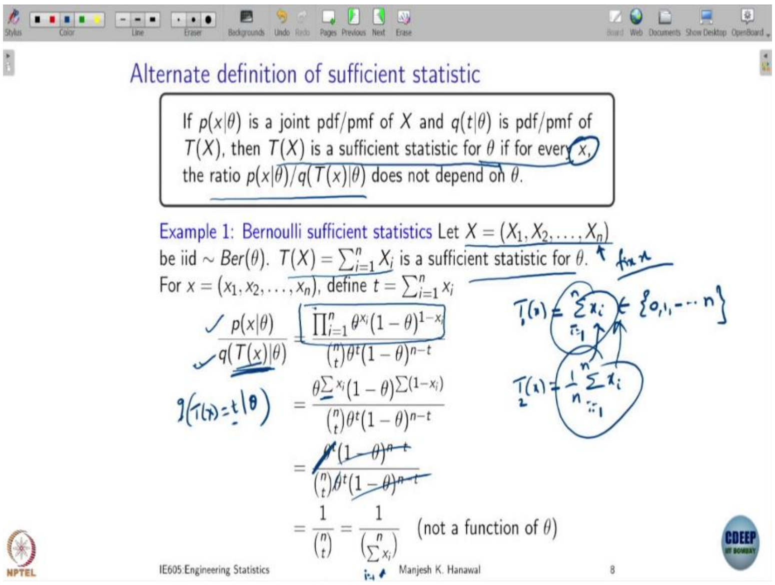
Engineering Statistics
Professor. Manjesh Hanawal
Industrial Engineering and Operation Research
Indian Institute of Technology, Bombay
Lecture 33
Characterization of Sufficient Statistics and Factorization Theorem

(Refer Slide Time: 00:15)



Alternate definition of sufficient statistic

If $p(x|\theta)$ is a joint pdf/pmf of X and $q(t|\theta)$ is pdf/pmf of $T(X)$, then $T(X)$ is a sufficient statistic for θ if for every x, the ratio $p(x|\theta)/q(T(x)|\theta)$ does not depend on θ.

Example 1: Bernoulli sufficient statistics Let $X = (X_1, X_2, \ldots, X_n)$ be iid $\sim Ber(\theta)$. $T(X) = \sum_{i=1}^{n} X_i$ is a sufficient statistic for θ. fix x

For $x = (x_1, x_2, \ldots, x_n)$, define $t = \sum_{i=1}^{n} x_i$

$$\frac{p(x|\theta)}{q(T(x)|\theta)} = \frac{\boxed{\prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}}}{\binom{n}{t}\theta^{t}(1-\theta)^{n-t}}$$

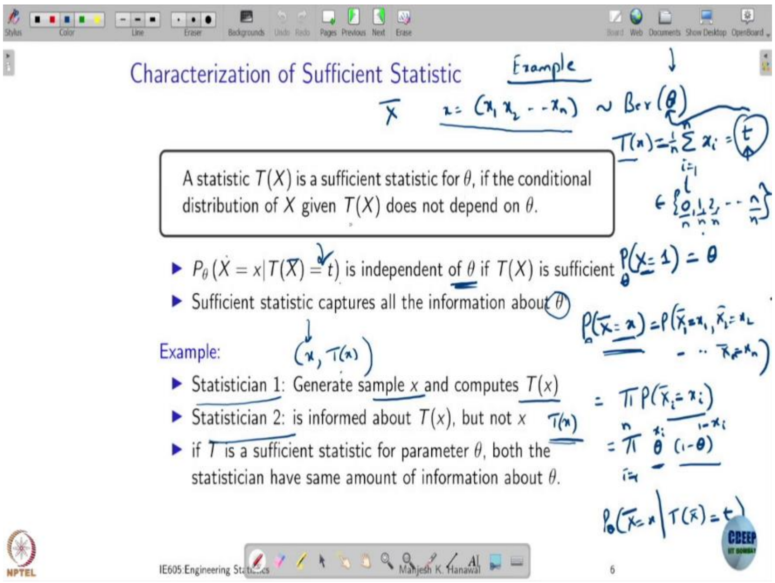
$$q(T(x)=t|\theta) \quad = \frac{\theta^{\sum x_i}(1-\theta)^{\sum(1-x_i)}}{\binom{n}{t}\theta^{t}(1-\theta)^{n-t}}$$

$$= \frac{\cancel{\theta^t}\cancel{(1-\theta)^{n-t}}}{\binom{n}{t}\cancel{\theta^t}\cancel{(1-\theta)^{n-t}}}$$

$$= \frac{1}{\binom{n}{t}} = \frac{1}{\binom{n}{\sum_{i=1}^{n} x_i}} \quad \text{(not a function of } \theta)$$

$T_1(x) = \sum_{i=1}^{n} x_i \in \{0, 1, \ldots n\}$

$T_2(x) = \frac{1}{n}\sum_{i=1}^{n} x_i$



Characterization of Sufficient Statistic

Example

$\bar{X}$ $\quad x = (x_1, x_2, \ldots x_n) \sim Ber(\theta)$

$T(x) = \frac{1}{n}\sum_{i=1}^{n} x_i = t$

$\in \{\frac{0}{n}, \frac{1}{n}, \frac{2}{n}, \ldots \frac{n}{n}\}$

A statistic $T(X)$ is a sufficient statistic for θ, if the conditional distribution of X given $T(X)$ does not depend on θ.

