

# Sustainability Through Green Manufacturing System: An Applied Approach

Prof. Deepu Philip

Department of Industrial & Management Engineering

Indian Institute of Technology, Kanpur

Dr. Amandeep Singh Oberoi

National Institute of Technology, Jalandhar

## Lecture – 06

### Basic Statistical Concepts for Sustainable Manufacturing Analysis

Good morning, welcome to today's lecture on sustainable manufacturing. And we are going to study today basic statistical concepts that are required for conducting analysis in related to sustainable manufacturing. We, already seen the hierarchical structure or sustainable manufacturing, where we will look at from the factory view, to the production line view, to the unit manufacturing processes and the need of analyzing them other process. And we have also seen the usage of the importance of simulation and its usages and important tool for doing analysis of sustainable manufacturing systems. Since, many of you are not from the probable does not have the probability in view point or exposure to that course. Today, we will kind of cover the very basic concepts of it in a very simplified manner; and it might help you in understanding some of the concepts are represented in this course

(Refer Slide Time: 01:15)

The slide is titled 'AGENDA' and contains a list of topics with handwritten notes in red ink. The topics are:

- Random Variables — Study system state
- Independence & Randomness — Simulation analysis
- Covariance Stationary Process — Steady state analysis, transient analysis
- Law of Large Numbers — replications & averaged estimates
- Probability Distributions & Means — model various manufacturing process
  - processing times
  - setup times
  - down times
- Test of Independence

Handwritten notes at the bottom left:

- (1) Probability and Statistics for Engineers & Scientists. (Jay L. Devore)
- (2) Applied probability and statistics ... (John Miller et al)

The date 8/24/2017 is written in the bottom right corner.

So, the major agenda of today's discussion would be the random variables and why it is necessary, and this is required to we basically uses to study system state, especially in

this case we are going to study this is state of this sustainable manufacturing system. We need to study the independent and randomness for simulation analysis. We need to study the covariance stationary process that the importance of it for what we call as the steady state analysis and also the transient analysis. We study the law of large numbers for what we call as the replications and averaged estimates, this concept is important here.

Then we look at the probability distributions and mean to model various manufacturing process. So, in this, we model the process times or processing times as we call it setup times and we also talk about it as the down times, there are many aspects of this region model as part of this. And we also look into the test of independence as well as part of this. So, for detailed information about this I would recommend you two text books one is Probability and Statistics for Engineers and Scientists, the author is Jay L Devore. And the second book is we would call it is Applied Probability and Statistics for something engineers and scientist by John Miller et al; the multiple authors, typically known as Miller and John text book. So, with this let us get into today's concepts and let us see whether we can understand basics aspects of it.

(Refer Slide Time: 04:28)

3

## BASIC TERMINOLOGIES

- Experiment
  - Process whose outcome is not known with certainty  
*There is uncertainty associated with the outcome.*
- Sample space
  - Collection of all possible outcomes of experiments
 

$S = \{1, 2, 3, 4, 5, 6\}$   
*rolling a die*

$S = \{H, T\}$   
*tossing a coin*
- Sample points
  - Outcomes are sample points in sample space
 

$S = \{1, 2, 3, 4, 5, 6\}$   
 $S = \{H1, H2, H3, H4, H5, H6, T1, T2, T3, T4, T5, T6\}$   
*Size has increased.*

Tossing a fair coin  
Head or Tail

Rolling a fair die  
face values = 1, 2, 3, 4, 5, and 6

Tossed a fair coin  
and rolled a die.

8/24/2017

So, the first concept we are going to do today, the basic terminologies that we need be familiarized with this course which will be used repeatedly throughout this process will be the concepts of experiment. The first one will be experiment, then we will talk about this is sample space, then we talked about this is sample points. So, the experiment is

typically it is a process whose outcomes are not known with any certainty; and means in simplest way there is uncertainty associated with the outcome.

So, a simple example of this is tossing a coin. So, in a fair coin you have you make a toss, you can either get a head or a tail. So, if you make a toss now, you do not know whether it would be head or a tail, it can be head, it can be tail. So, the outcome is not known with quite a large is not known with certainty, you cannot really say it is going to be head, it can say it is head, it could be head or tail. So, experiment these any process like that. So, this is an example of an experiment.

