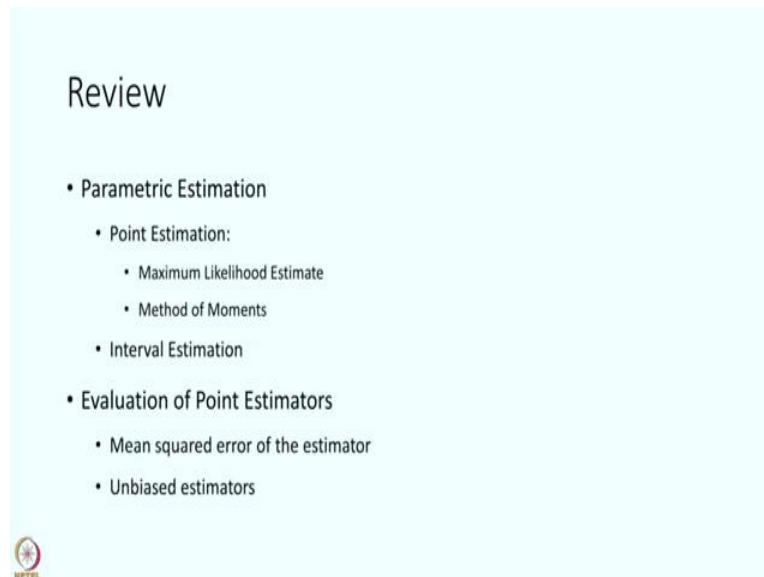


**Dealing with Materials Data**  
**Professor M P Gururanjan**  
**Professor Hina A Gokhale**  
**Department of Metallurgical Engineering and Materials Science**  
**Indian Institute of Technology, Bombay**  
**Lecture 54**  
**Bayesian Estimation**

Hello and welcome back to the course on Dealing with Materials Data. Presently we are going through sessions on estimation problem for the parametric distributions and let us quickly recall.

(Refer Slide Time: 00:35)

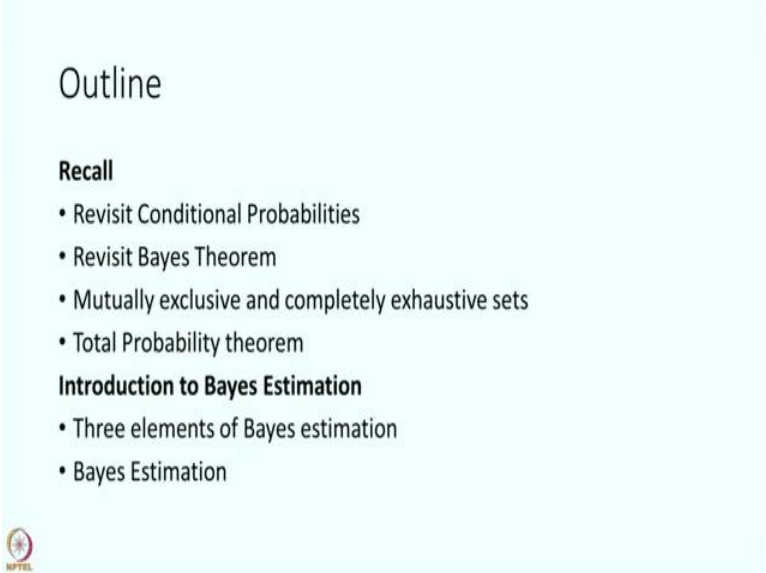


We are going to, we have studied the parametric estimation there are two kinds of parametric estimation. One is a point estimation which for which we talked about maximum likelihood estimation and the method of moments estimation and we also show the weakness of the method of moments estimation.

And we showed that in one particular case of uniform distribution how the results of the method of moments estimate may not give you the correct answer they tend to you know deviate from the basic understanding. So one has to be careful while working with method of moments. We also talked about interval estimation then we talked something about how to evaluate point estimator.

In this case we talked about mean squared error of estimator and what is known as what is called an unbiased estimator.

