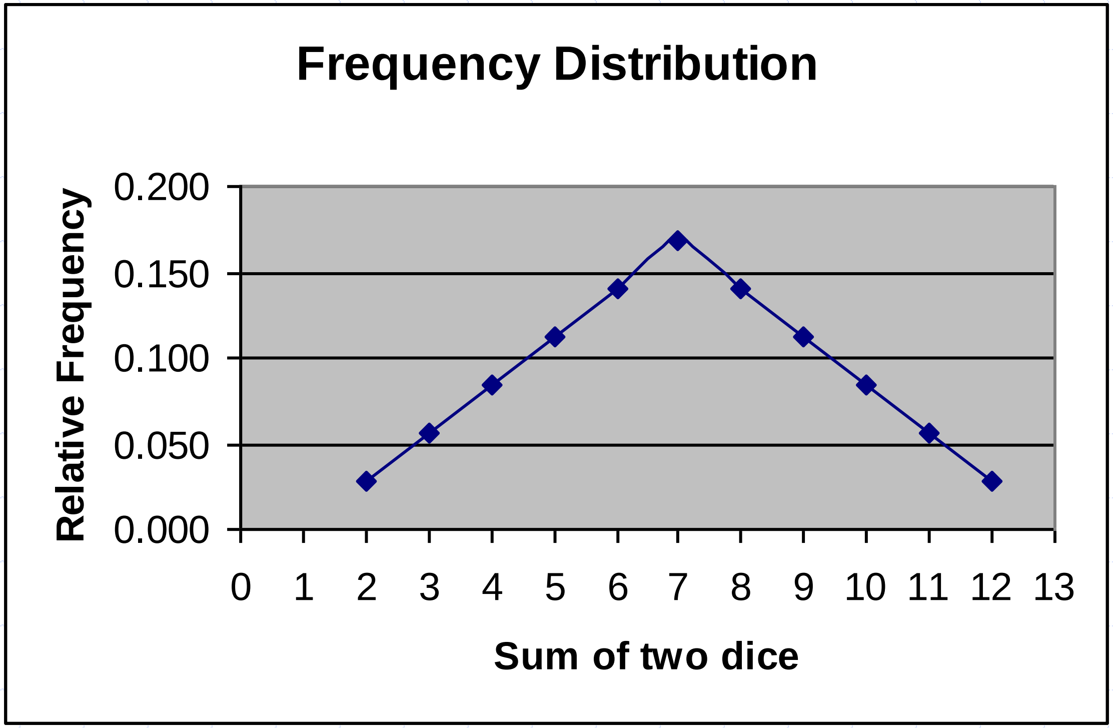
?

Probability Distribution of Rolling Two Dice

Sum	Frequency	Relative Frequency
2	1	1/36
3	2	2/36
4	3	3/36
5	4	4/36
6	5	5/36
7	**6**	**6/36**
8	5	5/36
9	4	4/36
10	3	3/36
11	2	2/36
12	1	1/36

Frequency distribution

If just the first and last columns were written, we would have a probability distribution. The relative frequency of a frequency distribution is the probability of the event occurring. This is only true, however, if the events are equally likely.



The Normal (Gaussian) Distribution

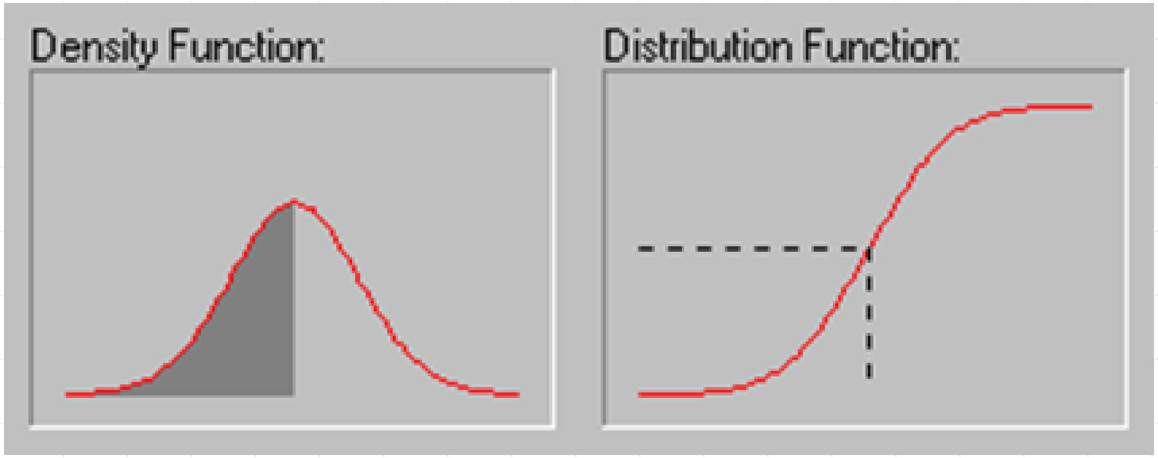
The probability **density function** of the normal distribution

