ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
XII, 1 (1971)

R. BARTOSZYŃSKI (Warszawa)

ON THE CONSTRUCTION AND EVALUATION OF SUBJECTIVE CLASSIFICATIONS

0. Introduction. The aim of this paper is to present some theoretical results concerning the methods of construction and evaluation of subjective classifications, and also, to present the practical application of the suggested theory to a certain empirical problem. Though the theoretical concepts introduced in this paper may be applied not only to the particular empirical problem which inspired their introduction and analysis, it seems best to start from a short outline of the questions posed by practice.

The National Bank of Poland (NBP) replaces systematically used banknotes by new ones. The banknotes replaced are those which enter a branch of NBP performing such a replacement (there are several hundreds of such branches) and which are deemed as used up in a degree exceeding the admissible norm (in NBP terminology such banknotes are called "destructs").

The norms of admissible waste are defined rather vaguely; in practice, the bank employees whose principal duty is to count the packages of banknotes reject from these packages those banknotes which, according to their subjective judgment, should be withdrawn from the circulation, and replace them by banknotes taken out of a special package. Elimination of destructs is only a secondary task of these employees: they are interested primarily in the correct counting the total number of banknotes in packages, as they are financially responsible for the correctness of this counting.

As mentioned above, the criteria which distinguish destructs from non-destructs are rather vague and at present cannot be changed in any controllable manner. The bank instruction specifies that one should reject the banknotes which are torn, have distinct stains, or are "too much used up". While the first categories are defined in a more or less satisfactory manner, the basic difficulty lies in the lack of a sufficiently precise definition of the "degree of normal waste" beyond which the banknotes should be eliminated as destructs. Generally, the bank would like to create the possibility of introducing a flexible policy of replacement by which the "critical level of waste" beyond which the banknotes are eliminated could be changed by decisions of NBP. Such a flexible policy in an obvious prerequisite for any subsequent research aimed at the choice of a policy being optimal from the point of view of suitable criteria.

Thus, to create the possibility for search of an optimal policy of replacement, it is necessary to formulate objective criteria for the concept of destructs, criteria which could be changed in a controllable way; moreover, it is necessary to create methods of objective evaluation of various policies of replacement in terms of the quality of banknotes in circulation. These two problems are closely related and will be treated jointly.

For technical reasons one cannot use any criteria of the degree of waste of banknotes other than those for subjective evaluations. Hence, measuring the number of creases crossing a given line, amount of light absorbed, or other more or less obvious indices, are eliminated from among the considerations at the beginning. Therefore all concepts introduced in this paper will be based solely on subjective evaluations by persons performing the classifications, and the estimates of parameters introduced will be based on observations of those subjective classifications. In short, we will try to create methods of objective evaluations and of objective control of subjective classifications.

1. Preliminaries. We postpone the consideration of classifications to the following sections and begin with presenting two simple theorems which will be used in Section 2.

Consider a system $\mathcal{G} = \{G_1, \ldots, G_n\}$ of n independent experiments. Assume that each experiment can lead either to "success" or to "failure" and let a_i be the (unknown) probability of success in the experiment G_i. Our problem consists in constructing methods of inference about the probabilities a_i in situations where for some reasons one is allowed to make not more than two independent observations of each experiment G_i. These methods of inference will be given in the form of estimates of the quantities

$$\bar{a} = \frac{1}{n}\sum_{i=1}^{n} a_i \quad \text{and} \quad \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(a_i - \bar{a})^2$$

which characterize to some extent the vector $(a_1, \ldots, a_n)$ corresponding to the system $\mathcal{G}$.

Assuming two independent observations of each experiment G_i, let X_i be 1 or 0 depending on whether the first trial in the experiment G_i

resulted in success or not, and let Y_i be 1 or 0 depending on whether the second trial in the experiment G_i resulted in success or not. Thus, if the first and the second trials are independent, and if different experiments G_i are independent, the random variables in the pairs (X_i, Y_i), (X_i, X_j), (Y_i, Y_j) and (X_i, Y_j) are independent for $i \neq j$.

Write

$$Z_{ij} = X_i Y_j$$

and set

$$U = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad V = \frac{1}{n}\sum_{i=1}^{n} Y_j,$$

$$L = \frac{1}{2}(U+V), \qquad W = \frac{1}{n}\sum_{i=1}^{n} Z_{ii} - UV.$$

We prove

THEOREM 1. *The random variable L is an unbiased estimate of parameter $\bar{a}$, and the random variable W is an unbiased estimate of parameter σ^2. Moreover, $D^2 L \leqslant 1/8n$, and $D^2 W \leqslant a_n \sim 3/4n$.*

Proof. By construction, U and V are independent and equally distributed. We have $EX_i = P\{X_i = 1\} = a_i = EY_i$, hence

$$EU = EV = \frac{1}{n}\sum a_i = \bar{a},$$

which implies that $EL = \bar{a}$. Next, $EZ_{ii} = EX_i Y_i = EX_i EY_i = a_i^2$, and consequently,

$$EW = \frac{1}{n}\sum EZ_{ii} - EUV = \frac{1}{n}\sum a_i^2 - EUEV = \frac{1}{n}\sum a_i^2 - \bar{a}^2 = \sigma^2,$$

which completes the proof of unbiasedness.