(Refer Slide Time: 01:42)



The slide is titled "Outline" and is set against a light blue background. It contains two main sections: "Recall" and "Introduction to Bayes Estimation". The "Recall" section lists four items: "Revisit Conditional Probabilities", "Revisit Bayes Theorem", "Mutually exclusive and completely exhaustive sets", and "Total Probability theorem". The "Introduction to Bayes Estimation" section lists two items: "Three elements of Bayes estimation" and "Bayes Estimation". A small circular logo with the letters "NPTEL" is located in the bottom left corner of the slide.

Today what we would like to do is learn about Bayesian estimation and before we do that we would like to revisit the conditional probabilities. We will revisit Bayes theorem we will understand mutually exclusive and completely exhaustive sets. we will once again visit the total probability theorem this is only to recall what we have already done in the previous sessions.

And then we will introduce the base estimate and we will talk about the three basic elements of the base estimation and what is based estimation itself.

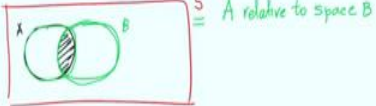
(Refer Slide Time: 02:20)

Conditional probability & Bayes Theorem

- Probability of occurrence of event A given that event B has already occurred:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$
$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

$\therefore P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$



MPTEL

So let us start, Conditional probability and Bayes theorem are related. First let us talk about conditional probability. A probability of occurrence of an event a given that an event B has already occurred is given by this formula. What does it really say? If you look at it very carefully what it really says is that suppose you have a sample space s in this sample space you have a set or an event A and you have an event B.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$\therefore P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

This is an event B, what this says is that the probability of A Given B is probability of this region which is A Intersection B it is a probability of this region with reference to only the region, complete region B, okay. What does this mean? What we are trying to do is generally when we find our probability of A it is with reference to the sample space S.

Here what we are trying to say is you look at the probability of occurrence of A as if your sample space is only restricted to B, okay, So you have to look at the occurrence of event A relative to space B. Now when you do this you get this is the formula for that this formula can also be understood which is simplified in here. This formula can be easily understood because you are taking it relative to space B so you have divided the probability with respect to probability of B.

This further simplifies into the fact that if you just take the probability of B on the other side it says that probability of interaction of A and B is probability of A given B, this is as we said in the past this can be read as probability of A given B multiplied by probability of B and this as well be written as you know this as well be written as probability of B given A will be probability of same A interaction B divided by probability of A.

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

Because now the reference set is A and we are looking at of occurrence of B given that the A has occurred, so instead of looking at relative to s we are going to look into the probability of occurrence of B relative to A and therefore only the portion which occurs along with A can be counted. So that is how this is what becomes and therefore we get this formula.

(Refer Slide Time: 06:19)

• for any events A and B:

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

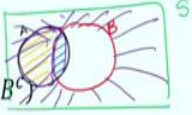
*Handwritten note:*  $\therefore P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$

*Handwritten note:*  $E_1, E_2, \dots, E_n \ni E_i \cap E_j = \emptyset \quad i \neq j$

*Handwritten note:*  $\bigcup_{i=1}^n E_i = S$

Mutually Exclusive events: At the most only one event can occur

Collectively Exhaustive: There are no other event that can occur

$$P(A) = P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + \dots + P(A|E_n)P(E_n)$$


Now, you agree that for any event A probability of A interaction B plus probability B interaction B complement gives you the complete probability.

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

$$\therefore P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

So what I am trying to say is that if you again look at the sample space  $s$  and then you look at set  $A$  and then there is a set  $B$  which is this, this I call  $B$  then what we are looking at is actually  $A$  interaction  $B$  as an  $A$  and rest of the part let us use another colour.

So rest of the part that is this portion is the interaction of  $A$  along with the outer area of  $B$ , so it is actually an interaction with respect to this complete area which is outside  $B$ , I think this is well understood and therefore it is this yellow area which has been shown here, this yellow area is what you get and therefore this area plus this area together makes it a mutually exclusive sets.