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

The cumulative probability or **distribution function** is given by

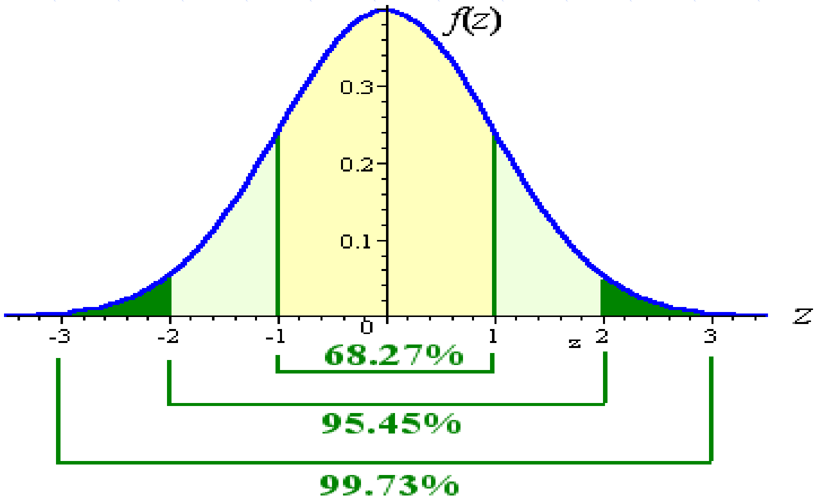
$$F(x) = \int_{-\infty}^{x} f(x) dx$$

where x is a random variable, μ is the mean, and σ is the standard deviation.



Characteristics of a Normal Distribution

- Bell-shaped
- Symmetric about mean
- Continuous
- Never touches the x-axis
- Total area under curve is 1.00



Area under a Normal Distribution

- Approximately **68.27%** lies within 1 standard deviation of the mean, **95.45%** within 2 standard deviations, and **99.73%** within 3 standard deviations of the mean.

Standard Normal Distribution

Same as a normal distribution, but also ...

- Mean is zero
- Variance is one
- Standard Deviation is one.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad \Longrightarrow \quad f(x) = \frac{1}{\sqrt{2\pi}} exp\left[-\frac{x^2}{2}\right]$$

Why the "Normal distribution" is important?

- The "Normal distribution" is important because in most cases, it well approximates the normal density function.
- The distribution of many test statistics is normal or follows some form that can be derived from the normal distribution.
- In this sense, philosophically speaking, the Normal Distribution represents one of the empirically verified elementary "truths about the general nature of reality," and its status can be compared to the one of fundamental laws of natural sciences.
- The exact shape of the normal distribution (the characteristic "bell curve") is defined by a function which has only two parameters: mean and standard deviation (how nice!).
- A characteristic property of the Normal distribution is that 68% of all of its observations fall within a range of one standard deviation from the mean, and a range of 2 standard deviations includes 95% of the scores.
- In other words, in a normal distribution, observations that have a standardized value of less than -2 or more than +2 have a relative frequency of 5% or less. (Standardized value means that a value is expressed in terms of its difference from the mean, divided by the standard deviation.)

Are Data Normally Distributed?

This is a question asked most often. How to determine?

?

Are Data Normally Distributed?

This is a question asked most often. How to determine?