- $P_\theta(X = x|T(\bar{X}) = t)$ is independent of θ if $T(X)$ is sufficient $\quad P_\theta(X = 1) = \theta$
- Sufficient statistic captures all the information about θ

$P(\bar{X} = x) = P(\bar{X}_1 = x_1, \bar{X}_2 = x_2, \ldots \bar{X}_n = x_n)$

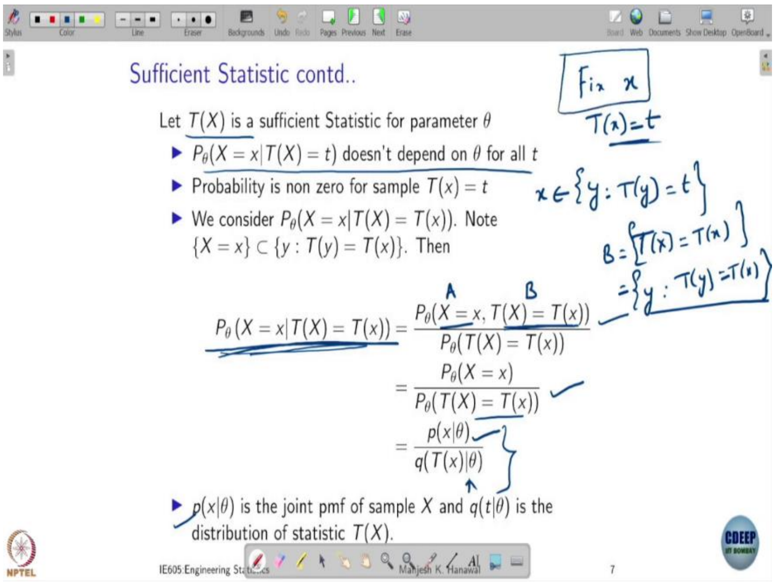
Example: $\quad (x, T(x))$

- Statistician 1: Generate sample x and computes $T(x)$
- Statistician 2: is informed about $T(x)$, but not x $\quad T(x)$
- if T is a sufficient statistic for parameter θ, both the statistician have same amount of information about θ.

$= \prod P(\bar{X}_i = x_i)$

$= \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$

$P_\theta(\bar{X} = x \mid T(\bar{X}) = t)$



Sufficient Statistic contd..

Let $T(X)$ is a sufficient Statistic for parameter θ

- $P_\theta(X = x|T(X) = t)$ doesn't depend on θ for all t
- Probability is non zero for sample $T(x) = t$
- We consider $P_\theta(X = x|T(X) = T(x))$. Note $\{X = x\} \subset \{y : T(y) = T(x)\}$. Then

$$P_\theta(X = x|T(X) = T(x)) = \frac{P_\theta(\overset{A}{X = x}, \overset{B}{T(X) = T(x)})}{P_\theta(T(X) = T(x))}$$
$$= \frac{P_\theta(X = x)}{P_\theta(T(X) = T(x))}$$
$$= \frac{p(x|\theta)}{q(T(x)|\theta)}$$

- $p(x|\theta)$ is the joint pmf of sample X and $q(t|\theta)$ is the distribution of statistic $T(X)$.

$\boxed{\text{Fix } x}$

$T(x) = t$

$x \in \{y : T(y) = t\}$

$B = \{T(x) = T(x)\}$
$= \{y : T(y) = T(x)\}$

IE605:Engineering Statistics Manjesh K. Hanawal 7

So, now let us do this example how to verify are given statistics sufficient or not? So, naturally we have one characterization, look into the ratio we just defined. For that, maybe what we want to now based on what we have just derived, we said that a statistic T is a sufficient statistics for this parameter theta, if it is conditional distribution of x given T(x) does not depend on theta. Now, the same thing we are going to say in an alternate manner we are going to say that T(x) is a sufficient statistic for theta, if for every x this ratio p(x|θ) divided by q(T|θ), this does not depend on theta.

Notice that now, this ratio has to be independent of theta for every possible x. It is not that if you show it for one x then you are done. This whatever we said this, this should not depend on which x you are considering. So, let us look into this example, we just discussed, let us say this x I do not have some time putting bar and not bar do not get confused. This is x I am going to treat as a random vector here, which is coming from underlying population which has a Bernoulli parameter theta. Now, I am interested in this sum notice that n is fixed, I am fixing n. Now our claim is if I am going to take the sum of the sample, that is a sufficient statistics for theta.