And therefore what you have is a probability of  $A$  as probability of  $A$  interaction  $B$  plus probability  $B$  interaction,  $A$  interaction  $B$  complement let us try to generalize it. So we try to generalize it, my spelling is mistake it is there is no  $G$  here. So let us try to generalize it, so we say that there are event  $E_1, E_2, E_n$  such that they are mutually exclusive. So  $E_i$  intersection  $E_j$  is null set for all  $i$  not equal to  $j$  and union of  $i$  is equal to 1 to  $n$   $E_i$  gives you the sample space.

Such an event are called mutually exclusive and collectively exhaustive, these events are called mutually exclusive and collectively exhaustive. So if we have this then looking at this which can be generalized as shown here that probability of  $A$  is equal to probability of  $A$  interaction,  $B$  is probability  $A$  given  $B$  probability  $B$ , plus probability of  $A$  given  $B$  complement probability of  $B$  complement.

We can write here that probability of any event  $a$  in that case can be written as probability of  $A$  given  $E_1$  probability  $E_1$  plus probability of  $A$  given  $E_2$  multiplied by probability of  $E_2$  and so on and so forth, probability of  $A$  given  $E_n$  multiplied by probability of  $E_n$ , this is known as a total probability rule or total probability theorem.

So if we have a mutually exclusive and collectively exhaustive units, sets or events then any event  $A$  can be divided into collection of or a summation of several events which are conditioned events and you have to multiply with a respective probability of the event on which you have conditioned it. So now we come to Bayes theorem, it is very simple.

(Refer Slide Time: 11:32)

Bayes Theorem

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$
$$\therefore P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$P(A|B) \propto P(B|A)$

A relative B :  $P(A|B)$   
B relative A :  $P(B|A)$

The Bayes theorem puts it in a very simple way, from this formula it simply says that probability of A given B is probability of B given A multiplied by probability of A divided by the probability of B. So if you put this in the denominator you get this formula. Let us try to understand what does this formula say, because this has a great importance and a great bearing on what the Bayesian inference that is very important in today's time.

So let us spend a little bit of time, once again let us try to understand through Venn diagram. This is a sample space S, there is a set A, event A and this is an event B. If you are looking at the A interaction B why am I looking at an interaction B, because of this. See this whole Bayes theorem has come out because A interaction probability of A interaction B can be written two types of conditional probabilities.

So here we are saying that we are looking at this event, the A interaction B in two different ways. So on one way we are saying that occurrence of A relative to B is what we call probability of A given B. On the other hand you have a probability occurrence of B relative to A which is what gives probability of B given A. And here you see these are the two things which are being connected here.

Note that these are the two things are connected here that is if you are looking at occurrence of A relative to B or you are looking into occurrence of B relative to A they are sort of proportional or

they are weighted down by a factor. So it is you can very easily say that probability of A given B is some kind of proportional to probability of B given A.

So if you know one, you should be able to find the other is what is being indicated here. First let us try to understand this whole thing how it becomes useful in application by taking a very simple problem.

(Refer Slide Time: 14:42)

### Ice-cream problem

- Near a college, an ice-cream shop sells three flavors of ice cream: Chocolate, Vanilla and Strawberry. It has been observed that both girl students and boy students frequent the ice cream shop equally. It is also observed that 55% of the girls prefer chocolate flavor, 20% buy vanilla flavor and 25% prefer strawberry ice cream. While the boys prefer chocolate, vanilla and strawberry flavors 50%, 30% and 20% of times respectively.

- 1) Find probability that strawberry flavor ice cream will be ordered by a random student visiting the shop.
- 2) Suppose a student comes to the shop and orders strawberry flavor ice cream, what is the probability that the student is a girl?



Let us look at the problem of an ice cream shop. There is ice cream shop that sells only 3 flavors of ice cream for our convenience, chocolate, vanilla and strawberry and since it is near the college only boys and girls tend to visit, the students tend to visit is more often. We have not taken care of the faculty which may be going and eating ice dream dead but we are talking about the students only here.

