Sufficient Statistics

Anuradha Uggi

Department of Electrical Engineering
Indian Institute of Technology Hyderabad

April 23,2021

Contents

Sufficient Statistic and Fisher Fcatorization

- Sufficient Statistic holds information needed to compute unknown.
- Distribution of X conditioned on θ,Statistic can be any $T(X)$.
- Sufficiency is stated mathematically by Fisher factorization theorem.
- $f_X(x) = h(x)g(\theta, T(x))$
- $h(x)$ is any constant and $g(\theta, T(x))$ tells θ interact to X only through $T(X)$.

Bernoulli Distribution

- Let $X = \{X_1, X_2, .., X_n\}$ be i.i.d $Bernoulli(p)$.
- Let $T(X) = X_1 + X_2 + .. + X_n$ which is sum of $1's$.
- $\Pr(X = x) = \Pr\{X_1 = x_1, X_2 = x_2, ..., X_n = x_n\}$
- above can be written as
- $p^{x_1}(1-p)^{1-x_1}p^{x_2}..p^{x_n}(1-p)^{1-x_n}$
- $p^{\sum x_i}(1-p)^{n-\sum x_i} = p^{T(X)}(1-p)^{n-T(X)}$
- $h(x) = 1$ and $g(p, T(x)) = p^{T(X)}(1-p)^{n-T(X)}$
- $T(X)$ is alone enough to find p

Uniform Distribution

- Let $X = \{X_1, X_2, .., X_n\}$ be i.i.d and Uniformly Distributed over $(0, \theta)$.
- $T(X) = \max(X_1, X_2, ..., X_n)$
- $f_X(x_1, x_2, .., x_n) = \frac{1}{\theta} 1_{\{0 \leq x_1 \leq \theta\}} \cdots \frac{1}{\theta} 1_{\{0 \leq x_n \leq \theta\}}$
- $f_X(x_1, x_2, .., x_n) = \frac{1}{\theta^n} 1_{\{0 \leq \min x_i\}} 1_{\{\max x_i \leq \theta\}}$
- $h(x) = 1_{\{0 \leq \min x_i\}}$ remaining part is $g(x, T(x))$
- Therefore $T(X)$ is sufficient to find θ.

Poisson Distribution

- Let $X = X_1, X_2, .., X_n$ be i.i.d Poisson(λ)
- $T(X) = X_1 + X_2 + .. + X_n$ for λ
- $\Pr(X = x)$=$\Pr(X_1 = x_1, X_2 = x_2, .., X_n = x_n)$.
- $\Pr(X = x)$=$\frac{e^{-\lambda}\lambda^{x_1}}{x_1!}\frac{e^{-\lambda}\lambda^{x_2}}{x_2!}..\frac{e^{-\lambda}\lambda^{x_n}}{x_n!}$.
- $e^{-n\lambda}\lambda^{(x_1+x_2+..x_n)}\frac{1}{x_1!x_2!..x_n!}$
- $h(x) = \frac{1}{x_1!x_2!..x_n!}$
- $g(x, T(x))$=$e^{-n\lambda}\lambda^{T(X)}$

Question

- Suppose $X_i = X_1, X_2, ..., X_n$ are i.i.d Uniform $(\theta, 2\theta), \theta > 0$. Let $X_{(1)}$=min$\{X_1, X_2, ..., X_n\}$ and $X_{(n)}$=max$\{X_1, ..., X_n\}$.then which of the following statements are correct.
    1. ($X_{(1)}$,$X_{(n)}$) is jointly sufficient and complete for θ
    2. ($X_{(1)}$,$X_{(n)}$) is jointly sufficient but not complete for θ
    3. $\frac{X_{(n)}}{2}$ is maximum likelihood estimate for θ
    4. $X_{(1)}$ is maximum likelihood estimate for θ

Solution

- $T(X) = \{X_{(1)}, X_{(n)})\}$.
- $f_X(x) = \frac{1}{\theta^n} 1_{\{\theta \leq X_{(1)}\}} 1_{\{X_{(n)} \leq 2\theta\}}$.
- $h(x) = 1$.
- $g(x, T(x) = \frac{1}{\theta^n} 1_{\{\theta \leq X_{(1)}\}} 1_{\{X_{(n)} \leq 2\theta\}}$.
- $T(X)$ is sufficient to find θ.

$1_{\{\theta \leq X_{(1)}\}} 1_{\{X_{(n)} \leq 2\theta\}}$

- Let $x_1, x_2, ..., x_n$ are random samples from random variables $X_1, X_2, ..X_n$.

$$1_{\{\theta \leq x \leq 2\theta\}:} = \begin{cases} 1 & \text{if } \theta \leq x \leq 2\theta \\ 0 & \text{otherwise} \end{cases}$$

- Above equation implies that every random sample lies in the defined range.
- It can be generalized by taking max and min among all sampled values.
- i.e $\max\{X\} \leq 2\theta$ always and
- $\theta \leq \min\{X\}$ always

Complete Statistic

- T(X) is said to be complete for θ if for every measurable function g,
- if $E_\theta(g(T)) = 0$ for all θ then
- $P_\theta(g(T) = 0) = 1$ for all θ.

Complete Statistic

- Let $g(T) = X_{(1)} - X_{(n)}$
- $E[g(T)]{=}E[X_{(1)} - X_{(n)}] = c$ for all θ
- $E[X_{(1)} - X_{(n)} - c] = 0$ for all θ
- $\int (X_{(1)} - X_{(n)} - c)\frac{1}{\theta^n} dx = 0$
- We can say that $X_{(1)} - X_{(n)} - c = 0$ for all θ
- Therefore $T(X)$ is Sufficient and Complete for θ

MLE of θ

- Estimate of θ which Maximizes Likelihood i.e $f_X(x)$.
- $f_X(x) = \frac{1}{\theta^n}$ where θ and $f_X(x)$ are inversely related.so Minimum value of θ maximizes the Likelihood.
- and we also know that $\theta \leq X_{(1)}$ and $X_{(n)} \leq 2\theta$
- i.e $\frac{X_{(n)}}{2} \leq \theta \leq X_{(1)}$
- Therefore MLE of θ=$\frac{X_{(n)}}{2}$

References

- https://en.wikipedia.org/wiki/Sufficient_statistic
- https://en.wikipedia.org/wiki/Completeness_(statistics)

Thank you