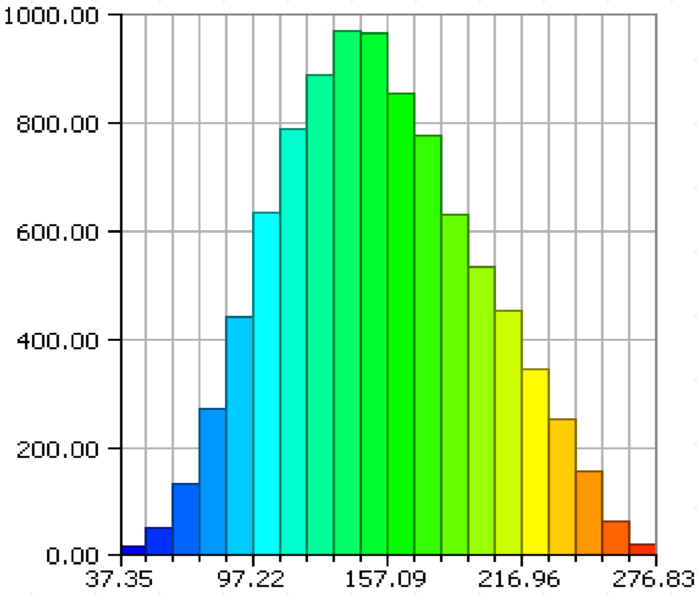
There are at least **two approaches**:

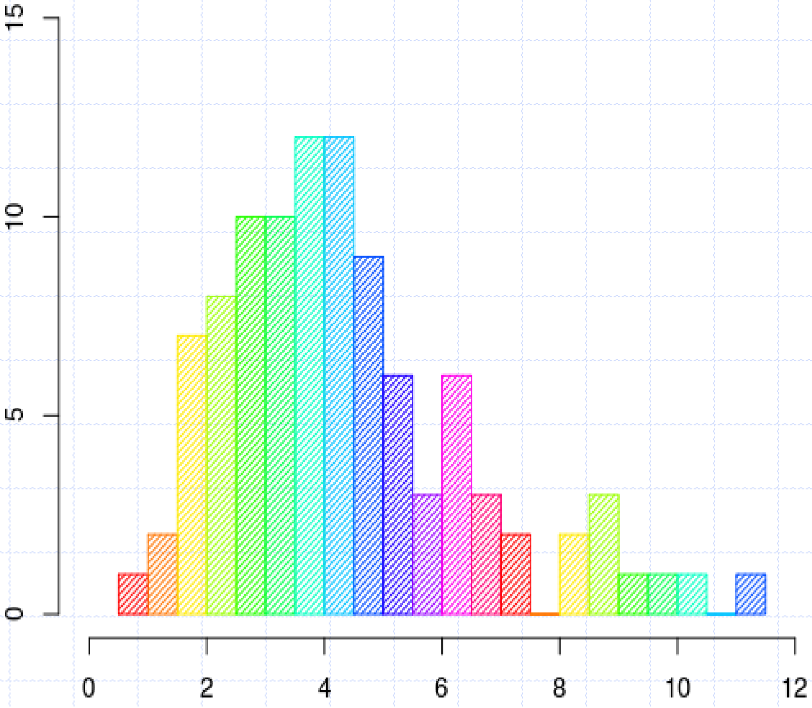
1) To plot a **histogram** of the data and compare with a Normal distribution;

2) To plot **the normal probability plot** or the cumulative curve on a special type of graph paper called probability paper which has an ordinate scale that converts a Normal cumulative curve from a sigmoid curve to a straight line. We can then estimate by eye whether or not the data show a close approximation to a Normal distribution.;

Are Data Normally Distributed?

