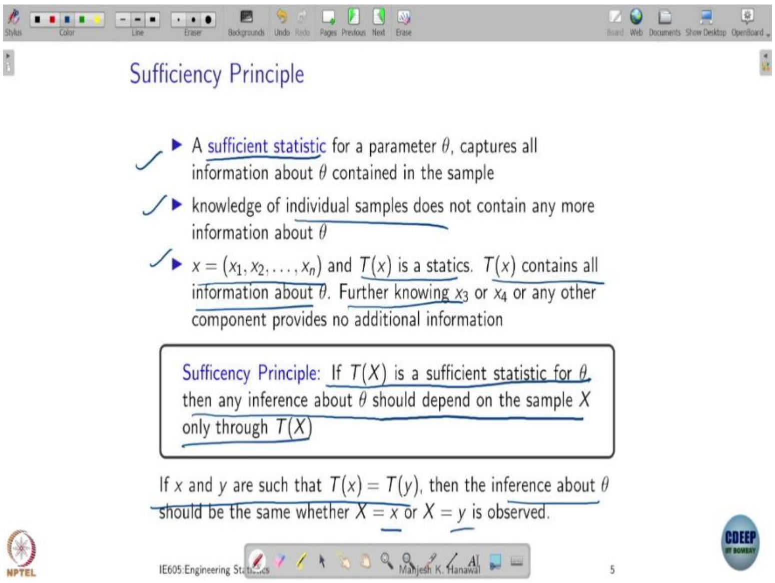
Engineering Statistics
Professor Manjesh Hanawal
Department of Industrial Engineering and Operation Research
Indian Institute of Technology, Bombay
Lecture 32

Sufficient Statistics and Characterization of Sufficient Statistics

(Refer Slide Time: 00:15)



Sufficiency Principle

- A sufficient statistic for a parameter θ, captures all information about θ contained in the sample
- knowledge of individual samples does not contain any more information about θ
- $x = (x_1, x_2, \ldots, x_n)$ and $T(x)$ is a statics. $T(x)$ contains all information about θ. Further knowing x_3 or x_4 or any other component provides no additional information

Sufficency Principle: If $T(X)$ is a sufficient statistic for θ, then any inference about θ should depend on the sample X only through $T(X)$

If x and y are such that $T(x) = T(y)$, then the inference about θ should be the same whether $X = x$ or $X = y$ is observed.

IE605:Engineering Statistics Manjesh K. Hanawal 5

Now, talking about sufficiency principles, recall that by sufficiency principle we are trying to say that when I reduce my data does it capture sufficient amount of information in the best possible way about my parameter theta. Now, we will be interested in you may have come up with the different statistics, but what is that statistic which will capture the best possible information about your parameter theta.

So, maybe we can different parameters different statistics we know. We can we have sample mean, we have sample variance, we have minimum value, maximum value all these statistics are there, but for my given population which are these statistics are capturing the best possible way the information about my underlying parameter theta.

So, we need to know make this which is the best possible statistics we need to kind of formalise, for that we are going to introduce something called sufficient statistic. Now, sufficient statistic for a parameter theta capture all information about theta contained in the samples. So, in a way it is giving me the best reduction or best possible way capturing information about theta from your samples. And it is also about reduction you have n number of samples and you have reduced it using your statistics to one value and now, in that value you are looking into how best information about my theta is captured.

Now, if I give you that reduced value and if in that reduced value information is well contained, then the in addition to that reduced value, if I give you any further information, that is individual samples, there should not add any more information about that particular parameter theta. So, that is we are just leading to the what should be the formal these are kind of like these points are like what should be the properties of a sufficient statistic.

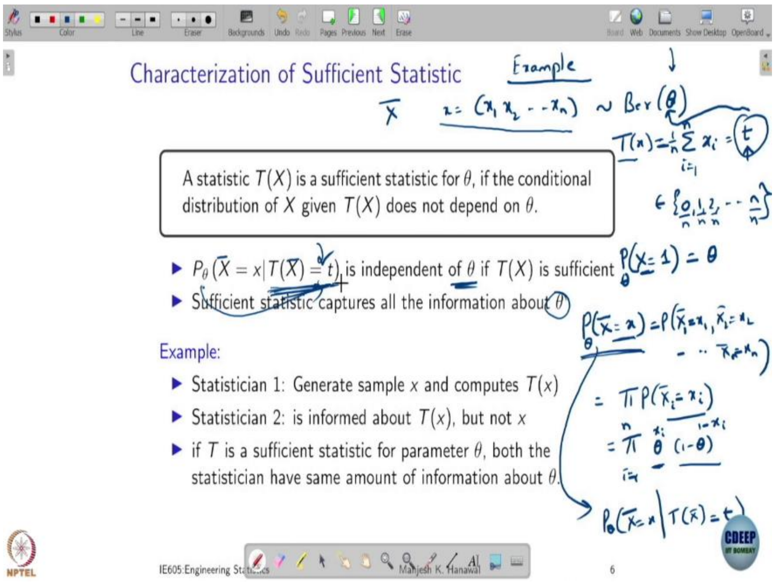
So, an example of the second case is suppose you have $x=(x_1,\ldots,x_n)$ and it is a statistic suppose, now, T(x) contains all information about theta what does that mean? If I give you any individual samples like let us say x3 or x4, they are not going to benefit you to get more information about theta.