Now, $D^2 X_i = D^2 Y_i = a_i(1-a_i) \leqslant 1/4$, and using the assumption of independence we can write

$$D^2 U = D^2 V = \frac{1}{n^2}\sum D^2 X_i \leqslant 1/4n;$$

consequently,

$$D^2 L \leqslant \frac{1}{4}(D^2 U + D^2 V) \leqslant 1/8n.$$

Evaluation of the variance of W is somewhat messy. We write

$$\begin{aligned}
D^2W &= D^2\left(\frac{1}{n}\sum Z_{ii}-UV\right)\\
&= \frac{1}{n^2}D^2\left(\sum Z_{ii}\right)+D^2UV-\frac{2}{n}\operatorname{Cov}\left(\sum Z_{ii}, UV\right)\\
&= \frac{1}{n^2}\sum D^2Z_{ii}+\frac{1}{n^4}D^2\left(\sum X_i \sum Y_j\right)-\\
&\qquad -\frac{2}{n^3}\operatorname{Cov}\left(\sum Z_{ii}, \sum X_j \sum Y_k\right)\\
&= \frac{1}{n^2}\sum D^2Z_{ii}+\frac{1}{n^4}D^2\left(\sum Z_{ij}\right)-\frac{2}{n^3}\operatorname{Cov}\left(\sum Z_{ii}, \sum Z_{jk}\right)\\
&= \frac{1}{n^2}\sum D^2Z_{ii}+\frac{1}{n^4}\sum \operatorname{Cov}(Z_{ij}, Z_{km})-\frac{2}{n^3}\sum \operatorname{Cov}(Z_{ii}, Z_{km})\\
&= \frac{1}{n^2}\sum D^2Z_{ii}+\frac{1}{n^4}\sum_{\substack{i\neq j\\ k\neq m}} \operatorname{Cov}(Z_{ij}, Z_{km})-\\
&\qquad -\left(\frac{1}{n^3}-\frac{1}{n^4}\right)\left[\sum \operatorname{Cov}(Z_{ii}, Z_{jk})+\operatorname{Cov}(Z_{ij}, Z_{kk})\right]\\
&= S_1+S_2-S_3.
\end{aligned}$$

Now, we have

$$D^2Z_{ii} = P\{Z_{ii}=1\}[1-P\{Z_{ii}=1\}] = a_i^2(1-a_i^2) \leqslant 1/4,$$

hence the first term, S_1, is bounded from above by $1/4n$. In the second sum all terms vanish except those corresponding to systems of indices of the form $(i,j;i,j)$ with $i\neq j$ and either $(i,j;i,k)$ or $(i,j;k,i)$ with i,j,k all distinct. For $i\neq j$ we have

$$\operatorname{Cov}(Z_{ij}, Z_{ij}) = D^2Z_{ij} = a_ia_j(1-a_ia_j) \leqslant 1/4.$$

Next, for i,j,k all distinct, we have

$$\begin{aligned}
\operatorname{Cov}(Z_{ij}, Z_{ik}) &= EZ_{ij}Z_{ik}-EZ_{ij}EZ_{ik} = EX_i^2Y_jY_k-EX_iEY_jEX_iEY_k\\
&= EY_jEY_kD^2X_i = a_ja_ka_i(1-a_i) \leqslant 1/4.
\end{aligned}$$

Now, the number of systems of indices of the form $(i,j;i,j)$ with $i\neq j$ is $n(n-1)$; the number of systems of indices of the form $(i,j;i,k)$ or $(i,j;k,i)$ with distinct i,j,k is $2n(n-1)(n-2)$. Thus, the term S_2 is bounded from above by

$$\frac{1}{4}\frac{1}{n^4}[n(n-1)+2n(n-1)(n-2)] \sim \frac{2}{4n}.$$

To obtain the desired upper bound of order $3/4n$ for D^2W it remains to prove that all terms in the sum S_3 are non-negative. Now, all terms in S_3, corresponding to systems $(i, i; j, k)$ with $j \neq i$, $k \neq i$, vanish. There remain covariances of the form $\mathrm{Cov}(Z_{ii}, Z_{ii})$ equal to $D^2 Z_{ii} \geqslant 0$ and covariances of the form $\mathrm{Cov}(Z_{ii}, Z_{ij})$ with $i \neq j$. We have

$$\begin{aligned}\mathrm{Cov}(Z_{ii}, Z_{ij}) &= EZ_{ii}Z_{ij} - EZ_{ii}EZ_{ij} = EX_i^2 Y_i Y_j - EX_i Y_i EX_i Y_j \\ &= a_j[EX_i^2 EY_i - (EX_i)^2 EY_i] = a_j(a_i^2 - a_i^3) \geqslant 0.\end{aligned}$$

Thus, $S_3 \geqslant 0$ which completes the proof of Theorem 1.