And the shopkeeper has noticed that 55 percent of the girls prefer chocolate flavor, 20 percent of the of them prefer vanilla flavor and 25 percent of them prefer strawberry ice cream. While among the boy this distribution is slightly different. The boys like chocolate flavor 50 percent of the time, vanilla flavor 30 percent of the time and 20 percent of the time they buy the strawberry flavor.

We are also given the information here that both girls and boys student frequent ice cream shop equally, okay. Now we are asked the question that find a probability that a strawberry flavor ice cream will be ordered by a random student visiting a shop, that is any student comes what is the

probability that he or she would be ordering ice cream which is a strawberry flavor? Second question is that suppose a student comes has ordered a strawberry ice cream, what are the chances that student is a girl student? So these, let us try to solve this problem.

(Refer Slide Time: 16:49)

$P(G) = P(B) = 0.5$  ✓  
 $P(C|G) = 0.55$  &  $P(C|B) = 0.5$   
 $P(V|G) = 0.20$  &  $P(V|B) = 0.3$   
 $P(S|G) = 0.25$  &  $P(S|B) = 0.2$   
 $P(S) = ?$   
 $P(G|S) = ?$

*Total Probability rule*  
 $P(S) = P(S|G)P(G) + P(S|B)P(B)$   
 $= (0.25)(0.5) + (0.2)(0.5)$   
 $= 0.225$

*Bayes Rule*  
 $P(G|S) = \frac{P(S|G)P(G)}{P(S)}$   
 $= \frac{(0.25)(0.5)}{0.225}$   
 $= 0.56$

So here we are we have simplified this here, as to what we want. We say that the boy student and girl students frequent the shop equally it means that both probability of boys going to the shop or the girl going to the shop is equal and therefore it is half. As I said we are not considering the case that the faculty also might be visiting the shop, okay.

So we are restricting ourselves only to the students. Now C is for chocolate ice cream, V is for vanilla ice cream and S is for strawberry ice cream, okay. So then it says that if a girl visits then the chance that she buys a chocolate ice cream is 55 percent, it means that probability of a chocolate ice cream given that a girl is has visited is 0.55 and similarly the same thing will be 0.5 with respect to boy.

So these are the conditional probabilities given to us, so then let us find out what the question is. The question says that if anybody wants to go and buy a strawberry ice cream, what is the possibility probability that a ice cream bought will be strawberry? So first question is that what is probability of S? This is question number 1, second it says that suppose a student comes and buys the strawberry ice cream.



So given that it has bought a strawberry ice cream, what is the chance that the student is a girl student? So this is what we are looking for, okay. So this is what we are actually looking for, how shall we solve this? You must have done this sometime in your other courses or other high school level courses, so here we come to we say that let us use total probability rule okay.

Where do we apply total probability rule? We say that probability the of strawberry any person coming and buying a strawberry ice cream is probability that a strawberry item is bought by a girl multiplied by probability of girl plus probability of strawberry item bought by a boy given that probability of boy visiting the shop.

So we say that suppose a girl visit a shop and buys the strawberry ice cream then this is the portion which defines that probability that a girl visits the ice cream shop is  $P_G$ , probability of a girl visiting ice-cream shop and she goes and buys a strawberry ice-cream. So these two are multiplied, similarly these are multiplied and therefore the answer here I think this is 0.25, this is 0.5 plus 0.2 multiplied by 0.5. If you simplify it is 0.2 to 5, okay.

Now let us try to answer the second question, student has come and bought a strawberry ice-cream what is the probability that given that strawberry ice-cream has been bought the student is a girl student? Now you note that we do not have this probability, either we have to look for this but we already have this probability, right.

So if you use the Bayes theorem, it says that probability of a girl buying ice cream strawberry given that a strawberry ice-cream has been bought the person is the girl is probability that a girl has come a strawberry ice cream is bought given that a girl has visited multiplied by probability of the girl visiting the ice cream shop divided by probability of the strawberry ice cream itself.

