Sufficient Statistic

Vaishnavi W

IITH AI

AI20BTECH11025

Sufficient Statistic

Consider a random sample $X = \{X_1, X_2, \ldots, X_n\}$ with joint probability density given as $f(X_1, X_2, \ldots, X_n; \theta)$, associated with an unknown parameter θ

Let $T(x) = t(X_1, X_2, \ldots, X_n)$ be a statistic. It is said to be sufficient for θ if it contains all the information about the parameter θ.

Conditional probability of the data given $T(x)$,

$$f(X_1, X_2, \ldots, X_n; \theta | T(x)) \tag{1}$$

is independent of θ

Question

UGC/MATH 2018 (June math set-a), Q.108

Let X_1 and X_2 be i.i.d. with probability mass function $f_\theta(x) = \theta^x (1-\theta)^{1-x}$; $x = 0, 1$ where $\theta \in (0,1)$. Which of the following statements are true?

1. $X_1 + 2X_2$ is a sufficient statistic
2. $X_1 - X_2$ is a sufficient statistic
3. $X_1^2 + X_2^2$ is a sufficient statistic
4. $X_1^2 + X_2$ is a sufficient statistic

Solution

Given that, X_1 and X_2 are i.i.d. with probability mass function $f_\theta\left(x\right) = \theta^x\left(1-\theta\right)^{1-x}$; $x = 0, 1$ where $\theta \in (0, 1)$

$$f\left(x\right) = \begin{cases} (1-\theta) & x = 0 \\ \theta & x = 1 \end{cases} \tag{2}$$

Sufficient Statistic

Definition

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

According to the question, $T(x)$ is a sufficient statistic if the conditional probability,

$$\Pr(X_1 = x_1, X_2 = x_2 | T = t) \tag{3}$$

is independent of θ for all values of x_1, x_2 and t

Solution contd.

1. $T = X_1 + 2X_2$

Consider a case where $x_1 = 0, x_2 = 0$ and $t = 0$

$$\Pr(T = 0) = \Pr(X_1 + 2X_2 = 0) \tag{4}$$

$$= \Pr(X_1 = 0, X_2 = 0) \tag{5}$$

As X_1 and X_2 are independent

$$\Pr(T = 0) = \Pr(X_1 = 0)\Pr(X_2 = 0) = (1 - \theta)^2 \tag{6}$$

The conditional probability,

$$\Pr(X_1 = 0, X_2 = 0 | T = 0) = \frac{\Pr((X_1 = 0, X_2 = 0) \cap (T = 0))}{\Pr(T = 0)} \tag{7}$$

From (5), $(X_1 = 0, X_2 = 0) \subseteq (T = 0)$

$$= \frac{\Pr(X_1 = 0, X_2 = 0)}{\Pr(T = 0)} = \frac{(1 - \theta)^2}{(1 - \theta)^2} = 1 \tag{8}$$

Solution contd.

Similarly, conditional probabilities for other values of x_1, x_2 and t are given in table 3

x_1	x_2	t $t = X_1 + 2X_2$	Conditional probability $P_\theta\left(X_1 = x_1, X_2 = x_2 \middle| T = t\right)$
0	0	0 otherwise	1 0
1	0	1 otherwise	1 0
0	1	2 otherwise	1 0
1	1	3 otherwise	1 0

Table: Conditional Probabilities

All the conditional probabilities are independent of θ

$\therefore X_1 + 2X_2$ is a sufficient statistic.

Solution contd.

2. $T = X_1 - X_2$

Consider a case where $x_1 = 0, x_2 = 0$ and $t = 0$

$$\begin{aligned}
\Pr(T = 0) &= \Pr(X_1 - X_2 = 0) && (9)\\
&= \Pr(X_1 = 0, X_2 = 0) + \Pr(X_1 = 1, X_2 = 1) && (10)\\
&= \Pr(X_1 = 0)\Pr(X_2 = 0) + \Pr(X_1 = 1)\Pr(X_2 = 1) && (11)\\
&= (1-\theta)^2 + \theta^2 && (12)
\end{aligned}$$

Solution contd.

The conditional probability,

$$\Pr\left(X_1 = 0, X_2 = 0 \middle| T = 0\right) = \frac{\Pr\left(\left(X_1 = 0, X_2 = 0\right)\right) \cap \left(T = 0\right)}{\Pr\left(T = 0\right)} \tag{13}$$

From (10), $\left(X_1 = 0, X_2 = 0\right) \subseteq \left(T = 0\right)$

$$= \frac{\Pr\left(X_1 = 0, X_2 = 0\right)}{\Pr\left(T = 0\right)} = \frac{\left(1-\theta\right)^2}{\left(1-\theta\right)^2 + \theta^2} \tag{14}$$

depends on θ.
$\therefore X_1 - X_2$ is not a sufficient statistic.

Solution contd.

3. $T = X_1^2 + X_2$

Consider a case where $x_1 = 1, x_2 = 0$ and $t = 1$

$$\Pr(T = 1) = \Pr\left(X_1^2 + X_2 = 1\right) \tag{15}$$
$$= \Pr(X_1 = 0, X_2 = 1) + \Pr(X_1 = 1, X_2 = 0) \tag{16}$$
$$= \Pr(X_1 = 0)\Pr(X_2 = 1) + \Pr(X_1 = 1)\Pr(X_2 = 0) \tag{17}$$
$$= \theta(1-\theta) + (1-\theta)\theta = 2\theta(1-\theta) \tag{18}$$

The conditional probability,

$$\Pr(X_1 = 1, X_2 = 0 | T = 1) = \frac{\Pr((X_1 = 1, X_2 = 0) \cap (T = 1))}{\Pr(T = 1)} \tag{19}$$

From (16), $(X_1 = 1, X_2 = 0) \subseteq (T = 1)$

$$= \frac{\Pr(X_1 = 1, X_2 = 0)}{\Pr(T = 1)} = \frac{\theta(1-\theta)}{2\theta(1-\theta)} = \frac{1}{2} \tag{20}$$

Solution contd.

Similarly, conditional probabilities for other values of x_1, x_2 and t are given in table 3

x_1	x_2	t $t = X_1^2 + X_2$	Conditional probability $P_\theta\left(X_1 = x_1, X_2 = x_2 \middle| T = t\right)$
0	0	0 otherwise	1 0
1	0	1 otherwise	1/2 0
0	1	1 otherwise	1/2 0
1	1	2 otherwise	1 0

Table: Conditional Probabilities

All the conditional probabilities are independent of θ

$\therefore X_1^2 + X_2$ is a sufficient statistic.

Solution contd.

4. $T = X_1^2 + X_2^2$

The conditional probabilities for all values of x_1, x_2 and t are given in table

x_1	x_2	t $t = X_1^2 + X_2^2$	Conditional probability $P_\theta\left(X_1 = x_1, X_2 = x_2 \middle\| T = t\right)$
0	0	0 otherwise	1 0
1	0	1 otherwise	1/2 0
0	1	1 otherwise	1/2 0
1	1	2 otherwise	1 0

Table: Conditional Probabilities

All the conditional probabilities are independent of θ

$\therefore X_1^2 + X_2^2$ is a sufficient statistic.