We state for further reference that the last covariance is bounded from above by $4/27$, i.e. that for $i \neq j$ we have

$$\mathrm{Cov}(Z_{ii}, Z_{ij}) \leqslant 4/27. \tag{1}$$

Suppose now that system $\mathscr{S}$ is partitioned into two disjoint subsystems $\mathscr{S}_1$ and $\mathscr{S}_2$ consisting of $n_1 > 0$ and $n_2 > 0$ experiments respectively (so that $n_1 + n_2 = n$). Define

$$\bar{a}^{(1)} = \frac{1}{n_1}\sum\nolimits_1 a_i, \qquad \bar{a}^{(2)} = \frac{1}{n_2}\sum\nolimits_2 a_i,$$

$$\sigma_1^2 = \frac{1}{n_1}\sum\nolimits_1 (a_i - \bar{a}^{(1)})^2, \qquad \sigma_2^2 = \frac{1}{n_2}\sum\nolimits_2 (a_i - \bar{a}^{(2)})^2,$$

where for simplicity, $\sum_1$ and $\sum_2$ denote sums extended over indices i from subsystems $\mathscr{S}_1$ and $\mathscr{S}_2$, respectively.

Using standard analysis of variance partitioning we shall devise a test for the hypothesis that $\bar{a}^{(1)} = \bar{a}^{(2)}$, i.e. for the hypothesis that the mean probabilities of success are the same in both subsystems $\mathscr{S}_1$ and $\mathscr{S}_2$. Let U_1, V_1, W_1 and U_2, V_2, W_2 be defined as before for subsystems $\mathscr{S}_1$ and $\mathscr{S}_2$, that is let

$$U_1 = \frac{1}{n_1}\sum\nolimits_1 X_i, \qquad V_1 = \frac{1}{n_1}\sum\nolimits_1 Y_i, \qquad W_1 = \frac{1}{n_1}\sum\nolimits_1 Z_{ii} - U_1 V_1,$$

and similarly for U_2, V_2 and W_2. The random variable W will be defined as before for the whole system $\mathscr{S} = \mathscr{S}_1 \cup \mathscr{S}_2$. We prove

THEOREM 2. *The random variable*

$$K = W - \frac{n_1}{n}W_1 - \frac{n_2}{n}W_2$$

has a non-negative expectation, $EK \geqslant 0$. *The equality* $EK = 0$ *holds if and only if* $\bar{a}^{(1)} = \bar{a}^{(2)}$. *Moreover, for* $\min(n_1, n_2) \to \infty$ *we have* $D^2K \leqslant \beta_n \sim 145/54n$.

Proof. Partitioning the expression for σ^2, we get

$$\sigma^2 = \frac{1}{n}\sum (a_i-\bar{a})^2$$
$$= \frac{1}{n}\sum\nolimits_1 (a_i-\bar{a}^{(1)}+\bar{a}^{(1)}-\bar{a})^2+\frac{1}{n}\sum\nolimits_2 (a_i-\bar{a}^{(2)}+\bar{a}^{(2)}-\bar{a})^2$$
$$= \frac{1}{n}\sum\nolimits_1 (a_i-\bar{a}^{(1)})^2+\frac{1}{n}\sum\nolimits_2 (a_i-\bar{a}^{(2)})^2+\frac{n_1}{n}(\bar{a}^{(1)}-\bar{a})^2+\frac{n_2}{n}(\bar{a}^{(2)}-\bar{a})^2$$
$$= \frac{n_1}{n}\sigma_1^2+\frac{n_2}{n}\sigma_2^2+\frac{n_1}{n}(\bar{a}^{(1)}-\bar{a})^2+\frac{n_2}{n}(\bar{a}^{(2)}-\bar{a})^2.$$

By Theorem 1, W, W_1 and W_2 are unbiased estimates of σ^2, σ_1^2 and σ_2^2, and we can write

$$EK = EW-\frac{n_1}{n}EW_1-\frac{n_2}{n}EW_2 = \frac{n_1}{n}(\bar{a}^{(1)}-\bar{a})^2+\frac{n_2}{n}(\bar{a}^{(2)}-\bar{a})^2 \geqslant 0$$

with the equality sign holding if and only if $\bar{a}^{(1)} = \bar{a}^{(2)}$, as in this case $\bar{a} = \bar{a}^{(1)} = \bar{a}^{(2)}$. To evaluate the variance of K note that W_1 and W_2 are independent, hence

$$D^2K = D^2W+\frac{n_1^2}{n^2}D^2W_1+\frac{n_2^2}{n^2}D^2W_2-2\frac{n_1}{n}\operatorname{Cov}(W, W_1)-2\frac{n_2}{n}\operatorname{Cov}(W, W_2).$$