Whatever T(x) telling you information about theta the individual samples are not complementing that. So, in summary, we are going to take sufficiency principle as if T(x) is a sufficient statistics for some parameter theta then any inference we are going to make about theta should depend on the sample X only through your T(X).

So, this is the sufficiency principle we are going to talk so, we are saying that the statistics should be enough to get information about that the data summary or the data reduction I am going to get through my statistic should be capturing enough information or what all the possible information about my parameter theta; that's it; any additional things will not help me. If that is happens, then we are going to say yes we are done we are following this sufficiency principle.

We already said that, if x and y are such that they have the same statistic value, then the inference about theta should be the same, whether it is x sample or y sample because, x and y individual does not matter to me, what matters is T(x) or T(y) in this case T(x) and T(y) are same. So, that means I have same amount of information about theta irrespective of whether it is x or y.

(Refer Slide Time: 05:52)



Characterization of Sufficient Statistic

Example

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

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

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

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

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

$P_\theta(X = 1) = \theta$

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

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

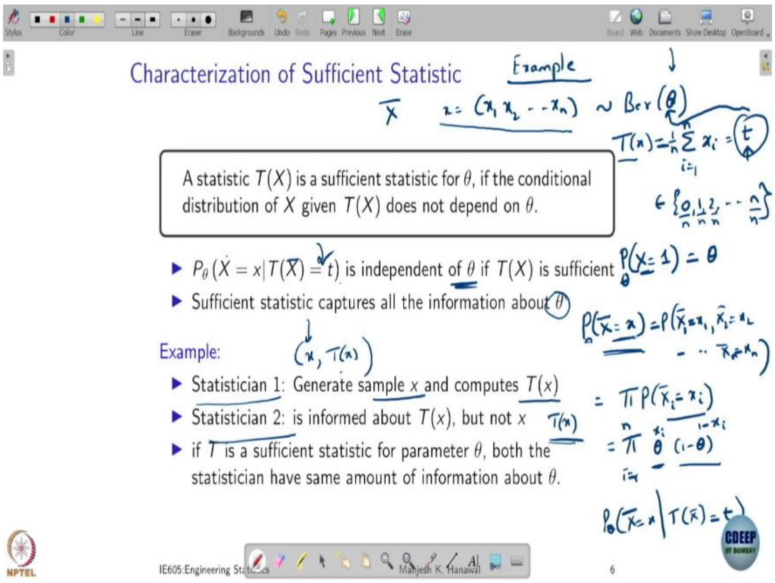
$= \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$

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

Example:

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

IE605:Engineering Statistics Manjesh K. Hanawal 6



Characterization of Sufficient Statistic

Example

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

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

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

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

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

$P_\theta(X = 1) = \theta$

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

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

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

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

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

- 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.

IE605:Engineering Statistics Manjesh K. Hanawal 6

Now, let us take an example. Let us take X_1, X_2,….X_n they are coming from Bernoulli parameter theta, everybody fine. Now, I am interested in some statistic which is summation of xi and we know that what is the possible values of T, T(x) we know that this is going to be between 0 to n, this is like a 0, 1, 2 up to n.

Now, my underlined parameter the unknown parameter here is theta. So, rather than this let me take it as yes whatever, let it be like this. Now, let us say that my x is going to be 1 what is this probability under parameter theta. So, x is this Bernoulli that is the populations it is going to be theta? x is going to be 1 is going to be theta. Now, maybe let me write this as 1 by n.

So, now this values are 1 by n, 2 by n, like this level, the value of T(x) is going to be this. Now, I am going to tell you this is x equals to 1 now let us say I am going to take this x this x this is a 1 value, maybe I should have. This is one I am going to make it vector now this is going to be x bar means vector. And let us call this x bar is let us say.

So, the first component x1 is x1 x bar of these 2 is x2 and like this xn bar equals to xn you all of you in segment three, so now x bar, let me call this as x bar, x bar is the sample. And this is like a one realisation of that random vector. We know that this is nothing but because we are talking about random samples, that means they are independent. So, this is like xi bar equals to xi. And we discussed this last time I can write it as i equals to 1 to n.

How can I write it? I can write it as theta xi, and 1 minus theta 1 minus xi let us do a sanity check when x equals to 1 I only get this term theta and when x equals to 0, I will get this term. So, this is the probability this is like an unconditional probability of observing sample x under my parameter theta, this is the probability, but now I am telling you T(X) has taken some value t, T(X) this is taken let us Say I am calling it here small t.