This is where we are applying Bayes rule or Bayes theorem and therefore here the answer is going to be this is given to us as 0.25 multiplied by 0.5 divided by 0.225 and this roughly becomes 0.56, so more than half the time you will find that, that is a girl and that makes sense because if you look at it the higher percentage of buying ice cream is by a girl but it is the difference is 0.25 and 0.2 while here it says that more than half the time it will be a girl, so this is how the Bayes theorem gets applied.

Now let us come to estimation part of it, so far we talked about we recall what we had done in the past. Parametric estimation let us look at in a different way.

(Refer Slide Time: 22:46)

Parametric Estimation:

- Data:  $X_1, X_2, \dots, X_n \sim \text{iid}(f(\theta))$ :  $x_1, x_2, \dots, x_n$  data observed
- Parameter  $\theta$ : unknown, fixed value
- Background information on  $\theta$ 
  1.  $X_1, X_2, \dots, X_n \sim \text{iid } N(\mu, \sigma_0^2)$ :  $\mu$  is unknown &  $\sigma_0^2$  is known
  2. past  $\bar{y}$  estimate of  $\mu$ : past info.  $\bar{y} \sim N(\mu_0, \sigma_1^2)$   $\mu_0$  is known  $\sigma_1^2$  is known

How to use this background inf. to estimate  $\mu$  from fresh data.

In parametric estimation we have a data, right, what is the data? We say that  $X_1, X_2, \dots, X_n$  is iid that is independent and identically distributed sample from a function  $f(\theta)$ , okay,  $f(\theta)$  where what we really mean is that  $x_1, x_2, \dots, x_n$  is the data observed okay. So this is a sample and this is the data we have observed. Then there is this parameter  $\theta$ , this parameter  $\theta$  we say that it is unknown, okay.

But we also assume in parametric estimation as we said this parameter  $\theta$  actually is a value that the population takes, we do not know exactly how the population looks like we are trying to understand the population through the small sample of data we have collected. So data observed is  $x_1, x_2, x_3, \dots, x_n$  is our small data and from that we have to find out what is the value of  $\theta$  and when we do the parametric estimation so far.

What we did so far, we assume that  $\theta$  is some unknown and fixed value right, we say that oh well the population is unknown but it has some mean value. For example, it has some mean value and it is a fixed mean value for the population. Now we want to ask ourselves a question is this a good assumption? Now we want to ask ourselves a question is this a good assumption? So that is our next.

Suppose this is not the first time that you have collected the data to understand this  $\theta$  you have not collected this data for the first time you had a background information on  $\theta$  in other words in the past some time you had collected the data and you had estimated  $\theta$ , okay.

So what when you had already done this exercise once why not use that information and improve upon your new estimation, again I repeat I have some background information on  $\theta$ , how do I get the background information? Very simple I have already done this exercise in the past and from the past exercise I have some information as to what the (thea) generally the value  $\theta$  can take, that is also an estimation.

And now the question is that I having already done this exercise one and having got this background information, why should I simply repeat the infer, the process to get another estimate of  $\theta$ ? Why cannot I use the background information that I already have? Let us take an example to understand it better. Suppose I have a data coming from a population of normal distribution with mean  $\mu$  and variance  $\sigma^2$ , okay.

I say  $\mu$  is unknown and  $\sigma^2$  is known, this is only to make life simple for the example. In the past I have done the analysis and found an estimate of  $\mu$ , okay. I will denote it by  $\hat{\mu}$ , okay. So in the past I have founded a  $\hat{\mu}$  estimate of  $\mu$  and I have done it repeatedly and you also know that when you estimate  $\bar{x}$ ,  $\bar{x}$  itself,  $\bar{x}$  is a good estimator of  $\mu$  but then  $\bar{x}$  itself has its distribution.