The another example would be rolling a fair die. When you do this, again the thing is you could get any of the values from face values of 1, 2, 3, 4, 5, and 6 are the possible values that you can get out of this you can get anyway any of those faces. You are not sure which value you are going to get. Next concept that we need to go through is a sample space concept the concept of sample face, which by definition is the collection of all possible outcomes of the experiment. So, as we said earlier if the experiment is rolling a fair die then all possible outcomes, if you make write it as a set then the sample space is denoted by capital S or uppercase S, the set would be 1, 2, 3, 4, 5, and 6, this would be when you are a rolling a die.

Similarly, if you are tossing a coin this set would be head and tail this is tossing a coin. So, sample space is the set the collection of all possible outcomes of the experiment is when X plus does a set is what you call as the sample space. Then we talk about sample points, so all the outcomes are the sample points in the sample space, so you rolling a fair die and you get the face value is one. So, this is a sample point, this is another sample point. So, when you say S equal to 1, 2, 3, 4, 5 and 6 for rolling if I die then 1, 2, 3, 4, 5 and 6 are the sample points.

So, I hope you gets understand that we can also create and experiment where you can say that I am rolling a die and tossing a coin at the same time. So, then you can have a head or a tail and along with that any value that is possible. So, we can have if you write both of them at the same time then the sample space would be you can have head. And then one on the die, you can have head and two on the die, you can have head and three on the die, head and four on the die head and five on the die head and six on the die. When you have tail and one on the die, tail which is on the coin and two one the die, tail and three

on the die, tail and four on the die, tail and five on the die, tail and six on the die these are the possible options. So, you can see that as the experiment gets complicated. So, here what we did is we tossed a coin, tossed a fair coin, coin and rolled a die. So, we can see this is the sample space, the size is increased. Similarly, as the experiments get complicated your sample space keeps on increasing.

(Refer Slide Time: 09:10)

**4**

*Rolling a die*  
 $P(Y=2) = 1/6$   
 $Y = 1, 2, 3, 4, 5, 6$   
 $P(Y=y)$   
*RV value*

## RANDOM VARIABLES

- Random Variable is a function that assigns a real number to each point in sample space
- Probability density function or mass function
  - Discrete  
*It represents the values the RV can take along with its respective probabilities.*  

x	1	2	3	4	5	6
$P(X=x)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

*discrete*
  - Continuous  
*pdf*
- Cumulative density function or distribution function
  - Discrete  
 $P(X \leq t) = \{1, 2, 3, 4\}$   
 $= P(X=1) + P(X=2) + P(X=3) + P(X=4)$   
 $= 1/6 + 1/6 + 1/6 + 1/6 = 4/6$   
*accumulated probability*
  - Continuous  
 $P(X \leq x) = \int_{-\infty}^x f(x) dx$   
*cdf*

*Tossing a fair coin*  
 $S = \{H, T\}$   
 Head  $\Rightarrow$  win = 1  
 Tail  $\Rightarrow$  lose = 0  
 $P(X=1) = 0.5$   
*RV value*

*tossing a coin*  
 $P(X=2) = 0.5$   
*RV value*

8/24/2017

So, then we talk about the concept of random variable we did kind of cover this in the yesterday's lecture, but more formally random variable we define as it is a function random variable is a function that is the first important thing that you need to know. It is a function that assigns a real number to each point in the sample space. So, it is a function that assigns a real number to each point in the sample space is what a random variable is or all about. So, if you are talking about the problem of rolling a tossing a coin let us take this again tossing a fair coin the sample space is equal to head or tail.

So, now, what does random variable do random variable assigns real number to each point in the sample space. So, if I say that I get a head, we said as win which is equal to one; and tail equal to applies it is a loose and that is equal to 0. Then the random variable of getting ahead can be described us probability of capital X, X is the tossing coin equal to 1, which means the probability that I when I toss a coin fair coin and get the head and I say this 0.5. So, this is what this function that maps the random variable which maps

the outcome to a real value that is what we call as. So, in this case the random variable is the  $X$ . So, this is the random variable.

So, random variable allows you to map the outcomes of an experiment to a real value in the sample space. So, I can also say see on the same line I can also say that probability of  $X$  equal to 0 equal to 0.5 which means if I toss a fair coin and like that a tail and the probabilities also 0.05. So, this is again this is this random variable is denotes the experiment of tossing a coin. Now, so similarly some people expressed it in different way. So, if you talk about rolling a die, so if we talk about the rolling a die and I say let be the random variable be  $Y$  and I get a face value of 3, so then this probably equal to 1 over 6. So, the values this random variable  $Y$  can take are equal to 1, 2, 3, 4, 5 and six. So, we typically represent this as probability of big  $Y$  equal to little  $y$ , or here you will write it as probability of big  $X$  capital  $X$  equal to little  $x$ . So, this is the value, this is the random variable; this is the value, this is the random variable. So, you know that now random variable is a function that assigns a real number to each point in the sample space.

