North Asian International Research Journal of Multidisciplinary

CIF IMPACT FACTOR:4.465

ISSN: 2454-2326 Vol. 7, Issue-11 November-2021

Index Copernicus Value: 58.12 DOI NUMBER: 10.5948/2454-2326.2020.00008.0 Thomson Reuters ID: S-8304-2016

A Peer Reviewed Refereed Journal

ON NEYMAN'S FACTORIZATION THEOREM CONCERNING A SUFFICIENT STATISTICS FOR A PARAMETER

BILAL AHMAD BHAT*

****Division of Social Science, Faculty of Fisheries, S.K. University of Agricultural Sciences and Technology Kashmir (J&K)***

ABSTRACT

In this paper, it is shown that if $X_1, X_2, \ldots, X_n$ denote a random sample of size n from a distribution that has pdf $f(x; \theta)$, $\theta \in \Omega$, then the statistics $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ is a sufficient statistics for θ if

$$f(x_1; \theta)\, f(x_2; \theta) \ldots f(x_n; \theta) = \Psi(y_1; \theta)$$

where $\Psi(y_1; \theta)$ is a function of y_1 and θ alone.

KEYWORDS: Best statistics, unbiased statistic, sufficient statistic and Fisher-Neyman Criterion.

1. INTRODUCTION AND STATEMENT OF RESULTS

A famous result known as Factorization Theorem of Neyman [1, 2, 3] concerning a sufficient statistic for a parameter θ can be stated as follows:

THEOREM A. Let $X_1, X_2, \ldots, X_n$ denote a random sample of size n from a distribution that has pdf $f(x; \theta)$, $\theta \in \Omega$. The statistic $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ is a sufficient statistics for parameter θ if and only if we can find two nonnegative functions, Φ_1 and Φ_2, such that

$$f(x_1; \theta)\, f(x_2; \theta) \ldots f(x_n; \theta) = \Phi_1(u_1(X_1, X_2, \ldots, X_n); \theta)\, \Phi_2(X_1, X_2, \ldots, X_n),$$

where for every fixed value of $y_1 = u_1(x_1, x_2, \ldots, x_n)$, $\Phi_2(X_1, X_2, \ldots, X_n)$ does not depend upon θ.

In order to verify that a certain statistic $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ is a sufficient statistic for a parameter θ with the help of factorization theorem of Neyman, we have to express the joint pdf of $X_1, X_2, \ldots, X_n$ as product of two nonnegative functions Φ_1 and Φ_2 where $\Phi_2(X_1, X_2, \ldots, X_n)$ does not depend upon θ for every fixed value of $y_1 = u_1(x_1, x_2, \ldots, x_n)$. Here we establish an elegant criteria pertaining to a sufficient statistic of a parameter θ. More precisely, we prove:

THEOREM 1. Let $X_1, X_2, \ldots, X_n$ denote a random sample of size n and let $Y_1 = u_1(X_1, X_2, \ldots, X_n)$. If

$$f(x_1; \theta)\, f(x_2; \theta) \ldots f(x_n; \theta) = \Psi(y_1; \theta)$$

where $\Psi(y_1; \theta)$ is a function of y_1 and θ alone, then $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ is a sufficient statistic for θ.

We illustrate Theorem 1 with the help of the following examples:

EXAMPLE 1. Let $X_1, X_2, \ldots, X_n$ denote a random sample of size n from a distribution having the pdf

$$f(x; \theta) = \theta \exp(-\theta x),\ 0 < x < \infty,\ \theta > 0;$$
$$= 0 \text{ elsewhere.}$$

Consider the statistic $Y_1 = X_1 + X_2 + \ldots + X_n$. Since the joint pdf of $X_1, X_2, \ldots, X_n$ is

$$f(x_1; \theta)\, f(x_2; \theta) \ldots f(x_n; \theta) = \theta^{-\theta x}{}_1\, \theta^{-\theta x}{}_2 \ldots \theta^{-\theta x}{}_n$$

$$= \theta^n e^{-(x_1 + x_2 + \ldots + x_n)}$$

$$= \theta^n e^{-\theta y_1}$$

$$= \Psi(y_1; \theta) \qquad \text{(say)}$$

where $\Psi(y_1; \theta)$ is a function of y_1 and θ only. Hence by Theorem 1, $y_1 = \sum_{i=1}^{n} x_i$ is a sufficient statistic for θ.

