The Annals of Statistics
1977, Vol. 5, No. 6, 1245–1248

ON A CHARACTERIZATION OF RIDITS

BY PATRICK L. BROCKETT AND ARNOLD LEVINE

Tulane University

In this paper characterization theorems for Bross' ridits are proved with the help of functional equations by assuming intuitively reasonable postulates.

1. Introduction. Bross (1958) proposed a method for assigning quantitative values to ranked categorical data. This method is now widely used, especially in epidemiological studies (Williams and Grizzle (1972)). His assignment function is as follows: let p_i, $i = 1, \cdots, k$, denote the empirical proportion of N items which fall into the ith ranked category when there are k categories; then

$$\tfrac{1}{2}(\sum_{j<i} p_j - \sum_{j>i} p_j) + .5$$

is the value (called the ridit value) assigned to category i.

Bross' paper was designed to elucidate the use of ridits and so did not contain the mathematical ideas underlying its derivation. Bross promised to provide this derivation but it never appeared. Since ridit analysis is an important tool of statistical analysis, it is useful to characterize mathematical structures which lead to this assignment method. In this paper we develop two different sets of postulates which lead essentially to ridits. We justify each postulate in terms of characteristics desired for an assignment function.

2. Postulates on an assignment function leading to ridits. Let $\mathbf{p} = (p_1, p_2, \cdots, p_k)$ be the empirical probability distribution of the N items over k ranked response categories, $p_i \geqq 0$, $\sum_{i=1}^{k} p_i = 1$. Based upon the empirical probabilities $\mathbf{p}$ we wish to assign a weight to each category. It is reasonable to assume that the assignment function should depend upon the number of categories present, so we shall write $h_k(i, \mathbf{p})$ for the weight assigned to category i when $\mathbf{p}$ is the empirical distribution and there are k response categories.

POSTULATE 1. $h_1(1, 1) = 0$.

This postulate corresponds to the belief that if there is only one response, then there can be no ordering possible and we cannot distinguish a response pattern.

POSTULATE 2. $0 \leqq h_2(2, p, 1 - p) = -h_2(1, 1 - p, p)$.

This assumption reflects the idea that if the empirical distribution over the two responses is reversed, then by symmetry, the *absolute* numerical values assigned to the two responses should be switched right along with the switch in the empirical distribution. However, to preserve the ranked character of the responses, the sign should change. Hence $0 \leqq h_2(2, p, 1 - p) = -h_2(1, 1 - p, p)$.

Received April 11, 1977.

AMS 1970 *subject classification.* Primary 62H05.

Key words and phrases. Ridits, mathematical characterizations.

POSTULATE 3 (branching property). Suppose there are more than two responses and for reasons of a statistical or computational nature we wish to combine two adjacent response categories. We assume that the unaffected response categories retain their same assigned values, and the value given to the new combined category is the weighted average of the values given to the original response categories. Symbolically, this says that if the i and $(i+1)$st categories are combined, and $\mathbf{p} = (p_1, p_2, \cdots, p_k)$ is the original empirical distribution and $\mathbf{q} = (q_1, \cdots, q_{k-1})$ is the new empirical distribution ($q_t = p_t$ for $t < i$, $q_i = p_i + p_{i+1}$, $q_t = p_{t+1}$ $t > i$), then

$$
\begin{aligned}
h_{k-1}(i, \mathbf{q}) &= \frac{p_i}{p_i + p_{i+1}} h_k(i, \mathbf{p}) + \frac{p_{i+1}}{p_i + p_{i+1}} h_k(i+1, \mathbf{p}) \\
h_{k-1}(t, \mathbf{q}) &= h_k(t, \mathbf{p}) \quad t \leqq i - 1 \qquad (1) \\
h_{k-1}(t, \mathbf{q}) &= h_k(t+1, \mathbf{p}) \quad t \geqq i + 1
\end{aligned}
$$

This assumption usually simplifies computational work in data reduction efforts.