Now when you have the entire random variable and the associated probability along with it when you describe that then that becomes a probability density or probability mass function or multiple lens people use a probability density, probability distribution, probability mass function many names for it. But at the end of the day they are all pretty much the same, it represents the values the random variable can take along with its respective probabilities. So, if I do like the probability density function of rolling inside a die. So, I would say like this  $x$ , so this is the value this is the value that random variable can take and that will be 1, 2, 3, 4, 5 and 6. I will write as probability of big  $X$  equal to little  $x$ .

So, this is the probability that the random variable take any of these values and I write those probabilities as 1 over 6 this becomes the probability density or probability mass function distribution function whatever we want to call it. It typically known as the PDF. This tells you what are the values random variable, you can take and the associated probabilities with that. So, this is a case of a discrete. So, this is discrete. Why is it called discrete, because the out outcomes are separate countably definite that can also be a continuous process in which the values of random variable can change continuously.

A classic example of this would be a scenario where the random variable, here is your numerical value  $x$  and here is the probability  $x$  equal to  $x$  and you can have the random variable take any value between  $a$  and  $b$ . So, here this kind of a system is called a what you call say uniform distribution, but here the random variable could take any value in between this with equal probability, so that is again  $a$ . So, this case random variable is continuously distributed between  $a$  and  $b$ . So, any value between  $a$  and  $b$ , it can take.

Now the cumulative density function or distribution function which is typically known as the CDF, this is another aspect the we get CDF from the PDF with there again there both can be discrete or continuous. So, if you want to do probability of a CDF for the rolling if a die when I say probability of  $x$  less than or equal to 4 what that means, is what is the probability that I get an outcomes of 1, 2, 3 and 4 that is what it means. So, how do I get that that is equal to probability of big  $X$  equal to 1 plus probability of big  $X$  equal to 2 plus probability of big  $X$  equal to 3 plus probability of big  $X$  equal to 4, which is equal to  $\frac{1}{6}$  plus  $\frac{1}{6}$  plus  $\frac{1}{6}$  plus  $\frac{1}{6}$  equal to  $\frac{4}{6}$ . So, this is the probability that if I roll a fair die and I get any value less than or equal to 4, so that cumulated probability, this is the cumulated probability.

So, the cumulated probability of this happening any of the face less than or equal to four happening is. So, this is a discrete case that is the CDF cumulated density function. In the case of a continuous distribution, what you will be doing will be, so here in this case in a continuous the PDF can be typically represented as integral. So, there you will write it as  $\int_a^b f(x) dx$  this  $f(x)$  will be a function. So, in this case, it will be like if you are going to go to a particular value, you say probability of  $X$  less than or equal to little  $x$  you will be integral of you know minus infinity to  $x$   $f(x) dx$ . So, integrate from negative all the way up to  $x$ .

So, here in this case if you are looking at that discrete uniform distribution, again if you take a look into this and you say that this is  $a$ , this is  $b$ , and somewhere here is  $x$ . So, then this is the total area that you are talking about what is a probability that the random variable can take any value between this is  $x$ ,  $x$  to  $a$  that is what you are looking at. So, that is the cumulative probability function. I would request you to read it more I am trying to give it in a much simpler version, so that everybody can have an fair idea, but please read the text book for accurate information and as well as getting the proper terminologies and as well as looking at the solid solutions or examples.

(Refer Slide Time: 18:29)

5

$P(X=x) = \frac{1}{6}$

$E(X) = \sum x \cdot P(X=x)$

$= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6}$

$= \frac{1+2+3+4+5+6}{6}$

$= \frac{21}{6}$

$= 3.5$

$\bar{x} = \frac{\sum x_i}{n}$

$\frac{2+3+3+10}{3}$

$\frac{20}{3}$

Sample mean

## MEAN & VARIANCE

*Measures on the behavior of the data*

*if you repeat the experiment infinitely (long time) what will the value converge to?*

- Mean or Expected value of a random variable
- Mean provides the central tendency *(converging)* *What is that value which is the centroid of the data?*
- Variance of a random variable measures its dispersion (or deviation)  $(\mu - x)$
- Variance is always greater than or equal to zero
- Standard deviation  $\sigma_x = \sqrt{\text{Var}}$  *Sample Unit of variance is not the same as data.*
- Has the same unit as that of the data