EXAMPLE 2. Let $X_1, X_2, \ldots, X_n$ denote a random sample of size n from a distribution having the pdf

$$f(x; \theta) = \theta x^{\theta-1},\ 0 < x < 1,\ \theta > 0;$$
$$= 0 \text{ elsewhere.}$$

By theorem 1, it can be easily seen that the statistic $y_1 = x_1, x_2, \ldots, x_n$ is a sufficient statistics of θ because

$$f(x_1; \theta)\, f(x_2; \theta) \ldots f(x_n; \theta) = \Psi(y_1; \theta)$$

$$= \theta^n y_1^{\theta-1}.$$

REMARK 1. The advantage of Theorem 1 is that it tells us at a first glance that a statistic $y_1 = u_1(x_1, x_2, \ldots, x_n)$ is a sufficient statistic for a parameter θ if the product

$$f(x_1; \theta)\, f(x_2; \theta) \ldots f(x_n; \theta)$$

is a function of y_1 and θ only.

2. PROOF OF THE THEOREM

PROOF OF THE THEOREM 1. We shall prove the theorem when the random variables are of the continuous type. In case of random variables of the discrete type, we have to replace integrals by sums only. Suppose that

(1) $f(x_1; \theta)\, f(x_2; \theta) \ldots f(x_n; \theta) = \Psi[u_1(X_1, X_2, \ldots, X_n); \theta]$.

Consider one-to-one transformation $y_1 = u_1(x_1, x_2, \ldots, x_n)$, $y_2 = u_2(x_1, x_2, \ldots, x_n)$, …, $y_n = u_n(x_1, x_2, \ldots, x_n)$. Let the inverse functions of this transformation be $x_1 = w_1(y_1, y_2, \ldots, y_n)$, $x_2 = w_2(y_1, y_2, \ldots, y_n)$, ... , $x_n = w_n(y_1, y_2, \ldots, y_n)$. If

$$J = \frac{\partial(x_1, x_2, \ldots, x_n)}{\partial(y_1, y_2, \ldots, y_n)}$$

denotes the Jacobian of Transformation, then the joint pdf of the statistics $Y_1, Y_2, \ldots, Y_n$ is

$g(y_1, y_2, \ldots, y_n; \theta) = f(w_1(y_1, y_2, \ldots, y_n; \theta))\, f(w_2(y_1, y_2, \ldots, y_n; \theta)) \ldots f(w_n(y_1, y_2, \ldots, y_n; \theta))|J|$

(2) $= \Psi(y_1; \theta)|J|$.

The marginal pdf of $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ is given by

$$g(y_1; \theta) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(y_1, y_2, \ldots, y_n; \theta)\, dy_2\, dy_3 \ldots dy_n$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \psi(y_1; \theta)\, |J|\, dy_2\, dy_3 \ldots dy_n$$

(3) $$= \psi(y_1; \theta) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |J|\, dy_2\, dy_3 \ldots dy_n .$$

Since the Jacobian J is a function of $Y_1, Y_2, \ldots, Y_n$ and θ is neither involved in the Jacobian nor in the limits of the integration, therefore, the (n-1) fold integral

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |J|\, dy_2\, dy_3 \ldots dy_n$$

is a function of y_1 alone say $\Phi(y_1)$. Thus from (3), we obtain

$$g(y_1; \theta) = \psi(y_1; \theta)\, |\, \phi(y_1) .$$

If $\phi(y_1) = 0$, then $g(y_1; \theta) = 0$. If $\phi(y_1) > 0$, we can write

$$\psi(u_1(x_1, x_2, ..., x_n); \theta) = \frac{g(u_1(x_1, x_2, ..., x_n); \theta)}{\phi(u_1(x_1, x_2, ..., x_n))}$$