POSTULATE 4. If there are two categories, then $h_2(2, p, 1-p) - h_2(1, p, 1-p)$ is nondecreasing as p increases.

This assumption reflects the desired characteristic that the assigned values for two categories should not become closer as the difference in the proportion of respondents in each category becomes greater; that is, the weighting should be consistent with the empirical response pattern.

With these reasonable postulates we obtain the following characterization theorem.

THEOREM 1. *An assignment function h_k satisfies Postulates* 1–4, *if and only if* $h_k(i, \mathbf{p}) = c(\sum_{j<i} p_j - \sum_{j>i} p_j)$, *where c is an arbitrary constant.*

PROOF. Define $f(p) = h_2(2, p, 1-p)$, $0 \leqq p \leqq 1$. Then Postulates 1, 2 and 3 together imply

$$0 = h_1(1, p + 1 - p) = pf(1-p) - (1-p)f(p)\,. \qquad (2)$$

Let $p = (1+t)/2$ and define $\phi(t)$ as $\phi(t) = \phi(2p-1) = (2/(1+t))f((1+t)/2) = (1/p)f(p)$. One easily calculates $\phi(-t) = (1/(1-p))f(1-p)$ so that (2) implies that ϕ is an even function of t, i.e.,

$$f(p) = p\phi(|2p-1|)\,. \qquad (3)$$

Now let $x = \sum_{j<i} p_j$ and $y = \sum_{j>i} p_j$. Applying (1) of Postulate 3 we note that

$$h_k(i, \mathbf{p}) = h_3(2, x, p_i, y) \qquad p_i > 0\,. \qquad (4)$$

We now calculate

$$
\begin{aligned}
h_3(3, x, p_i, y) &= h_2(2, x + p_i, y) = f(x + p_i) \qquad (5) \\
&= (x + p_i)\phi(|2(x + p_i) - 1|) = (x + p_i)\phi(|x + p_i - y|)
\end{aligned}
$$

and

$$h_3(1, x, p_i, y) = h_2(1, x, p_i + y) = -h_2(2, p_i + y, x) = -f(y + p_i) = -(y + p_i)\phi(|y + p_i - x|) \tag{6}$$

where we have used Postulates 2 and 3 and equation (3).

From Postulates 3 and 1 we see that $0 = xh_3(1, x, p_i, y) + p_i h_3(2, x, p_i, y) + yh_3(3, x, p_i, y)$ or equivalently

$$h_3(2, x, p_i, y) = \frac{-xh_3(1, x, p_i, y) - yh_3(3, x, p_i, y)}{p_i}\,. \tag{7}$$

Utilizing (5) and (6) together with (4) in (8) yields

$$h_k(i, \mathbf{p}) = \frac{x(1 - x)\phi(|y + p_i - x|) - y(1 - y)\phi(|x + p_i - y|)}{p_i}\,. \tag{8}$$

Now since $\phi(t) = \phi(|2p - 1|)$ is an even function about $p = \frac{1}{2}$, it follows that $\phi(|2p - 1|) = h_2(2, p, 1 - p) - h_2(1, p, 1 - p)$ cannot satisfy Postulate 4 unless ϕ is constant. Call this constant c, and plug into (8) to obtain

$$h_k(i, \mathbf{p}) = \frac{x(1 - x)c - y(1 - y)c}{p_i} = c\left[\frac{x(y + p_i) - y(x + p_i)}{p_i}\right] = c(x - y)\,,$$

that is,

$$h_k(i, \mathbf{p}) = c[\textstyle\sum_{j<i} p_j - \sum_{j>i} p_j]\,. \qquad \square$$

Note that this is not exactly the formula obtained by Bross, but a rather simple translation of such. The formula obtained here is more suitable for weighting of the categories obtained from a ranked categorical response questionnaire since, in such a situation, one wishes to weight responses according to their ability to distinguish between groups responding to the questionnaire (see Brockett and Levine [1], [4]).