$\text{Var} = \frac{(6-3)^2 + (6-3)^2 + (6-10)^2}{n-1} = \frac{10}{2} = 5$

$\text{Var}(x) = \sigma_x^2 = S^2$

8/24/2017

So, then we get to the concept of mean and variance. Mean and variance are two important aspects, which actually used to measure certain behavior of the data. So, these are measures on the behavior of the data the most important thing is the mean typically what we call as it or it is an expected value of a random variable. So, when you say that when you talk about a probability scenario, where you are rolling a die and say probability getting little x, you said and you this is your PDF and you see in that the values of x can vary from 1, 2, 3, 4, 5 and 5 and this probabilities can be 1 over 6, 1 over 6, 1 over 6, 1 over 6, 1 over 6, 1 over 6, then the mean or the expected value of the random variable is all about in the long term what you will get to see.

So, this is also what we call as if you repeat the experiment infinitely, it is not infinite it mean likes long time if you are doing this for a long time, what will the value converge to. So, you roll a die get a value not it down here all another take a value not in down keep on repeating this. And you take those values and you average them. What would that value be that is called as the expected value. So, expected value can be always found out easily. The typical formula of expected value it is not read it by E of x sometimes it is also known as mu, mu of x this is a notation for mean or expected value. It is given by x times p of x or this p of x is equal to probability of x equal to little x multiplied by those probabilities you will get the value.

So, in this case, it will be equal to 1 times 1 over 6 plus 2 times 1 over 6 plus 3 times 1 over 6 plus 4 times 1 over 6 plus 5 times 1 over 6 plus 6 times 1 over 6. And if you sum all of this, you will probably get a value close to 3.5 I believe. So that means, in the long term you keep on rolling a die, take the values, not them down on a piece of paper, sum them up divide by the number of observations you have. When I am talking about let us say roll this die and do it up to may be let say thousand times or something like that you will see the values slowly converging to 3.5, so that is the long time average or expected behavior of the system. So, or that is what the expected behavior of the random variable. So, that is what the concept of expectation is at the same applies for continuous distribution as well.

So, what does the mean says or what does the mean provides information about it provides what information which is called as a central tendency. Or what is the central tendency is in a way what is that value which is the centroid of the data or what is the central point of the data which shows the central tendency of the data on the long time that is what this date the data values should be converging to. So, this also kind of says converging, where will this value converge to. So, many a times the mean we all know the statistical it is called as  $\bar{x}$  which we always says as  $\frac{\sum x}{n}$ . So, if you have values of  $x$  is equal to 7, 3, 10 then  $\bar{x}$  is equal to 7 plus 3 plus 10 divided by  $n$ ,  $n$  is number of variable number of observations which is 3 that is 20 by 3. So, that will be the  $\bar{x}$  with or which is the mean which is the sample mean in this case. So, there is a difference between sample mean, population mean which we can read from the example, but here we are talking about pretty much this expected value as the case.

The other hand you also you measure the central tendency and the next thing you measure is the dispersion or deviation. Two ways we can measure it, one is by using the concept called variance. The variance of a random variable measures is dispersion. So, what really measures is it measures the deviation from the expected value. So, in a way what you trying to do is if you have the  $\mu$  and minus  $x$  is what you trying to measure. So, if you know what this expected value and this is  $x$  that is in the measure from the individual. So, then variance typically used square of this because you know sorry actually not that case in the case of sample variance.

So, the  $\mu$  minus  $x$  in the way what you are trying to say is how far is if you look at a set of data like this, and let us say this is the sample mean then these deviations are what you



are looking at something like this, can we shall we think about that in that way. So, important thing is variance is always greater than or equal to 0; it is never a negative value because you end up squaring it as I said earlier. If you just sum the deviations then it can all sum to 0 that is one of the reasons we square because there will be a positive and negative deviations. So, when you square it actually it becomes nonzero positive values and then (Refer Time: 25:03) possible value you can get is 0.

Whereas the problem with the variance is the unit of variance is not as same as data. So, let us say if you assume these values to be the unit will be degree Celsius then the variance of this will be given by it can be calculated typically by like 20 by 3 that will be about approximately 6.8 something like that. 6.8 it will be  $6.8 \text{ minus } 7 \text{ square plus } 6.8 \text{ minus } 3 \text{ square plus } 6.8 \text{ minus } 10 \text{ whole square minus } n \text{ minus } 1$  that will give you the variance which is  $n \text{ minus } 1$  will be  $3 \text{ minus } 1$ . So, in this case you the unit of this variants will be degree Celsius square this is squaring the whole thing this cannot be compared with the data value.