For large n_1 and n_2, the sum of the first three terms can be bounded from above by

$$\frac{3}{4n}+\frac{n_1^2}{n^2}\cdot\frac{3}{4n_1}+\frac{n_2^2}{n^2}\cdot\frac{3}{4n_2} = \frac{3}{4n}+\frac{3(n_1+n_2)}{4n^2} = \frac{3}{2n}.$$

Next, we need bounds from below for $\operatorname{Cov}(W, W_1)$ and $\operatorname{Cov}(W, W_2)$. We have

$$\operatorname{Cov}(W, W_1) = \operatorname{Cov}\left(\frac{1}{n}\sum Z_{ii}-UV, \frac{1}{n_1}\sum\nolimits_1 Z_{jj}-U_1V_1\right)$$
$$= \frac{1}{nn_1}\operatorname{Cov}\left(\sum Z_{ii}, \sum\nolimits_1 Z_{jj}\right)+\frac{1}{n^2n_1^2}\operatorname{Cov}\left(\sum Z_{ij}, \sum\nolimits_1 Z_{km}\right)-$$
$$-\frac{1}{nn_1^2}\operatorname{Cov}\left(\sum Z_{ii}, \sum\nolimits_1 Z_{jk}\right)-\frac{1}{n^2n_1}\operatorname{Cov}\left(\sum\nolimits_1 Z_{ii}, \sum Z_{jk}\right).$$

From the proof of Theorem 1 it follows that all covariances of the form $\operatorname{Cov}(Z_{ij}, Z_{km})$ are non-negative, thus we can concentrate on the last two terms of the above sum only. We have

$$\operatorname{Cov}\left(\sum Z_{ii}, \sum\nolimits_1 Z_{jk}\right) = \operatorname{Cov}\left(\sum\nolimits_1 Z_{ii}, \sum\nolimits_1 Z_{jk}\right) = \sum\nolimits_1 \operatorname{Cov}(Z_{ii}, Z_{jk}),$$

since for i in $\mathscr{G}_2$, j and k in $\mathscr{G}_1$, Z_{ii} and Z_{jk} are independent.

Now, in the last sum all terms for which both j and k differ from i are zero; terms for which exactly one of indices j, k is equal to i are, by (1), bounded by $4/27$, and all terms with $j = k = i$ are bounded by $1/4$. The number of the latter terms is n_1; the number of terms with exactly one of the indices j, k equal to i is $2n_1(n_1-1)$. We have therefore

$$\frac{1}{nn_1^2}\operatorname{Cov}\left(\sum Z_{ii}, \sum\nolimits_1 Z_{jk}\right) \leqslant \frac{1}{nn_1^2}\left[\frac{n_1}{4}+\frac{4}{27}2n_1(n_1-1)\right] \sim \frac{8}{27n}.$$

Finally, let us consider the covariance $\operatorname{Cov}\left(\sum_1 Z_{ii}, \sum Z_{jk}\right)$. We may write it in form of a sum,

$$\operatorname{Cov}\left(\sum\nolimits_1 Z_{ii}, \sum\nolimits_1 Z_{jk}\right)+\operatorname{Cov}\left(\sum\nolimits_1 Z_{ii}, \sum\nolimits' Z_{jk}\right),$$

where $\sum'$ denotes the sum extended over pairs j, k for which at least one term is outside of $\mathscr{G}_1$. The first covariance is bounded by

$$\frac{n_1}{4}+\frac{4}{27}2n_1(n_1-1)$$

as before. In the second sum, the terms with both j and k outside of $\mathscr{G}_1$ vanish. If only one index, say j, is in $\mathscr{G}_1$ and k is in $\mathscr{G}_2$, then the covariance is zero unless $j = i$. In the latter case it is bounded by $4/27$, as stated in (1). The number of such terms is obviously equal to $2n_1n_2$. Thus we have

$$\frac{1}{n^2n_1}\operatorname{Cov}\left(\sum\nolimits_1 Z_{ii}, \sum Z_{jk}\right)$$

$$\leqslant \frac{1}{n^2n_1}\left[\frac{n_1}{4}+\frac{4}{27}2n_1(n_1-1)+\frac{4}{27}2n_1n_2\right] \sim \frac{8}{27}\left[\frac{1}{n}\cdot\frac{n_1}{n}+\frac{1}{n}\cdot\frac{n_2}{n}\right] = \frac{8}{27n}.$$

Hence $\operatorname{Cov}(W, W_1)$ is bounded from above by $16/27n$, and, by symmetry, the same applies also to $\operatorname{Cov}(W, W_2)$. Combining these evaluations we can write

$$D^2K \leqslant \beta_n \sim \frac{3}{2n}+2\frac{n_1}{n}\cdot\frac{16}{27n}+2\frac{n_2}{n}\cdot\frac{16}{27n} = \frac{145}{54n},$$

which completes the proof.