So, how to verify this just apply the formulation we have compute their marginal sorry, compute their unconditional probability under the parameter theta, then compute unconditional probability of your statistics under the same parameter theta and then see that that ratio does not depend on theta. So, we just discussed that the distribution of x under parameter theta has this probability, anybody has any question on this, we just did that, we can write this now, that denominator is only we need to think through now T(x) is summation of x_i, we know that this value can take only 0 and n and we know that this is a sum of n Bernoulli random variables. So, sum of n Bernoulli random variables, which are identically distributed is binomial.

So, now so I should have written it like this like, I should have written this $q(T(x)=t|\theta)$. So, that a T(x) taking value t is now n choose t q to the power t 1 minus theta n minus t. So, you know now if you just simplify this, now it is simple algebra, you just you because of this product form, you take the you make this product into summation by taking this into the exponent and if you see that, what you are going to get is eventually all this theta gets knocked off. What remains is 1 upon n choose t, which does not depend on your theta parameter, it only depends on your data. So, this should be I equals to 1 to n xi.

So, now is this T is a sufficient statistic for your parameter theta and now, suppose instead of this I have taken T(x) equals to 1 by n this is the sample average earlier I was taking sum I am now taking the average let us call this one, T1 and T2, these two are different statistics? One is taking sum and others taking the average. But in terms of information about theta, are they any different?

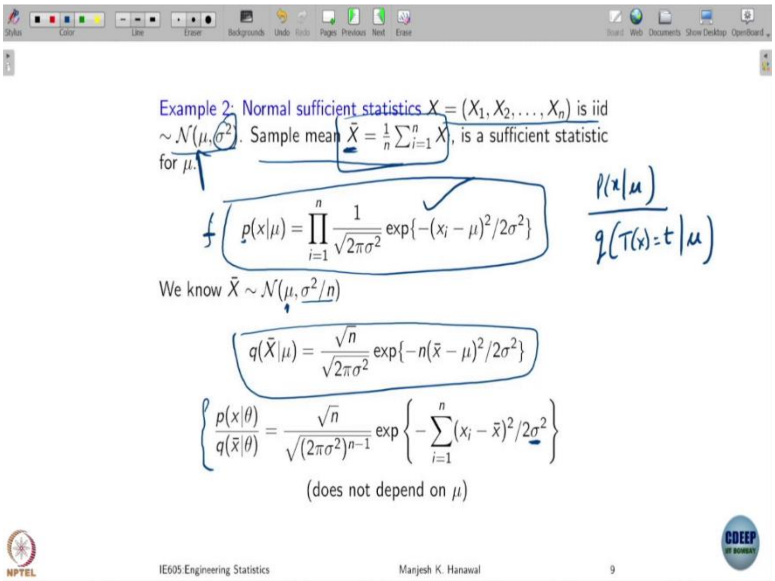
Student: (())(05:57)

Professor Manjesh Hanawal: No, why?

Student: (())(06:04)

Professor Manjesh Hanawal: Because in this case, n is a constant. I mean, if I know this quantity, I can easily get this quantity, even if I know this quantity, if I know n I can go back and get this quantity. So, what if I just by knowing one it is not that I do not know anything about other or I can have less information about that other. So, because of this, these are fine.

(Refer Slide Time: 06:30)



Example 2: Normal sufficient statistics $X = (X_1, X_2, \ldots, X_n)$ is iid $\sim \mathcal{N}(\mu, \sigma^2)$. Sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X$, is a sufficient statistic for μ.

$$f \quad p(x|\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-(x_i - \mu)^2/2\sigma^2\}$$

$$\frac{p(x|\mu)}{q(T(x)=t|\mu)}$$

We know $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$

$$q(\bar{X}|\mu) = \frac{\sqrt{n}}{\sqrt{2\pi\sigma^2}} \exp\{-n(\bar{x} - \mu)^2/2\sigma^2\}$$

$$\frac{p(x|\theta)}{q(\bar{x}|\theta)} = \frac{\sqrt{n}}{\sqrt{(2\pi\sigma^2)^{n-1}}} \exp\left\{-\sum_{i=1}^{n}(x_i - \bar{x})^2/2\sigma^2\right\}$$

(does not depend on μ)

IE605:Engineering Statistics Manjesh K. Hanawal 9

So, another quick example. Let us say, now I am saying that data is coming from, Gaussian, population with a parameter mu and sigma square. I am fixing now the underlying population, which is Gaussian with parameter mu and sigma square. But I am interested I am in this example, let us say that I know sigma square, but I do not know mu. Mu is the parameter of my interest. Now, I want to get information about this mu, which is a statistic which will give me good information about this mu. Any guess? I have this underlying parameter mu? From data, I want to get information about that parameter mu? How, what function I should apply? And my data are what should be my statistic so that I get a good information about new from the data?