So, hence to find the find a deviation which is in the same unit as that of a data value we do what we call as standard deviation. So, standard deviation typically denoted by sigma the by the way variance is typically do not by either variance of  $x$  or  $\sigma x$  when it comes to sample we basically say  $s^2$ , so that is the notation that is used which basically says it is a variance anyway. Standard deviation on the other hand is sorry it is not sigma, sigma square. Standard deviation is sigma of  $x$  or  $s$  denote this is a sample standard deviation, this is the population. So, this is given by square root of variance that is what a standard deviation is. So, that you get the unit has the unit same as that of the data, you get the unit back of, back of the data

(Refer Slide Time: 27:15)

The slide is titled "INDEPENDENCE" in green. It contains several bullet points and handwritten notes in red and black ink. The first bullet point states: "Two events are independent if the occurrence of one event does not affect the probability of occurrence of the other". A handwritten note next to it says: "Independence Concept is important to Conduct Simulation Analysis". The second bullet point states: "Independent events must be able to occur at the same time". A handwritten note next to it says: "When output data of Simulation is analysed appropriate can be created". The third bullet point states: "Disjoint events cannot be independent". A handwritten note next to it says: "Disjoint events cannot happen at the same time". The fourth bullet point states: "Not disjoint events may or may not be independent". A handwritten note next to it says: "Not Disjoint events may or may not be". There are three diagrams: 1. A Venn diagram with two overlapping circles labeled A and B. 2. A flowchart showing a box labeled "Milling" with an arrow pointing to a box labeled "Drilling". 3. A flowchart showing a box labeled "Drilling" with an arrow pointing to a box labeled "Milling".

6

## INDEPENDENCE

Two events are independent if the occurrence of one event does not affect the probability of occurrence of the other

Independence Concept is important to Conduct Simulation Analysis

$P(A/B) = P(B) \Rightarrow$  Bayes theorem

Independent events must be able to occur at the same time

When output data of Simulation is analysed appropriate can be created

Disjoint events cannot be independent

Disjoint events cannot happen at the same time

Not disjoint events may or may not be independent

Not Disjoint events may or may not be

Two events are happening and one event happening is not decided by the happening of other. They are independent events.

Drilling can only happen for this part. Drilling can only happen after milling (because of precedence).

Milling can only happen for this part. Drilling can only happen after milling (because of precedence).

8/24/2017

So, moving on I hope I request you get to go ahead and read this all concepts of variance expectations, expected value,  $\mu$ , population mean, sample mean, sample standard deviation, sample variance from the textbooks. This is just overview of what multiple chapters of the textbook covers, but hopefully gives you an idea and motivation to read the whole thing. Now, we are going to study the concept of independence. So, the independence is a tricky concept because independence concept is important to conduct simulation analysis, this you guess need to remember why you need to learn the concept of independence. So, we can say two events are independent; if the occurrence of one event does not affect the probability of occurrence of the other. So, you have two events are happening. So, first thing is two events are happening and one event happening is not decided by the happening of other or the occurrence of other.

So, when you have such a two events those events are said to be, they are independent events; mathematically it is given the probability of A given B. So, that probability of A said that the problem is that B has occurred is basically you can this you can think about in the case of Bayes theorem. So, the independent events must be able to occur at the same time that is the most important aspect that you need to would say. So, if you have a manufacturing system in which there is a milling there is going on and drilling there is going on. They are connected with each other up. Let us say part is coming here one part is here this. Drilling of this part can only happen for this part drilling can only happen after milling why because of precedence for those this particular part. But if you have

another part two different part this is part two, this part one. For part two drilling can happen, if milling is happening for part one, assuming that it is over.

So, what I am trying to say is that, but if you have the factory like this, we draw it here milling and drilling. And parts come and go to a milling, parts come and go to a drilling then you can say that the occurrence of milling does not influence occurrence of drilling. So, then you can think about it as an independent event, wherein this case we cannot do anything about us and independently. Then there is another event, which is called about disjoint events. So, please read that in the textbook disjoint events cannot be independent; disjoint events the region they cannot be independence because they cannot happen at the same time.