We may also establish a set of axioms which lead directly to Bross' ridits.

THEOREM 2. *Let h be a real-valued function satisfying*

(a) *h is continuous*

(b) *(branching property) $h(x, y) = \lambda h(x, y + (1 - \lambda)c) + (1 - \lambda)h(x + \lambda c, y)$ for any x, y, c, and λ.*

Then $h(x, y) = a + b(x - y)$ for some choice of a and b.

PROOF. Let us first assume $f \in C^1$. Differentiating (b) with respect to λ yields

$$0 = h(x, y + (1 - \lambda)c) - h(x + \lambda c, y) - \lambda c h_y(x, y + (1 - \lambda)c) + (1 - \lambda)c h_x(x + \lambda c, y) \tag{9}$$

where h_x and h_y represent the partial derivatives of h with respect to x and y respectively.

Setting $\lambda = 1$ in (9) we find $-h_y(x, y) = (h(x + c, y) - h(x, y))/c$ so that $-h_y(x, y) = h_x(x, y)$, i.e., $(d/ds)h(x + s, y + s) \equiv 0$. Integrating from $-y$ to 0 yields $h(x, y) = h(x - y, 0) = \phi(x - y)$. Placing $\phi(x - y)$ into (9) and setting $\lambda = 0$, $x = y$ we find $\phi(-c) = -c\phi'(0) + \phi(0)$. Upon setting $\lambda = 1$ and $x = y$ we find $\phi(c) = c\phi'(0) + \phi(0)$, so that $\phi(z) = a + bz$, a and b constants. Thus the conclusion holds when $h \in C^1$.

Suppose now that h is only continuous, and let

$$h_n(x, y) = \iint h(x + s, y + t)\omega_{1/n}(s, t)\, ds\, dt \tag{10}$$

where ω_r is a mollifier, i.e., $\omega_r \in C^\infty$ and $\omega_r(x, y) = 0$ if $x^2 + y^2 > r^2$, and $\iint \omega_r(x, y)\, dx\, dy = 1$. Using hypothesis (b) on h in (10) we find that h_n also satisfies (b) and is C^1, hence $h_n(x, y) = a_n + b_n(x - y)$. Since $h_n \to h$, it follows that $h(x, y) = a + b(x - y)$ also.

REMARKS 1. If we additionally assume $h(x, y) = -h(y, x)$ in Theorem 2, then $h(x, y) = b(x - y)$.

2. Define the assignment function $h_k(i, \mathbf{p}) = h(\sum_{j<i} p_j, \sum_{j>i} p_j)$ with h satisfying the conditions in Theorem 2 and we obtain another characterization of the ridit assignment function.

3. The hypothesis (a) in Theorem 2 is desired since small changes in the empirical distribution should cause only small changes in the assigned values. Hypothesis (b) is a branching condition which is justified in the same way as in Postulate 3 of Theorem 1.

REFERENCES

[1] BROCKETT, PATRICK L. and LEVINE, ARNOLD (1977). Analysis of categorical questionnaire data by use of pridits. Unpublished manuscript.

[2] BROSS, IRWIN (1958). How to use ridit analysis. *Biometrics* **14** 18–38.

[3] FORTHOFER, RONALD N. and KOCH, GARY G. (1973). An analysis for compounded functions of categorical data. *Biometrics* **29** 143–157.

[4] LEVINE, A. (1972). A mathematical method for analysing questionnaires. *Bull. World Health Organization* vol. **47**, no. 1.

[5] WILLIAMS, O. DALE and GRIZZLE, JAMES E. (1972). Analysis of contingency tables having ordered response categories. *J. Amer. Statist. Assoc.* **67** 55–63.

DEPARTMENT OF MATHEMATICS
TULANE UNIVERSITY
NEW ORLEANS, LOUISIANA 70118