Student: (())(07:44)

Professor Manjesh Hanawal: Others?

Student: (())(07:52)

Professor Manjesh Hanawal: Oh, no, why I cannot take expectation? I only have this samples. I have only data. Now, I do not have that luxury like if you, want you are talking about expectation, you need to know distribution with what respect to but I do not have the full information about the distribution, I do not know what are the new parameter here.

So, we have earlier we have discussed that sample mean is one good estimator for your mean. And we have said that is a consistent estimator, and also unbiased estimator. But now, let us see that now, instead of thinking about in terms of estimator consistency, and all now we are thinking in terms of the sufficiency principle.

Now is, which is a good data reduction method here to get information about mu? Sample mean is a data reduction method, maybe let us try that. Let us try that sample when I do the sample reduction. Sorry, when I do the sample mean, is that a good statistic? Believe it or not a sufficient statistic.

Now so, notice that I am always using p here, I am not using f even it is a continuous distribution, just to make sure that this is same for both discrete and continuous case. I mean, it should be clear from you for all of you whether it is continuous or discrete. I may just use simply p. So, this is the probability density functional of Gaussian. Now, I am using this statistic x bar. Now, what I know about x bar?

So, now to do compute, to verify whether it is a sufficient statistic or not, I need to compute that ratio has two things. One is the unconditional probability under that parameter mu and what was the denominator about the distribution of statistic let that it is taking some value under the parameter mu. Now, let us compute I know that the unconditional probability is this under the parameter mu.

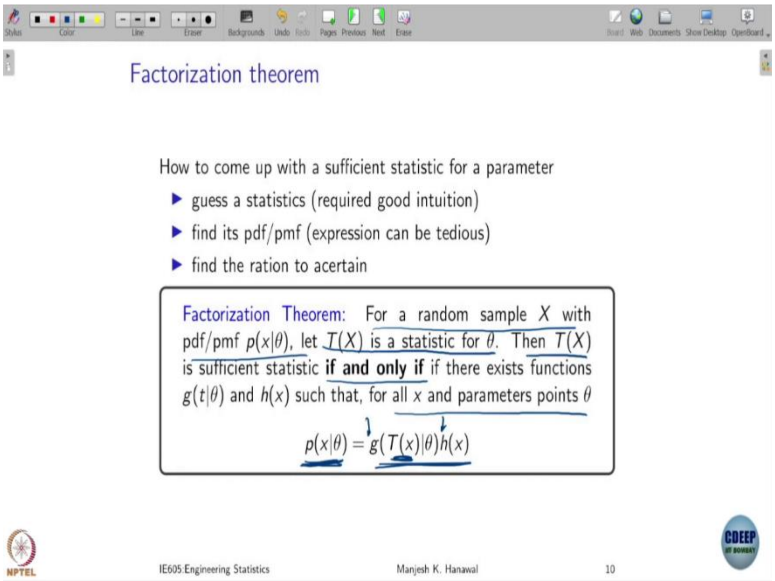
Now compute, let us compute the distribution of this statistic x bar what we know about x bar if I average and IID Gaussian samples, it is going to be still Gaussian, but with mean mu and variance sigma square by n. Now, I know distribution I can write a distribution only thing I have done is replace sigma square by sigma square by n which is given me this value.

Now, compute the ratio if you now compute your ratio, you will see that this entire quantity mu does not appear here. Even though sigma square appears here sigma squared is not my unknown parameter sigma square is known only unknown parameter is mu in this ratio, mu does not appear what does this indicate? What sample mean is a for what? For mu of?

Student: (())(11:49)

Professor Manjesh Hanawal: Gaussian distributions. So, we need to tell all these, like, it is a sufficient strategy for which parameter under which population distribution.

(Refer Slide Time: 12:04)



Factorization theorem

How to come up with a sufficient statistic for a parameter

- guess a statistics (required good intuition)
- find its pdf/pmf (expression can be tedious)
- find the ration to acertain

Factorization Theorem: For a random sample X with pdf/pmf $p(x|\theta)$, let $T(X)$ is a statistic for θ. Then $T(X)$ is sufficient statistic **if and only if** if there exists functions $g(t|\theta)$ and $h(x)$ such that, for all x and parameters points θ

$$p(x|\theta) = g(T(x)|\theta)h(x)$$

IE605:Engineering Statistics Manjesh K. Hanawal 10

Now, if you want to know something is a sufficient statistic or not. So, first thing is, if you want to know some parameter, now, you feel that having a sufficient statistics is good? Because it can essentially capture all the information about the parameter of your interest. But then, what should be that? Is it always easy to get that sufficient statistic? Is it that sample mean always happens to be such sufficient statistics? We do not know a priori. Now, that then the question comes how to find it. What the previously we just saw is given a statistic to check whether that is a sufficient statistics or not, somebody is claiming somebody has given you T. And he is claiming that see this is a sufficient statistic.