So, an example of this is when you toss a coin, you can get a head or you can get a tail you cannot get a head and tail in the same time, so that disjoint so that means, there is. So, if you think about a disjoint events and if you think about it the probability of event A and probably event B, there is no connection between them, they are not probability of intersection is actually equal to 0. So, this is A and this B. So, disjoint even been they cannot happened with the same time, there is no probability of now happening at the same time; not disjoint events may or may not be independent, remember this, this is important.

The disjoint events, they cannot be independent. If the event, so that does not mean that not disjoint events; if the event is not disjoint that does not mean there independent, they are may or may not be not disjoint events may or may not be independent that is the other important aspect. Anyway, the concept of independence is important for simulation analysis is purely because we analyze systems like this, we analyze systems like this. So, for such systems we would really like to you know study the concept of independence or at least understand the concept of independence, so that when output data of simulation is analyzed appropriateness can be a created.

(Refer Slide Time: 33:38)

7

*Probability =  $\frac{\text{Outcomes favorable}}{\text{Total outcomes}}$*

*(Red Color Card) & Spade cannot happen simultaneously.*

*Ace of diamonds*

## ILLUSTRATIVE EXAMPLES

*Disjoint events are not independent = they cannot occur at same time.*

Experiment

- Draw a card from deck and bet on whether the color is red
- Before placing bet, it is told that the card is spade – “red” and “spade” are disjoint events, but that information influence the event bet.
 

*this information that spade is black in color always*
- Before placing bet, it is told that card is ace – “ace” and “red” are not disjoint, but they are independent.
 

*Red and ace can happen at the same time*

$P(\text{Red}) = \frac{26}{52} = \frac{1}{2}$   
 $P(\text{Red/Ace}) = \frac{2}{4} = \frac{1}{2}$

The occurrence of “ace” has no influence on “red”

*Total Cards in a deck = 52  
13 Cards in 4 Suits. Black = 13+13 = 26  
Red = 13+13 = 26*

8/24/2017

So, continuing on here is an illustrative example for what we presented earlier. Let us thing about an experiment. So, here is an experiment for you. So, this is an experiment. Draw a card from a deck and bet on whether the colour is red. So, the experiments is this, you draw a card from the deck and you say whether this red or back. Before placing so, if you think about the solved experiment before placing the bet it is told that the card is spade so that means you know what the colours are. You have in if you draw a card you have what you called as one is called as the clubs, which kind of looks like this excuse my drawing capabilities. Another one is the hearts, another one is called as the diamond and third one is the spade, which kind of looks somewhat like this yeah my drawings are bad.

So, these two are the red guys and these are the black guys. So, before placing the bet you told that the card is spade, spade means red and spade are disjoint events, but so that means, if it is a spade it can only been black it cannot been red. So, that means, red and spade they cannot happen at the same time they are disjoint events so that means, red colour card and is spade these two cannot happen simultaneously. So, the information the you know that disjoint that means, they cannot happen simultaneously and you have to bet whether the card is red or not. So, if somebody says you the card is spade, then you should not bet on red colour, because it is not going to be red, it is going to be black.

But that hence that information influences the bet this information that spade is black in color always influences the bet. But if I tell you that instead of telling is the spade, if I tell you that the card is ace, ace stands is represented in a card as like it will something like this A and A and if you are something like this that means, it is an ace of a diamond. So, this is ace of diamonds. So, if I say that the card is ace then the red is not disjoint because you have an ace for diamond there you have an ace for heart, these two are red; then you have an ace for a clubs, which will be something like this. And you have an ace for a spade, which will look something like this.

So, these two are black, these are red. So, two out of the aces are red to out of the aces are black. So, now, you have that ace it can be black, it can be red; that means, they are not disjoint or in a way red and ace can happen at the same time. So, they are not disjoint, but they are independent. How are the independent because red is independent because the probability of red is 1 over 2 because there are 52 cards in a deck. So, total cards in a deck equal to 52, thirteen13 cards in 5 suits, these are the suits, so that gives you the 52 then so that means, you have 13 plus 13. So, the black is 13 plus 13; red is 13 plus thir1teen. So, this is the clubs and spades club and spades this is the heart and diamonds. So, you have 26 cards that belong to the red colour see you and there after the total 52.

So, the probability is given by probability is equal to outcomes number for outcomes favorable by total out comes. So, in that if you follow that formula the 26 favorable outcomes for getting the card the red card and that total of 52 cards. So, you get the probability of 1 over 2. Probability of red given ace, there are four aces. So, in this case the total number is 4, this is the total outcome which is 4, and there are two out of the four are red aces. So, is two out of four which is equal to one over two. So, the occurrence of ace has no influence on the occurrence of red because it can be one of those, but they are independent events and their not disjoint.

