Jñānābha, Vol. 14, 1984

THE DISCREPANCY MEASURE: A CHARACTERIZATION

by

PARVINDER KAUR

Department of Statistics, Lucknow University,

Lucknow-226 007, India

(*Received : March 23, 1983; Revised: March 27, 1984*)

ABSTRACT

H. Kaufman, A. M. Mathai and P. N. Rathie [1] and the author [2] characterized the *discrepancy measure*, which is a well-known and useful concept in statistical problems. Here this discrepancy measure is characterized with the help of a functional equation.

1. Introduction

In statistical tests of goodness of fit and many other problems, a statistic usually employed is Pearson's χ^2 (Chi-square) statistic which is defined as

$$\chi^2 = \sum_{i=1}^{k} [(n_i - np_i)^2/np_i],\ n = \sum_{i=1}^{k} n_i, \tag{1.1}$$

where n_i and np_i are the observed and expected frequencies correspon-ding to the *i-th* group, respectively, and p_i is the probability of getting an observation in *i–th* group. Equation (1.1) can be simplified to the form

$$\chi^2 = n\,[\sum_{i=1}^{k} q_i^2/p_i - 1],\ q_i = n_i/n. \tag{1.2}$$

Proof. Let us first solve the functional equation (2.1) under the condition (2.2). Let k, t, r, s be any positive integers such that $1 \leqslant k \leqslant r$, $1 \leqslant t \leqslant s$. Setting $x_i = 1/k$, $u_i = 1/r$, $y_j = 1/t$, $v_j = 1/s$ $(i = 1, 2, \ldots, k;\ j = 1, 2, \ldots, t)$ in (2.1), we get

$$kt\, H(1/kt, 1/rs) = kt(1/k)^{-1}\,(1/r)^2\, H(1/t, 1/s) + kt(1/t)\, H(1/k, 1/r), \tag{2.4}$$

that is,

$$H(ab, cd) = a^{-1}\, c^2\, H(b, d) + bH(a, c), \tag{2.5}$$

where $a = 1/k$, $b = 1/t$, $c = 1/r$, $d = 1/s$.

Also

$$H(ab, cd) = H(ba, dc), \tag{2.6}$$

giving

$$a^{-1}\, c^2\, H(b, d) + bH(a, c) = b^{-1}\, d^2\, H(a, c) + aH(b, d), \tag{2.7}$$

so that

$$\frac{H(a, c)}{a^{-1}\, c^2 - a} = \frac{H(b, d)}{b^{-1}\, d^2 - b} = A(say), \tag{2.8}$$

where A is a constant. Therefore

$$H(a, c) = A(a^{-1}\, c^2 - a) \tag{2.9}$$

To extend this solution to the case of rational $a, c \in [0, 1]$,

let $x = k/t$ $(k < t)$ and $u = p/q$ $(p < q)$ be two rational numbers. Let m be any positive integer sufficiently large so that $mp \geqslant k$, $mq \geqslant t$,

$m \geqslant q(t-k)/t(q-p)$.

Taking k as $t - k + 1$ and t as k, and setting

The quantity

$$\chi^2/h = \sum_{i=1}^{k} q_i^2/p_i - 1 \tag{1.3}$$

is defined to be a measure of discrepancy between two discrete populations

$$P = (p_1, p_2, \ldots, p_k) \text{ and } Q = (q_1, q_2, \ldots, q_k);\ p_i, q_i > 0,$$

$$\sum_i q_i = \sum_i p_i = 1.$$

Kaufman, Mathai and Rathie characterized the quantity (1.2) using recursivity property and also by assuming a particular structure. The author proposed and characterized a generalized measure which leads to the characterization of (1.2). In this paper the quantity (1.2) is characterized with the help of a functional equation.

2. Characterization

Theorem *Let H: $[0,1]$ x $[0,1] \rightarrow R$ (reals) be a continuous function. For every positive k and t, if H satisfies the functional equation*

$$\sum_{i=1}^{k} \sum_{j=1}^{t} H(x_i, y_j, u_i v_j) = \sum_{i=1}^{k} \sum_{j=1}^{t} x_i^{-1} u_i^2 H(y_j, v_j)$$

$$+ \sum_{i=1}^{k} \sum_{j=1}^{t} y_j H(x_i, u_i) \tag{2.1}$$

then, under the condition that

$$H(1/4, 3/4) = 4/3, \tag{2.2}$$

$$\sum_{i=1}^{k} H(p_i, q_i) = [\sum_{i=1}^{k} \frac{q_i^2}{p_i} - 1]. \tag{2.3}$$

$$\begin{cases} x_1 = k/t,\ x_2 = \dots\ x_{t-k+1} = 1/t \\ y_1 = y_2 = \dots = y_k = 1/k \\ u_1 = p/q,\ u_2 = \dots = u_{t-k+1} = 1/mt \\ v_1 = v_2 = \dots = v_k = 1/pm \end{cases} \tag{2.10}$$

in (2.1), we get

$$kH(1/t,\ 1/qm) + k(t-k)\ H(1/kt,\ 1/ptm^2) = k(k/t)^{-1}\ (p/q)^2\ H(1/k,\ 1/pm) +$$

$$+\ k(t-k)\ (1/t)^{-1}\ (1/mt)^2\ H(1/k,\ 1/pm)\ +\ k(1/k)\ [\ H(k/t,\ p/q)\ +$$

$$+\ (t-k)\ H(1/t,\ 1/mt)\ ] \tag{2.11}$$

which, with (2.3), gives

$$H(k/t,\ p/q) = A[\ (k/t)^{-1}\ (p/q)^2 - (k/t)\ ] \tag{2.12}$$

or, equivalently,

$$H(x,\ u) = A[x^{-1}\ u^2 - x] \tag{2.13}$$

for all rationals $x,\ u \in\ ]0,\ 1]$.

Since H is continuous, (2.13) is valid for all real $x,\ u \in [0,\ 1]$.

Now the condition (2.2) with (2.13) gives $A = 1$ so that

$$H(x,\ u) = [\ x^{-1}\ u^2 - x] \tag{2.14}$$

Therefore

$$\sum_{i=1}^{k} H(\ p_i,\ q_i) = \sum_{i=1}^{k} \ (p_i^{-1}\ q_i^2 - p_i)$$

$$= \sum_{i=1}^{k} \ q_i^2/p_i - 1, \tag{2.15}$$

which is precisely (1.1). This completes the proof of our theorem.

REFERENCES

[1] H. Kaufman, A. M. Mathai and P. N. Rathie, A mathematical foundation for Pearson's χ^2 goodness of fit statistic, *Sankhyā Ser. A* **34** (1973), 441–442.

[2] P. Kaur, A combined information measure, *Pure Appl. Math Sci.* **18** (1983), 59–65.