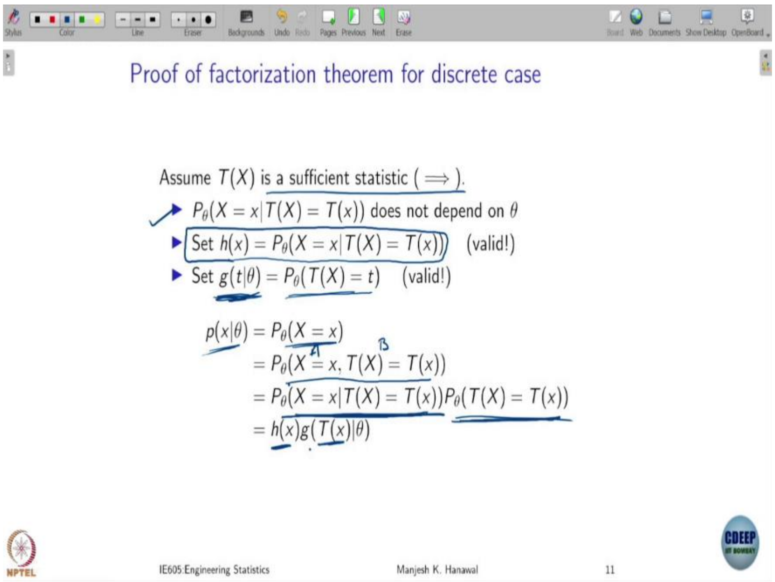
Now, what you are doing is you are telling him you are verifying whether his claim is correct or not by computing the ratio. But that guy who is coming up with the sufficient statistics, how he is going to come up. Your job was easy. You just verified it by applying finding this ratio, but that guy's job was hard to come up with a sufficient statistic.

Now, how to find a sufficient statistic, is there a method so factorization comes to rescue there to some extent. And it tells us when it is possible to have sufficient statistic for an underlying population. Now, let us see what is the factorization theorem states. Let us say you have a random sample x, which has the underlying pdf or pmf, which I am denoting p x given theta. And let us be T(x) be sufficient statistics for parameter theta.

Now, it is saying that T(x) is a sufficient statistic, if and only if there exist functions g(T(x)|θ).h(x) such that for every x and theta you our probability mass function should be factorizable in this form, it is an involved statement but try to pass it. It is saying that some statistic is going to be sufficient if and only if, if it is probability mass function or a probability density function can be factorise able into two parts were like factorizable into two things g function and h function, where h only depends on x and g only depends on t. It depends on x only through T, not explicitly on X but through T(x).

If you are able to factorise your probability density or probability mass function in this fashion for some statistic T, then that statistics is going to be sufficient statistics. Now, the guy who wants to use statistics now he has to do use this, he has to pick some statistics and see whether this holds. If this holds, then that guy is confident that that is sufficient to say, then he can confidently give you, take this and verify it is a sufficient statistics.

(Refer Slide Time: 16:09)



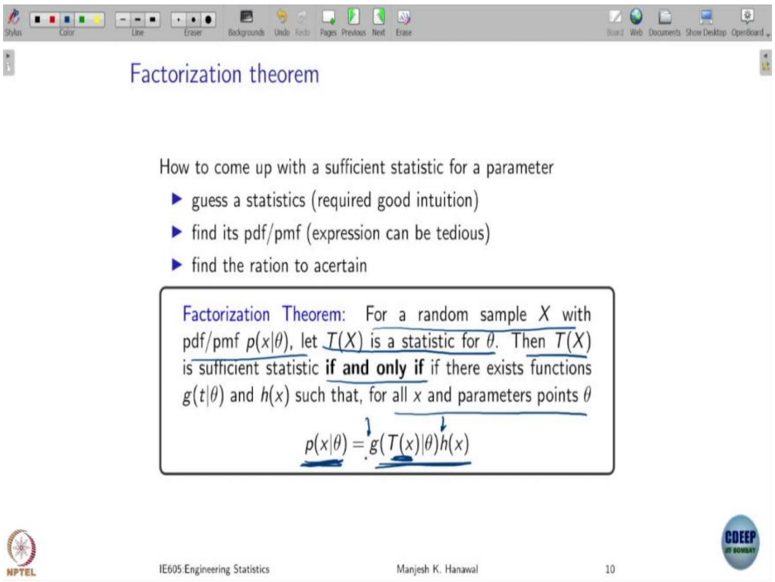
Proof of factorization theorem for discrete case

Assume $T(X)$ is a sufficient statistic ($\implies$).