So, if an even this disjoint, so think about this way, disjoint events are not independent why because they cannot occur at the same time. Independence do require the occurrence of the same time and in occurrence of one should not influence the occurrence of the other. So, the occurrence of the ace has no influence on the occurrence of red because ace can be in red or a black. So, the red and ace has the occurrence of an ace does not influence the occurrence of the red card, hence you can say that the occurrence of the ace

and occurrence of the red card are two independent events. But occurrence of a spade and occurrence of a red are not independent, because that disjoint events, because if it is a spade, it cannot ever be red. I hope there kind of makes the concept clear to you guys.

(Refer Slide Time: 41:10)

8

## ANOTHER EXAMPLE

*Not being able to identify the color*

- Colorblindness occur more often in men than in woman (women)
- But a woman can also be colorblind, hence the events are not disjoint
- Given that the probability of woman being colorblind is much lower, sex and colorblindness are not independent either.

*Disjoint = cannot happen at same time.*

*If the occurrence of something do influence the occurrence of another, then they are not independent events.*

Sex	male	female
$P(x)$	0.7	0.4

*$P(\text{color blind in men}) > P(\text{color blind in women})$*

*Sex influences the occurrence of color blindness.*

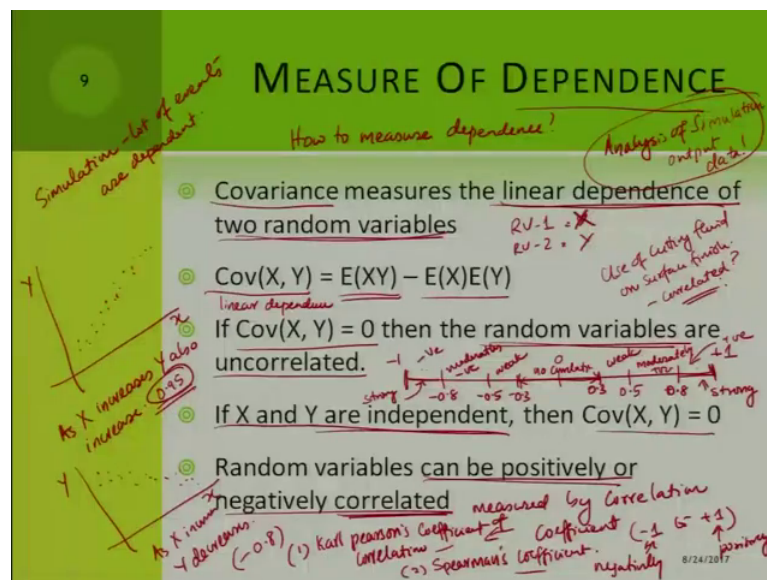
8/24/2017

So, here is another example little bit more realistic example. Let us forget about the cards and let us worry about this. Colorblindness, colorblindness is that you are not able to identify the colour. So, not being able to identify the colour, this is the colorblindness concept. So, colorblindness occur more often in men than in woman as this is not woman women w o m e n the plural I apologies. So, in the way there is best way to say this is men are more prone to colorblindness than women that means a woman can also be colour blind. It says that men or more prone does not mean that women can never be colorblind, she can also be color blind. And if there is a case then the events are not disjoint why because disjoint as remember earlier cannot happen at same time, but women can also be colour might they are not disjoint.

Given that the probability of a women being colorblind is much lower section colorblindness are not independent either, because you already said that if you say sex equal to let us say think about this way male and female or men and women. And if you put the probability of this x then let us say the male probability is if you put the female probability as let us say 0.4 or something like that the probability associated with men will be obviously, more than 0.4 maybe 0.6 or 0.7 or something like that some value.

The idea is that probability of colorblindness in men is greater than probability of colorblindness in women. So, given this information you can say that if you know the sex then the currency of colorblindness the sex and colorblindness are not independent neither because sex do influence, sex influences the occurrence of colorblindness. So, this is another illustrative example of how independence is important. So, if sometimes certain things say in a manufacturing system is something has happened earlier. So, the occurrence of something or if the occurrence or something do influence the occurrence of another then they are not independent events. So, then when you doing analysis, there should not be analyzed us independent events that is the most important aspect that you need to think about.