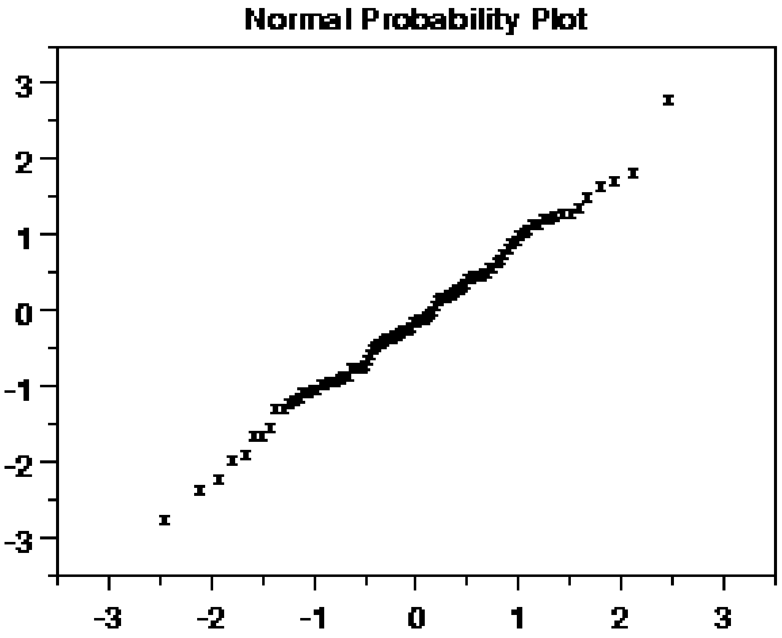
1) To plot a histogram of the data and compare with a Normal distribution;





Are Data Normally Distributed?

2) **The normal probability plot** (Chambers et al., 1983)



The normal probability plot is a graphical technique for assessing whether or not a data set is approximately normally distributed.

The data are plotted against a theoretical normal distribution in such a way that the points should form an approximate straight line.

Departures from this straight line indicate departures from normality.

Log-Normal Distribution

If *Y* = *log X* and Y has the Normal Distribution:

$$f(y) = \frac{1}{\sigma_Y \sqrt{2\pi}} exp\left[-\frac{(y-\mu_Y)^2}{2\sigma_Y^2} \right]$$

Then, we say that X is **log-normally distributed**.

Example: Permeability of soils and rocks.

Statistics of Hydraulic Conductivity

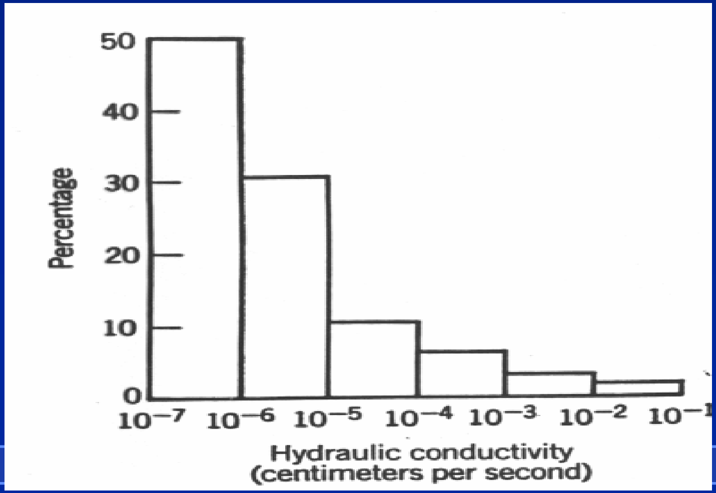
K is *log-normally* distributed

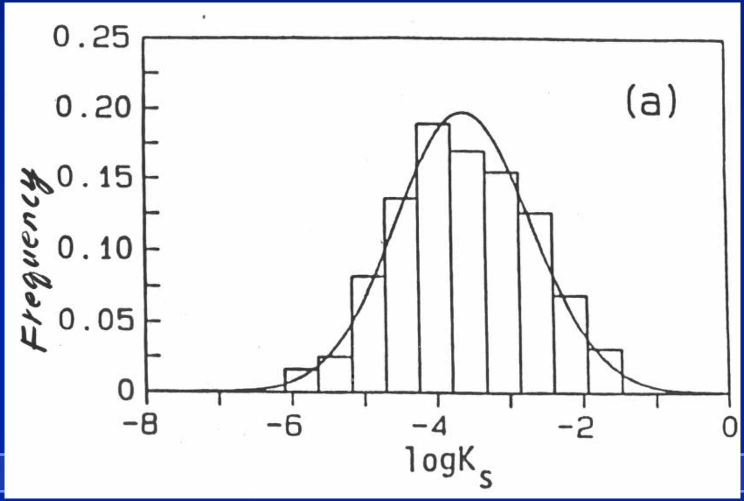
Y = log K is *normally* distributed.

Mean and Variance of *Y* (σ^2_Y)

Autocorrelation Function (ρ_Y)

Correlation length (λ)





Thanks !