(4) $$= g(u_1(x_1, x_2, ..., x_n); \theta)\psi(u_1(x_1, x_2, ..., x_n))$$

where $\psi(u_1(x_1, x_2, ..., x_n)) = \dfrac{1}{\phi(u_1(x_1, x_2, ..., x_n))}$

does not depend upon $\theta \in \Omega$ for every fixed value of $y_1 = u_1(x_1, x_2, \ldots, x_n)$. Using (1) in (4) we get

$$f(x_1;\theta) f(x_2;\theta)... f(x_n;\theta) = g(u_1(x_1, x_2, ..., x_n); \theta)\psi(u_1(x_1, x_2, ..., x_n))$$

or $$\frac{f(x_1;\theta) f(x_2;\theta)... f(x_n;\theta)}{g(u_1(x_1, x_2, ..., x_n); \theta)} = \psi(u_1(x_1, x_2, ..., x_n))$$

where $\psi(u_1(x_1, x_2, ..., x_n))$ does not depend upon $\theta \in \Omega$ for every fixed value of $y_1 = u_1(x_1, x_2, \ldots, x_n)$. Hence by definition of sufficient statistic, it follows that

$$y_1 = u_1(x_1, x_2, ..., x_n)$$

is a sufficient statistic for θ. This completes the proof of Theorem 1.

Acknowledgement: *The author would like to thank Professor A. Aziz-ul.Azeem, Ex. Dean Academics, Head, P.G. Dept. of Mathematics and Statistics, University of Kashmir for his guidance during the preparation of this paper and elder brother Mr. Mohd Altaf Bhat for his moral support in tough times. Recently, we lost this great Mathematician due to COVID-19 infection, may Allah grant him highest place in Jannah. This result I am dedicating to my teacher A. Aziz-ul.Azeem and brother Mr. Mohammad Altaf.*

REFERENCES

[1] Fisher, R.A. (1922): On the mathematical foundation of theoretical statistics. Philosophical Transactions of the Royal Society of London, Series A, 222A, 309-368.

[2] Fisher, R.A. (1925): Theory of statistical estimation. Proceedings of the Cambridge Philosophical Society, 22, 700-725.

[3] Neyman, J. (1935): Su un teorema concernente le considdette statistiche sufficienti. Giornale dell' Istitulo Italiano degli Attuari, 6, 320-334.

Professor A.Azizul-Auzeem former Dean Academics and Head, P.G. Department Mathematics and Statistics, University of Kashmir, the first PhD of Kashmir University in Mathematics always guided/supported author as his own younger brother/child. He was among the top mathematicians of Kashmir whose research work was recognized at world level. We lost this great personality due to COVID-19 infection recently. I thank sir from the core of my heart for his guidance and pray Almighty Allah to grant him highest place in Jannah along with my dear ones. Sir, you will remain in my heart always

Mr. Mohammad Altaf always supported author in research activities whether it was in JK, AMU Aligarh, Delhi University or ISI Calcutta. He used to author help always in field surveys in JK or outside. My words simply fail to thank you brother for your love and support always.

About Author:

Dr. Bilal Ahmad Bhat, is currently working as Associate Professor (Statistics) and Head, Division of Social Sciences, Faculty of Fisheries, SKUAST-Kashmir J&K since August, 2015. Dr. Bilal Ahmad (Phd Statistics, M.Sc. Mathematics, M.Sc. Statistics, Diploma in Foreign language (French), B.Ed.) was earlier working in Mathematics & Statistics Dept University of Kashmir and in SKUAST-Jammu as Assistant Professor

(Statistics) before joining SKUAST-Kashmir. He has organized National as well as International conference (s) and has participated and presented a number of research papers in various national/international conferences. He has published more than 220 research papers in various National/International Journals of repute and has published more than 400 popular articles. He has good knowledge of statistical software's like SPSS, MINITAB, SAS, R and has guided a number of research scholars for their PhD programme. His field of research is probability theory, information theory and applied statistics. Dr. Bilal Ahmad is a member of advisory board in some well known National and International journals of repute.