(Refer Slide Time: 45:03)



So, we talked about independence. So, if it is not independent then obviously, they should be dependent. So, if you are dependent how do you measure the dependence. So, this slide talks about how to measure dependence; in simulation lot of the events are dependent. So, we need to have a mechanism to measure the dependence. And what is the measure of dependents in this case, typically the measure is given by stuff called covariance. So, covariance measures the dependence, the linear dependence of two random variables. If you have two random variables x and y, so random variable one equal to X uppercase X, random variable two equal to Y uppercase Y, then covariance of X and Y. Or in a way the linear dependence of X and Y is given by the expected value of

X and Y expected value of X and Y occurring together minus expected value of X times expected value of Y minus the expected value of these X and Y separately.

So, that means, if variance of  $x$   $y$  equal to 0, then the random variable is un correlated; that means, there is no correlation associated with that. If X and Y are independent then the covariance is also 0, that is also true. So, if covariance X, Y equal to 0, then you can say that they are uncorrelated or other way they can also be independent. They are independent actually. Random variables can be positively or negatively correlated. So, what does that means is if you have a variable X, if you think about the scenario where we have X and Y like this and you draw the graph and you get something like this that means as X increases Y also increases. And you do the covariance and find the correlation behind that or the correlation coefficient you get a value and it is let us say 0.95 or something.

The correlation value usually is measured by correlation coefficient, I am not spending time on how to do correlation coefficient, I request you to study this. So, there are many of them is now Spears and Spearman Rho Karl Pearson's coefficient of correlation etcetera there are there. So, the correlation coefficient typically in an example is you know Karl Pearson's coefficient of correlation in varies between minus 1 to plus 1. So, minus 1 is it is negatively correlated, this is positively correlated.

So, what I am saying here is if you have a 0.95 value; that means, it is closed the 0.1 the is closed to plus 1 so that means, the X increases the value of Y is also increasing that means, there positively correlated. If you look at something like this is an X and Y and you have a data value doing this that means, as X increases Y decreases. So, we might get a value of minus 0.8 that means, X and Y are negatively correlated. That means, ask X increases Y decreases, X shows Y shows a negative linear dependence on X that is measured by this in this particular case also for engineers this correlation coefficient is distributed you think about this here is your minus 1 here is your plus 1 and here is 0. So, typically from if you take 0.8 here and let us say minus 0.8 here these are basically said as strongly correlated area this is a strong correlation which means this is a strong positive correlation this is a positive correlation, here is a negative correlation, but this is also strong correlation. So, this means is strong positive correlation as X increases Y increases, this means a strong negative correlation as X increases Y decreases.



Now, let us take another example of another point of 0.5 and minus 0.5. So, this area is what we called as moderate, this is moderately correlated this is moderately positive this is moderately negative correlation. So, then another value that is possible is 0.3 and minus 0.3, typically we call this as weak correlations, weak positive and weak negative correlations. And this area typically is taken as now correlation. So, you can think about the correlation coefficient, if you measure the correlation coefficient, you can identify whether strong positively correlated or strong negatively correlated that kind of an aspect.

So, it gives you that behavior of one random variable or it measures the linear dependence of two random variables how one random variable is linearly dependent on another. So, if covariance is equal to 0, if you measure covariance and covariance turns out to be 0, you can say that the random variables are not correlated. And if X and Y are independent events, no matter whatever it is, if X and Y are independent then covariance of X and Y should be equal to 0. So, for independent events you will get the covariance of X and Y where the variance of X along with the variance of Y comparing the covariance together variance of that you should be able to get a 0 value for that.

It said about the positive and negatively correlated and the correlation is measured by multiple correlation coefficients. So, I expected you to study at least two, one is the Karl Pearson's coefficient of correlation that is one, the other one is Spearman, Spearman's coefficient people called this as Spearman's rho and those kind of things, but these two are quite important for us for this class. And because these coefficients of correlations are usually measured when you are doing analysis simulation output data. When you doing that when you are analyzing the simulation data, which is a large set of data, you always need to see whether the data is dependent, independent, there is correlation between the data is it positively correlated, is it negatively correlated stuff like that. An example of simple example way down the road, you will actually read we are using a specific type of cutting fluid. So, use of cutting fluid on surface finish, if this correlated not correlated that ways one thing that is typically studied as part of the sustainable manufacturing analysis.

So, I think what we will do today is we will kind of stop here and take the rest of the topic as part of and second few more slides are left out. We will actually take this as a continuation in the next lecture. And with that we will complete the basic concepts or

introductions of statistical concept to it, but please take your time to read and understand this text books and cover up the basics, so that you can understand the futuristic analysis of how to make a manufacturing system sustainable.

Thank you for your patience here.