Theorem 2 can be extended without any change to the case of testing for homogeneity of a set of systems $\mathscr{G}_1, \ldots, \mathscr{G}_k$. In fact, let $\mathscr{G} = \mathscr{G}_1 \cup \ldots \cup \mathscr{G}_k$ with $\mathscr{G}_i \cap \mathscr{G}_j = \emptyset$ for $i \neq j$ and let $\mathscr{G}_i$ consist of $n_i > 0$ experiments. Put $n = n_1+\ldots+n_k$. Define $\bar{a}^{(i)}$, σ_i^2 and the estimates U_i, V_i, W_i in a manner analogous to that used before. We have

THEOREM 3. *The random variable*

$$K = W-\frac{1}{n}\sum_{j=1}^{k} n_jW_j$$

satisfies the condition $EK \geqslant 0$ *and* $EK = 0$ *if and only if* $\bar{a}^{(1)} = \bar{a}^{(2)} = \ldots$ $\ldots = \bar{a}^{(k)}$. *Moreover, for* $\min(n_1, \ldots, n_k) \to \infty$ *we have* $D^2K \leqslant \beta_n \sim 145/54n$.

The constant 145/54 can be further improved if one takes into account the fact that if terms of the form $a_i a_j a_k(1-a_k)$ are close to their maxima equal to 1/4, the terms involving $a_i(1-a_i)$ and $a_j(1-a_j)$ must be close to zero.

2. Evaluation of schemes of classification. In this section we show how the theorems of Section 1 may be used for constructing methods of evaluation of schemes of classification. Informally, by *classification* we shall understand the act of assigning elements of a given set (called *categories*) to elements of another set (of classified objects). Contrarily to the approach presented in [2], we make no assumption about the existence of a "true" category for any given object. Consequently, we shall evaluate classifications by means of some parameters characterizing interindividual and intraindividual variability. The underlying idea is that a good classification scheme satisfies the following informal requirement: if the same set of objects is classified twice (by different individuals or by the same individual), a majority of objects is assigned to the same category on both occasions.

It should be remarked that the above requirement constitutes only a necessary condition for a classification scheme to be "good", and it is by no means sufficient. However, as any sufficient condition must be based on the concept of "true" category for a given object and consists of requiring that the average number of "wrong" classifications is small in some sense or other, such conditions lie beyond the scope of this paper.

Formally, suppose that we are given a non-empty set B whose elements will be called *classified objects*, a non-empty set S whose elements will be called *individuals making classifications*, and a finite or countable set $\mathscr{C} = \{C_0, C_1, \ldots\}$ whose elements will be called *categories of classification*. To avoid trivialities, we assume that the set $\mathscr{C}$ contains at least two elements.

Given B, S and $\mathscr{C}$, by a *classification scheme* we shall mean a family of random variables (defined on some fixed probability space)

$$\{\xi_s^{(i)}(b),\ b \in B,\ s \in S,\ i = 1, 2, \ldots\}$$

which assume values in $\mathscr{C}$, and such that:

1° if $(b, s, i) \neq (b', s', i')$, then the random variables $\xi_s^{(i)}(b)$ and $\xi_{s'}^{(i')}(b')$ are independent;

2° for any $b \in B$ and $s \in S$, the random variables $\xi_s^{(i)}(b)$, $i = 1, 2, \ldots$, have the same distribution.

We shall interpret $\{\xi_s^{(i)}(b) = C_j\}$ as the event "on ith trial individual s classified object b into category C_j".

Write

$$P\{\xi_s^{(i)}(b) = C_j\} = p_{s,j}(b);$$

by condition 2°, the latter quantity is independent of index i.

For $b \in B$ and $s_1, s_2 \in S$, write

(2) $$u_{s_1,s_2}(b) = \sum_j p_{s_1,j}(b) p_{s_2,j}(b).$$

The quantity $u_{s_1,s_2}(b)$ will be used as our measure of quality of a classification scheme (with respect to object b and individuals s_1 and s_2); it is non-negative and attains its maximal value 1 if and only if the distributions $\{p_{s_1,j}(b)\}$ and $\{p_{s_2,j}(b)\}$ are identical and degenerate (i.e. concentrated at one value of j).