- $P_\theta(X = x|T(X) = T(x))$ does not depend on θ
- $\boxed{\text{Set } h(x) = P_\theta(X = x|T(X) = T(x))}$ (valid!)
- Set $g(t|\theta) = P_\theta(T(X) = t)$ (valid!)

$$\begin{aligned} p(x|\theta) &= P_\theta(X = x) \\ &= P_\theta(X \overset{A}{=} x, \overset{B}{T(X)} = T(x)) \\ &= P_\theta(X = x|T(X) = T(x))P_\theta(T(X) = T(x)) \\ &= h(x)g(T(x)|\theta) \end{aligned}$$

IE605:Engineering Statistics Manjesh K. Hanawal 11



Factorization theorem

How to come up with a sufficient statistic for a parameter

- guess a statistics (required good intuition)
- find its pdf/pmf (expression can be tedious)
- find the ration to acertain

Factorization Theorem: For a random sample X with pdf/pmf $p(x|\theta)$, let $T(X)$ is a statistic for θ. Then $T(X)$ is sufficient statistic **if and only if** if there exists functions $g(t|\theta)$ and $h(x)$ such that, for all x and parameters points θ

$$p(x|\theta) = g(T(x)|\theta)h(x)$$

IE605:Engineering Statistics Manjesh K. Hanawal 10

Now, let us too quickly go through it is proof. I am now going to do the forward part. So, what is the forward part here? If and only if there exists function g T, g, and h such that this holds. Now, for the sufficiency what I need to show T is a sufficient statistic if something happens. Now, I am going to start with saying that T is sufficient statistics.

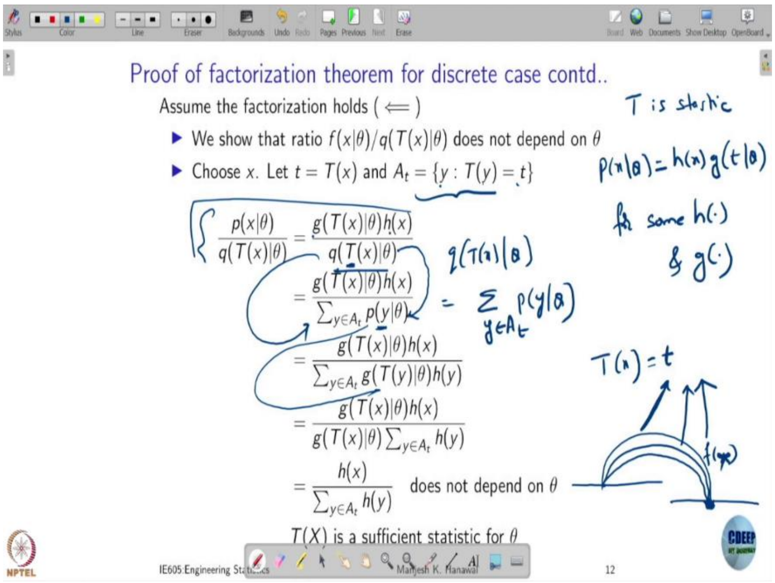
And now I am going to argue that if that is the case, my pmf is going to factorise in the way I want. If T is a sufficient conditions, sufficient statistics, I am going to use the properties. First thing is if T is a sufficient statistics, it is this conditional probability, it does not depend on theta. I am going to start with that.

Now, what I am going to do in this is, I am going to take h(x) to be this conditional probability. And can I claim that this h(x) does not depend on theta? Yes, or no? Yes. That is, by definition, if T is a sufficient statistic, I know that this conditional probability does not depend on theta. And I am calling theta h(x).

And now, the g of, say what I am going to do in the if thing, I am going to show that if T is a kind of sufficient condition, I need to show that there exists a g and h function such that this factorization holds, I need to show that now what I am trying to do in this proof is I am trying to show you such a g and h exist whenever T is a sufficient statistics. And now for g function, I am going to take this to be probability that T is going to take some this is basically distribution of my statistics itself.

Now, probability of x given theta is this, by definition. I know that I can write it like this. Why is that? Because I already told you, if I take this as set A and this has set B, A is a subset of B. That is why I could write like this. And now I am going to write this joint distribution as conditional probability and this marginal probability and not that is it now I have defined this portion has h and this portion as g and I know that this portion does not depend on T, it only depends on x. And this portion depends on x only through T(x) not directly on x. Now, is the sufficient part is complete. What are shown is if T is a sufficient statistic, I have this pmf split into h and g function where h depends only on x and g depends only on the T(x).

(Refer Slide Time: 19:37)



Proof of factorization theorem for discrete case contd..