So similarly this  $\mu$ , I am not saying that you have found a maximum likelihood estimator or which estimator but suppose you have found some estimator okay and let us assume that past information says that  $\hat{\mu}$  itself is distributed normally with some  $\mu_0$  and some variance  $\sigma_1^2$  where  $\mu_0$  is known and  $\sigma_1^2$  is known. So when I have done this exercise in the past I have observed that my mean, the estimator of mean also follows a normal distribution with a certain fixed value  $\mu_0$  and a variance  $\sigma_1^2$ .

So then the question is how to use this information? How to use this background information to estimate  $\mu$  from fresh data? Please note what is being said here is that you have done this exercise in the past and you have found that your estimator itself follows a distribution with a known mean and a known standard deviation in this case we assume it to be normal.

And now you have got a fresh data here, how to utilize this background information? This is the background information, how to utilize this information in order to get a better estimator? That certainly is good to use the background information if you already know something, if you already use it, it always gives, takes you closer to the reality. So this is what is a question and the answer lies in Bayes theorem.


(Refer Slide Time: 30:38)

Bayes Theorem is the answer

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad ; \quad P(\theta | \text{Data}) = \frac{P(\text{Data} | \theta) P(\theta)}{P(\text{Data})}$$

• In terms of density function:

$$f_{\theta|x_1, x_2, \dots, x_n}(\theta|x_1, x_2, \dots, x_n) = \frac{f_{x_1, x_2, \dots, x_n|\theta}(x_1, x_2, \dots, x_n|\theta) f_{\theta}(\theta)}{f_{x_1, x_2, \dots, x_n}(x_1, x_2, \dots, x_n)}$$

$$f(\theta|x_1, x_2, \dots, x_n) = \frac{f(x_1, x_2, \dots, x_n|\theta) f(\theta)}{f(x_1, x_2, \dots, x_n)}$$


The Bayes theorem says this is I have stated probability of A given B is probability of B given A multiplied by probability of A divided by probability of B. Let us translate it in terms of the parameter and the data.

So I am saying that in terms of density function instead of probabilities I can write it as a density function. So let me make it simple here, before we go next let us note it down that I can also write it as probability of estimate the unknown parameter theta given that I have got a fresh data is equal to probability that my data given the parameter theta multiplied by probability of theta divided by probability of only data without theta, this is what it translates into.

In terms of density function I would say that probability density of theta given X1, X2, X3, Xn is the probability of X1, X2, X3, Xn given theta multiplied by the probability density function of theta divided by the probability density of only X1, X2, X3, Xn without theta. Now you see these are my notations so I am just removing this, I am just removing these, so that you know this actually defines because everywhere I have written F.

So to distinguish  $f$  as a probability density function from this variable to this variable I had written a subscript. Let us remove it we understand that all these  $F$ s are different and they refer to their argument to take the form of it. So in simply simplified form it can be written the same thing what I have written up can be written in this fashion and this has the 3 parts into it which is what we want to talk about.

(Refer Slide Time: 33:20)

### Three elements of Bayes Estimation

$$f(\theta|x_1, x_2, \dots, x_n) = \frac{f(x_1, x_2, \dots, x_n|\theta)f(\theta)}{f(x_1, x_2, \dots, x_n)}$$

- Prior distribution (back ground information)
- Data
- Posterior distribution : obtained through incorporating prior with data

$$f(x_1, x_2, \dots, x_n) = \int \underbrace{f(x_1, x_2, \dots, x_n|\theta)}_{\text{likelihood function}} f(\theta) d\theta$$



### Bayes Estimator

Bayesian estimator of  $\theta$ ,  $\tilde{\theta}$  is defined as expected value of posterior distribution of parameter  $\theta$

$$\tilde{\theta} = E(\theta|x_1, x_2, \dots, x_n) = \int \theta f(\theta|x_1, x_2, \dots, x_n) d\theta$$

- Since prior is arrived at through past experience, it does not contain any unknown parameters.



There are three parts or three elements to this, we have used a Bayes estimation, so what we have got is a prior distribution that is the background information which is shown here in blue, you see?