In practical situations, the number of available independent classifications of the same object by the same individual is limited, primarily by learning effects. Thus, we shall use methods of Section 1 for estimation of quantities $u_{s_1,s_2}(b)$; as we shall see, it will be sufficient to take 2 or 4 observations of classifications by a given individual depending on whether $s_1 \neq s_2$ or $s_1 = s_2$.

Let $Q = \{b_1, \ldots, b_n\}$ be a finite subset of B, and let s_1 and s_2 be fixed elements of S (not necessarily distinct). Consider the following scheme of experiments $\{G_1, \ldots, G_n\}$: experiment G_i leads to success on the first trial if $\xi_{s_1}^{(1)}(b_i) = \xi_{s_2}^{(2)}(b_i)$; otherwise it leads to failure. Similarly, G_i leads to success on the second trial, if $\xi_{s_1}^{(3)}(b_i) = \xi_{s_2}^{(4)}(b_i)$; otherwise it leads to failure.

These definitions may appear somewhat artificial, but they allow to avoid tedious distinguishing of cases $s_1 = s_2$ and $s_1 \neq s_2$. As we see, in the first case one needs four observations, while in the second case one needs only two observations for each individual and each element $b \in Q$.

Clearly, by the assumed independence, the probability of success in experiment G_i is $u_{s_1,s_2}(b_i)$; moreover, different experiments and successive trials on the same experiment are independent. Thus, we are in the position to apply Theorem 1 of Section 1 to estimate the quantities

$$m_{s_1,s_2}(Q) = \frac{1}{n} \sum_{j=1}^{n} u_{s_1,s_2}(b_j)$$

and

$$\sigma^2_{s_1,s_2}(Q) = \frac{1}{n} \sum_{j=1}^{n} [u_{s_1,s_2}(b_j) - m_{s_1,s_2}(Q)]^2.$$

We can also apply Theorems 2 and 3 of Section 1 to design various tests for "homogeneity". We shall sketch the construction of two of them.

Suppose that each individual $s_1, \ldots, s_m$ $(m > 1)$ of set S performs four independent classifications of the same set $Q = \{b_1, \ldots, b_n\}$. In other words, we observe values of the random variables $\xi_{s_j}^{(i)}(b_k)$ for $j = 1, \ldots, m$, $k = 1, \ldots, n$ and $i = 1, 2, 3, 4$. We may now define the family $\mathcal{G} = \{G_{jk}, j = 1, \ldots, m, k = 1, \ldots, n\}$ of experiments defining "success" and "failure" on the first and the second trials in experiment G_{jk} depending (in the same manner as above) on the results of four classifications of object b_k by individual s_j. We may now proceed in two ways. First, by splitting the system $\mathcal{G}$ into subsystems $\mathcal{G}_{1.}, \mathcal{G}_{2.}, \ldots, \mathcal{G}_{m.}$ corresponding to different individuals and applying Theorem 3, we are able to test the hypothesis that the average

$$m_{s_j,s_j}(Q) = \frac{1}{n}\sum_{k=1}^{n} u_{s_j,s_j}(b_k)$$

is the same for all individuals s_j. Intuitively, this test would tell us whether all individulas in question are equally "reliable" in their classifications (with respect to the set Q).

We may also proceed differently and split $\mathcal{G}$ into subsystems $\mathcal{G}_{.1}$, $\mathcal{G}_{.2}, \ldots, \mathcal{G}_{.n}$ corresponding to different objects from Q. Theorem 3 could provide a test which (for large m) would inform us about the existence of "odd" objects in Q, i.e. objects which are significantly easier or significantly more difficult to classify. More precisely, we could test the hypothesis that all averages of the form

$$\frac{1}{m}\sum_{j=1}^{m} u_{s_j,s_j}(b_k), \quad k = 1, \ldots, n,$$

were equal.

In a similar manner, we may partition the set $\{s_1, \ldots, s_m\}$ into pairs (say, by random metching) and define experiment G_{jk} accordingly, depending on the results of classifications of object b_k by the jth pair of individuals. As before, two ways of splitting the system $\{G_{jk}\}$ so obtained lead to two tests: one, which would inform us whether among the classifying individuals there are such ones which "deviate" systematically from the rest of the group in their classifications, and the other, telling whether there exist objects which are either classified significantly "more unanimously" than others or cause significantly greater differences of opinion than others.

3. Construction of classification schemes. In this section we shall show that under certain conditions we can construct classification schemes satisfying some desirable properties. Intuitively, we shall assume that objects of the set B possess some quantitative property (either directly measurable, such as size, or a latent one, such as utility), and we shall assume that individuals from set S, when confronted with a pair of objects $a, b \in B$, are able to point out one of them as having "less" of this property. If ties are not allowed in these judgments, we may expect inconsistencies and, in general, randomness in these judgments; we shall see, however, that a relatively weak requirement of over-all consistency will suffice for constructing a reasonably good classification scheme.