Assume the factorization holds ($\Longleftarrow$)

- We show that ratio $f(x|\theta)/q(T(x)|\theta)$ does not depend on θ
- Choose x. Let $t = T(x)$ and $A_t = \underbrace{\{y : T(y)}= t\}$

$$\frac{p(x|\theta)}{q(T(x)|\theta)} = \frac{g(T(x)|\theta)h(x)}{q(T(x)|\theta)}$$

$$= \frac{g(T(x)|\theta)h(x)}{\sum_{y\in A_t} p(y|\theta)}$$

$$= \frac{g(T(x)|\theta)h(x)}{\sum_{y\in A_t} g(T(y)|\theta)h(y)}$$

$$= \frac{g(T(x)|\theta)h(x)}{g(T(x)|\theta)\sum_{y\in A_t} h(y)}$$

$$= \frac{h(x)}{\sum_{y\in A_t} h(y)} \quad \text{does not depend on } \theta$$

$T(X)$ is a sufficient statistic for θ

T is statistic

$p(x|\theta) = h(x)\,g(t|\theta)$

for some $h(\cdot)$

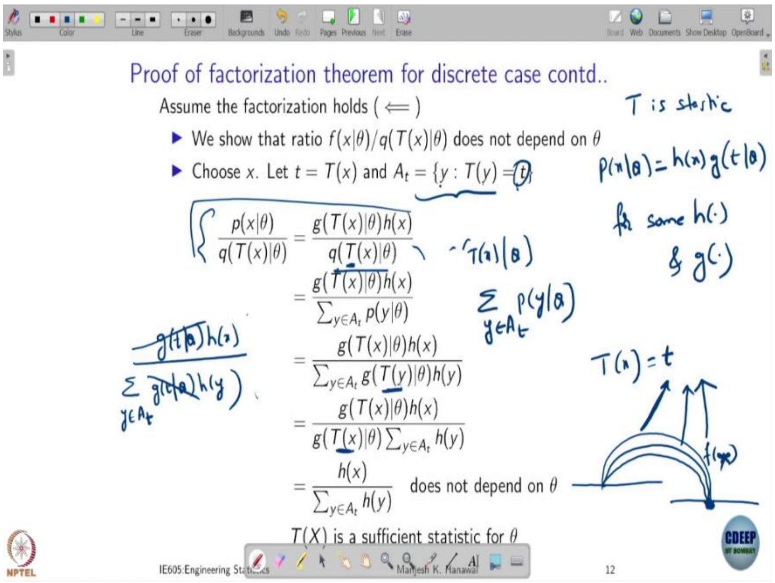
& $g(\cdot)$

$q(T(x)|\theta) = \sum_{y\in A_t} p(y|\theta)$

$T(x) = t$

$f(y)$

IE605:Engineering Statistics Manjesh K. Hanawal 12



Proof of factorization theorem for discrete case contd..

Assume the factorization holds ($\Longleftarrow$)

- We show that ratio $f(x|\theta)/q(T(x)|\theta)$ does not depend on θ
- Choose x. Let $t = T(x)$ and $A_t = \underbrace{\{y : T(y)}= t\}$

$$\frac{p(x|\theta)}{q(T(x)|\theta)} = \frac{g(T(x)|\theta)h(x)}{q(T(x)|\theta)}$$