The prior information is in blue, the data is given in green because data already has some assumption on theta so I have not shown this in green but this also is a part of data.

And then there is a posterior distribute which is obtained through incorporating the prior with data and that distribution is given in red. Now what is this F, the pure density function of X1, X2, X3, Xn without theta well it is integrating out the numerator with respect to D theta. You imagine now that the theta is also having a density function, it also has a distribution, it is a random variable.

$$f(\theta|x_1, x_2, \dots, x_n) = \frac{f(X_1, X_2, \dots, X_n|\theta)f(\theta)}{f(x_1, x_2, \dots, x_n)}$$

- Prior distribution (back ground information)
- Data
- Posterior distribution : obtained through incorporating prior with data

$$f(x_1, x_2, \dots, x_n) = \int f(X_1, X_2, \dots, X_n|\theta)f(\theta) d\theta$$

So I think this is a kind of confusing idea, so I want to clarify it. Our data is a random data, it has a unknown parameter that needs to be estimated. So far in parametric estimation we have been taking it only as a fixed unknown value but the question comes that if for the same population in the past I have got data and I have already estimated this parameter theta once or maybe many times then why do not I make use of the same theta, same information on theta to take a new and fresh estimate of my theta in the population.

Because in any case theta remains unknown, so instead of imaging theta as a fixed value, here we are trying to imagine that theta is also a random variable, it also varies depending on the sample, different samples you take it takes different values and suppose we know that this theta follows a certain distribution which has no unknowns in it, that distribution is completely known. In that case this integral is possible.

So if you integrate out this is like the likelihood function if you look at this, this part is the likelihood function of X, this is a likelihood function if you remember this is the function which we maximize as a function of X because X1, X2, X3, Xn are unknown to us. Now we keep that likelihood function, we know the distribution of theta and then we integrate it out we get a pure density of F of X.

These are the three important elements of Bayes estimation, there is a prior distribution there is a data and then there is a posterior distribution which is derived in this manner. And when this happens you have something called a Bayes estimator, so the Bayesian estimator of theta is then the expected value of the posterior distribution of parameter theta. Remember this is the posterior distribution, this is the posterior distribution of theta given X1, X2, X3, Xn data.

$$\tilde{\theta} = E(\theta|x_1, x_2, \dots, x_n) = \int \theta f(\theta|x_1, x_2, \dots, x_n) d\theta$$

You remember that I have given this as a smaller values because they are the actual data that we are calculating it here also they are smaller value, here only they are working as a random variable.

So with this what we get is we say that once you have a posterior distribution of parameter theta the expected value of theta from the posterior distribution is called the Bayesian estimator of theta. Generally it is seen that it improves upon the estimate of theta from the previous estimator that we have already managed.

So this is what is called Bayes estimator I think in many of your R-lessons you are going to work out certain Bayesian estimation and based on hypothesis testing procedures using Bayes estimator in your R-exercise and most important thing we have to realize here is that since prior is derived from the past experience it does not contain any unknown parameters.

(Refer Slide Time: 38:52)

## Summary

- Revisit to Conditional Probability and Bayes theorem
- Interpretation of Bayes theorem as a technique to get  $P(B|A)$  from  $P(A|B)$
- Re-look at Bayes theorem in terms of probability density function
- Introduced Bayes estimator, using Bayes theorem using back ground information on the unknown parameter and data, defining
  - Prior density function for unknown parameter from back ground information
  - Data obtained
  - Posterior Density obtained through Bayes theorem



So with this let us summarize, we revisited the conditional probability and Bayes theorem. Interpretation of Bayes theorem as a technique to get a probability of B given A from probability of A given B we had a relook at Bayes theorem in terms of probability density function and then we introduced Bayes estimator as the expected value of the posterior density obtained through Bayes theorem, what is the posterior density?

Well posterior density is obtained by using Bayes theorem on the prior information and the data you get the posterior density and from that when you estimate the parameter theta it is called a Bayesian estimator of theta. Thank you