In short, the construction of the classification scheme will be based on a suitably selected "standard sequence" of elements of B and will consist of assigning category C_j to all elements of B which are judged to fall "between" the jth and $(j+1)$th terms of this sequence. Our basic idea will be to construct the longest possible standard sequence satisfying the consistency property; with probability 1 no element of B will be judged "earlier" than the jth element and "later" than kth element of the sequence for $k > j$.

Formally, we shall assume that we are given a family of random variables

$$\{T_s^{(i)}(a, b),\ (a, b) \in B \times B,\ s \in S,\ i = 1, 2, \ldots\},$$

such that $T_s^{(i)}(a, b)$ assumes one of the values a or b and satisfying the following properties:

1° if $\big(s, i, (a, b)\big) \neq \big(s', i', (a', b')\big)$, then $T_s^{(i)}(a, b)$ and $T_{s'}^{(i')}(a', b')$ are independent;

2° for every $(a, b) \in B \times B$ the random variables $T_s^{(i)}(a, b)$, $s \in S$, $i = 1, 2, \ldots$, have the same distribution.

We shall interpret $\{T_s^{(i)}(a, b) = a\}$ as the event "on the ith presentation of pair (a, b) individual s pointed out to a as having "less" of the considered property than b".

Write

$$P\{T_s^{(i)}(a, b) = a\} = p(a, b);$$

by 2°, the last quantity does not depend on s and i.

We shall assume that the probabilities $p(a, b)$, $(a, b) \in B \times B$, satisfy the following axioms (¹):

(i) SYMMETRY. *For any $a, b \in B$ we have*

$$p(a, b) + p(b, a) = 1.$$

(ii) TRANSITIVITY. *For any $a, b, c \in B$, if $p(a, b) \geqslant 1/2$ and $p(b, c) \geqslant 1/2$, then*

$$p(a, c) \geqslant \max[p(a, b), p(b, c)].$$

(iii) PERFECT DISTINGUISHABILITY. *There exists a certain q satisfying the condition $\frac{1}{2} < q < 1$ such that for all $a, b, c \in B$, if $p(a, b) > q$ and $p(b, c) > q$, then $p(a, c) = 1$.*

(¹) This is a slight modification of the set of axioms given in [1].

To formulate the next axiom, write for fixed $a \in B$ and fixed h such that $\frac{1}{2} < h < 1$:

$$A_h^+(a) = \{x \in B\colon p(a, x) \geqslant h\},$$
$$A_h^-(a) = \{y \in B\colon p(y, a) \geqslant h\}.$$

(iv) CLOSURE. *For every $a \in B$ and every h such that $\frac{1}{2} < h < 1$:*

I. *If $A_h^+(a)$ is non-empty, then there exists a $u = u(h, a) \in A_h^+(a)$ such that $p(u, x) \geqslant 1/2$ for all $x \in A_h^+(a)$.*

II. *If $A_h^-(a)$ is non-empty, then there exists a $v = v(h, a) \in A_h^-(a)$ such that $p(y, v) \geqslant 1/2$ for all $y \in A_h^-(a)$.*

(v) ARCHIMEDEAN PROPERTY. *For every $a \in B$ and every h such that $\frac{1}{2} < h < 1$:*

I. *If there exists an infinite sequence $b_0, b_1, b_2, \ldots$ of elements of B satisfying the condition $p(b_j, b_{j+1}) \geqslant h$ for all j, then there exists $m = m(a)$ such that $p(a, b_m) \geqslant 1/2$.*

II. *If there exists an infinite sequence $c_0, c_1, c_2, \ldots$ of elements of B satisfying the condition $p(c_{j+1}, c_j) \geqslant h$ for all j, then there exists $n = n(a)$ such that $p(c_n a) \geqslant 1/2$.*

Note first that axiom (i) is not implied by the fact that the random variable $T_s^{(i)}(a, b)$ equals either a or b; indeed, as we consider ordered pairs (a, b), there is no a priori reason why the distributions of $T_s^{(i)}(a, b)$ and $T_s^{(i)}(b, a)$ should be related one to the other in any way.

Next, let Q be the set of q's satisfying the conditions of axiom (iii). Clearly, if $q \in Q$ and $q < q' < 1$, then $q' \in Q$. Write

$$q^* = \inf\{q\colon q \in Q\}. \tag{3}$$

Since strict inequalities are required in (iii), we have $q^* \in Q$, which implies that $Q = \langle q^*, 1\rangle$.

We shall precede the main theorems by some preparatory propositions.