$$= \frac{g(T(x)|\theta)h(x)}{\sum_{y\in A_t} p(y|\theta)}$$

$$= \frac{g(T(x)|\theta)h(x)}{\sum_{y\in A_t} g(T(y)|\theta)h(y)}$$

$$= \frac{g(T(x)|\theta)h(x)}{g(T(x)|\theta)\sum_{y\in A_t} h(y)}$$

$$= \frac{h(x)}{\sum_{y\in A_t} h(y)} \quad \text{does not depend on } \theta$$

$T(X)$ is a sufficient statistic for θ

$$\frac{\cancel{g(t|\theta)}h(x)}{\sum_{y\in A_t} \cancel{g(t|\theta)}h(y)}$$

T is statistic

$p(x|\theta) = h(x)\,g(t|\theta)$

for some $h(\cdot)$

& $g(\cdot)$

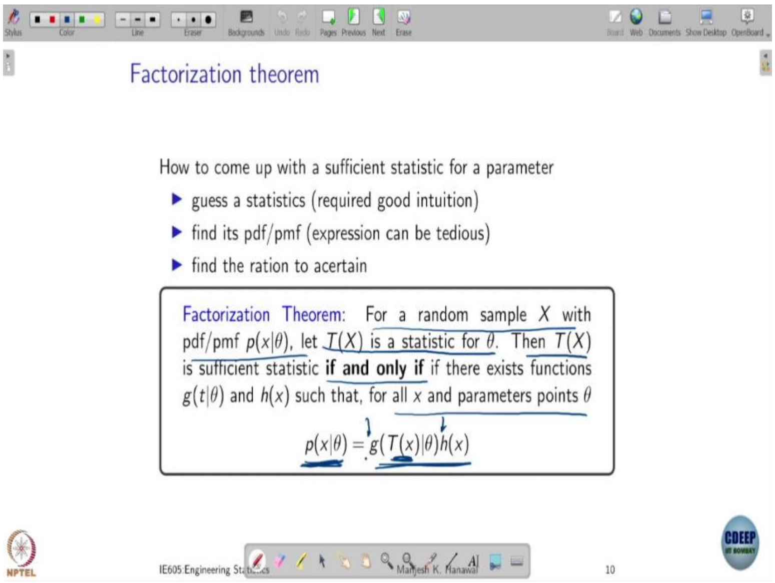
$T(x)|\theta)$

$\sum_{y\in A_t} p(y|\theta)$

$T(x) = t$

$f(y)$

IE605:Engineering Statistics Manjesh K. Hanawal 12



Factorization theorem

How to come up with a sufficient statistic for a parameter

- guess a statistics (required good intuition)
- find its pdf/pmf (expression can be tedious)
- find the ration to acertain

Factorization Theorem: For a random sample X with pdf/pmf $p(x|\theta)$, let $T(X)$ is a statistic for θ. Then $T(X)$ is sufficient statistic **if and only if** if there exists functions $g(t|\theta)$ and $h(x)$ such that, for all x and parameters points θ

$$p(x|\theta) = g(T(x)|\theta)h(x)$$

IE605:Engineering Statistics Manjesh K. Hanawal 10

Now, let us do the opposite direction? Now, I am assuming that factorization already holds. For some statistic, let us take some statistic T, and assume that the factorization holds. If that factorization holds, I need to show that T is a sufficient statistic and how to show that T is a sufficient statistic, I am going to go with that characterization of sufficient statistics where the ratio does not depend on theta. Now, let us see. So, T is my some statistic and I am assuming that p of x given theta factorises like this for some h function and g function now I am under the assumption that my p factorizes, like this, I am going to start with that. So, to do that, I am going to start looking into this ratio.

Now, I know that p factorises into g and h function which are written here and the denominator are just kept it like this. Everybody agree with the first step? I simply use the fact that my piece factorable, next what is this probability. So, $q(T(x)|\theta)$ is nothing but this is summation over all y which are mapping to At, that means they are all I am now going to take this T looking into all those values y, which are getting mapped to that value t.

And then this is p of y given theta. I do not know all of you are able to follow this step this step is you need to understand. So, T(x)? T(x) is a mapping here. And let us say this is going to be value t, this is some value t, but there are multiple value wise that could be mapping to this y. There could be multiple y's, which could be mapping to this value t.

So, I need to sum over all of them. And when I take the probability of all those y's, I will get that probability. So, what I do see that the distribution on the statistics are converted into the distribution on samples. So, basically what all the points that would have converged to map to T(x) I am look taking probability of their sum.

This was like, we have to start thinking about what we did in the first half of the course, we said that we did function mapping? We did that. Let us, let us take one function here and tell this is f of omega. So, we set f of x function mapping we did, and when I took on point here that could be multiple points, which could be mapping to the same things, and that is when you found a distribution of the function of random variables, you people use this property there.

So, that is why now this prop, so basically T is now a function of a random variable here. And that is the property that I have used here. And now once I do this, I am now go back and use my property again that this p function is factorizable, p could be written as a factor or a product of h and g function I have written here, now if you notice here, here are T(y), is a constant?

Because all these T of y are going to map to the same T. And so I can pull it out. I pull it out of the summation, because all these T(y)'s have the same value. Maybe what I could have done is simply, let me write this this portion here this is g of t given sorry g of t given theta h of x and in the denominator, y belongs to At, but this t of y is also t. So, this could I could write t given theta and h of y.

So, I wrote it T(x), but I could write it simply small, now, this guy can get knocked off with this and now, after knocking off what we left is h(x) divided by summation of h(y). Now, what did we say h(x) this h(x) does not depend on theta h(y) is also does not depend on theta, does this ratio depend on theta.

No, then what you have just showed is this ratio here on the left-hand side is also does not depend on theta if it is not depend on theta then what we conclude T is a sufficient statistic is the necessary condition is clear now. So, we are able to show both direction. So, this is So, the nice thing about this factorization theorem is it is like a complete characterization it is not just like a sufficient it is a necessary condition as well.

So, I said earlier two steps, somebody will come up with their statistics and ask going to give somebody to verify it is a sufficient statistics. But the first guy he do not know how to find a statistic, but he comes with a random statistics and tries to use this factorization theorem and chooses the one which follows satisfies this factorization theorem in the first step itself, you know, that what he has is sufficient statistics because you saw that actually sub second step is already used in the first step to verify that it is a sufficient statistic. So, any questions about factorization theorem.

So, you should be comfortable in applying this and to be clear about g and h function like what is that when we say we should come up with a g and h function? What should that h depends on what is that g depends on and what is that they should not depend on that should be clear. You cannot come up with some arbitrary g h function and say that okay product laelia Take this as your g there is a restriction on what and g and h function and when you give me g and h function you need to prove me that h only depends on x. g only depends on t all this part. So let us stop here.