And we are also saying that this T is a sufficient statistic, I am as I am telling I am taking these are sufficient statistics, by that what we mean what is our intuition this T is a statistic is a data summary in such a way that this T contains all the necessary information about this theta. So, that means this T is now a proxy for this theta because T is containing all the necessary information about theta, so, I could as well interpret this T is a proxy for theta.

Now, if I tell that, this is my T(X) has taken value t, that means in a way like I am actually passing you the proxy for this theta and if this theta is indeed capturing essence, all the essential information about theta once I tell you this is value t, then they should this unconditional when I move from this unconditional to conditional one given that this T(X_bar) equals to t, this should be independent of theta.

And they should only depend on this t value, because now the role of theta is played by this t because it has captured the essential information about that t. So, that is why we are saying that conditional distribution of X and here I am using this notation X bar here just to distinguish one random variable from a random vector.

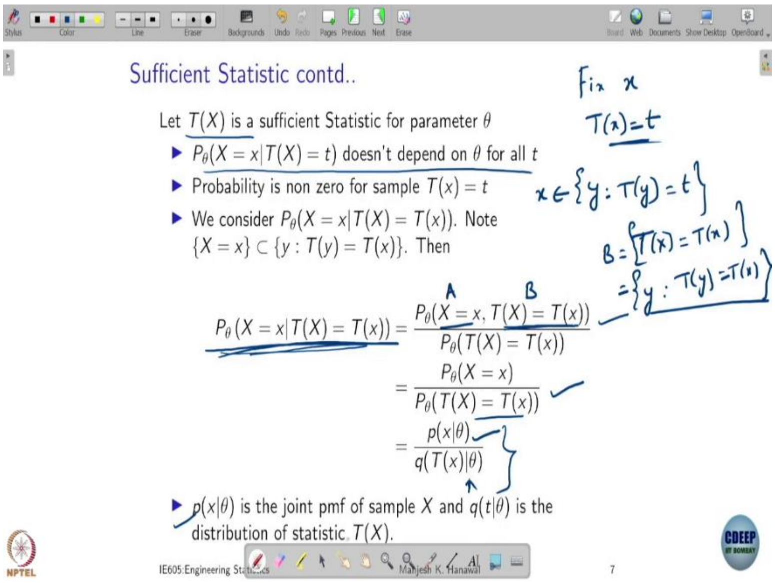
So, this conditional probability should be independent of theta. And that is because the sufficient statistics captures all the information about theta and now, I am telling you what is the theta true value t on which I am conditioning. So, now, we said that, let us take two statistician, one statistician has generated some sample x and he also compute T(x). So, statistician 1 has x as well as T(x) and the second statistician has only T(x).

Now, in terms of the parameter theta, who is better off between these two, 1 or 2, we said that both are same. Even though this guy has x as well as T(x), he is no better than the guy who knows only T(x), because having x if T is sufficient statistics, this having x is adding no more additional value about the parameter theta.

So, in this we are now talking about sufficiency principle when we say that when that is going to be sufficient, we are talking about what is the sufficiency principle is telling that sufficiency principle saying that, when you do a data reduction that reduced value will should contain information about theta.

Now, we are going to say it is going to be sufficient when whatever that reduction we are going to do it will contain a information rich enough so, that even if you go into give me any additional information about individual I am no better. Whatever is required is already contained in that reduction itself.

(Refer Slide Time: 14:33)



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. $T(X)$.

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, one quick thing we will do is quantify that. When we say that my conditional probability our goal is going to be independent of theta. Can I write it in a better way? Now let us assume T is a sufficient statistic. And now we said that if that is the case, this conditional probability does not depend on theta. And now let us say I am now going to compute I am going to fix an x arbitrary sample.

And my statistics is going to give me a value T on that. Now, let us try to compute the probability of conditional probability for this particular x conditioned on its statistic. Now, this is the conditional probability. What I have done here? I have simply applied the definition of conditional probability. That is the joint probability divided by the marginal.

Now we know that if I am going to let us take T(x) equals to t, and look into all the y's that will all the points that will map to the same t, we know that x is already going to take that value. So, x definitely belongs to this. So, because of this, now, if I am going to look into this event, so I will say let us call this event A and let us call this event B, B is my B event is T(X) is equals to T(x), that is means this is all y basically that are going to T(x) and we just know that this is A subset of B, because this is holds true. And now this is the case this this is x is going to take x and our parameter theta we are going to write it distribution of x I am enter parameter theta and the denominator is now the distribution of my statistic again under the parameter theta.

So, what we have done is this conditional probability we have segregated numerator is only on the distribution of x. And that determiner denominator is only on the distribution of your statistics, both under the parameter theta. Now, if this part is independent of theta, so much this ratio B, this ratio must be also independent of theta. So, I said, p(x) is the joint PM of your random samples and qt the denominator is the distribution of your statistic.