Define: $a \sim b$ if $p(a, b) = \frac{1}{2}$. We have

PROPOSITION 1. *Relation $\sim$ is an equivalence relation in B.*

Proof. By (i), relation $\sim$ is reflexive and symmetric. Suppose that $a \sim b$, and $b \sim c$, i.e. $p(a, b) = p(b, c) = \frac{1}{2}$. By (ii) we obtain $p(a, c) \geqslant \frac{1}{2}$. Next, by (i), we have also $p(c, b) = p(b, a) = \frac{1}{2}$, hence, by (ii), we get $p(c, a) \geqslant \frac{1}{2}$. Using (i) again, we obtain $p(a, c) = \frac{1}{2}$, which shows that $\sim$ is transitive, thus completing the proof of Proposition 1.

Let us fix an arbitrary $a^* \in B$ and h such that $\frac{1}{2} < h < 1$. We shall be dealing with finite or infinite sequences $\ldots, b_{-1}, b_0, b_1, b_2, \ldots$ of elements of B satisfying the conditions

$$b_0 \sim a^* \quad \text{and} \quad p(b_j, b_{j+1}) \geqslant h \quad \text{for all } j.$$

The class of all such sequences will be denoted by $L(h, a^*)$.

In the class $L(h, a^*)$ we shall distinguish the sequence $\ldots, b'_{-1}, b'_0, b'_1, b'_2, \ldots$ defined by the following recursive scheme:

1° select as b'_0 any element equivalent to a^*;

2^+. for $k \geqslant 0$, if b'_k is already defined, consider the set $A_h^+(b'_k) = \{x \in B: p(b', x) \geqslant h\}$. If $A_h^+(b'_k)$ is empty, then b'_k is the last term of the sequence with index $k \geqslant 0$. Otherwise, put $b'_{k+1} \sim u(h, b'_k)$ with $u(h, b'_k)$ given by axiom (iv), part I;

2^-. for $k \leqslant 0$, if b'_k is already defined, consider the set $A_h^-(b'_k) = \{x \in B: p(x, b'_k) \geqslant h\}$. If the set $A_h^-(b'_k)$ is empty, then b'_k is the last term of the sequence with index $k \leqslant 0$. Otherwise, put $b'_{k-1} \sim v(h, b'_k)$ with $v(h, b'_k)$ given by axiom (iv), part II.

We prove first

PROPOSITION 2. *Sequence $\{b'_k\}$ is defined uniquely up to the equivalence relation $\sim$.*

Proof. By construction, the assertion is true for $k = 0$. Suppose now that b'_k $(k \geqslant 0)$ is defined uniquely. The element $u \in A_h^+(b'_k)$ given by part I of axiom (iv) is unique, for if u' is another element such that $p(u', x) \geqslant 1/2$ for all $x \in A_h^+(b'_k)$, then $p(u, u') \geqslant 1/2$ and $p(u', u) \geqslant 1/2$, which by (i) implies that $u \sim u'$. It remains to show that if $b'_k \sim b''_k$, then $A_h^+(b'_k) = A_h^+(b''_k)$. Suppose that $x \in A_h^+(b'_k)$, i.e. $p(b'_k, x) \geqslant h$. By (ii) we get $p(b''_k, x) \geqslant \max[p(b''_k, b'_k), p(b'_k, x)] \geqslant h$, hence $x \in A_h^+(b''_k)$, which shows that $A_h^+(b'_k) \subset A_h^+(b''_k)$. By symmetry we obtain also the reverse inclusion. The proof for negative k is analogous.

We shall now prove that the sequence $\{b'_k\}$ is, in a sense, the "longest" among all sequences in $L(h, a^*)$. More precisely, we prove

THEOREM 4. *If $\{b_j\}$ is an arbitrary sequence in $L(h, a^*)$, then:*

1. *for any $k \geqslant 0$ for which b_k is defined, b'_k is also defined and satisfies the relation $p(b'_k, b_k) \geqslant 1/2$;*

2. *for any $k \leqslant 0$ for which b_k is defined, b'_k is also defined and satisfies the relation $p(b_k, b'_k) \geqslant 1/2$.*

Proof. By construction, we have $b_0 \sim b'_0 \sim a^*$, hence the theorem holds for $k = 0$. Suppose that the assertion is true for some $k \geqslant 0$ and that b_k is not the last term of the sequence with positive index. We have then $p(b_k, b_{k+1}) \geqslant h$ and, by inductive assumption, $p(b'_k, b_k) \geqslant 1/2$. Using (ii) we obtain $p(b'_k, b_{k+1}) \geqslant h$, hence $b_{k+1} \in A_h^+(b'_k)$. Thus, $A_h^+(b'_k)$ is not empty and b'_{k+1} is defined. By definition, b'_{k+1} satisfies the relation $p(b'_{k+1}, x) \geqslant 1/2$ for all $x \in A_h^+(b'_k)$, hence also $p(b'_{k+1}, b_{k+1}) \geqslant 1/2$, which completes the proof for $k \geqslant 0$. The proof for $k \leqslant 0$ is analogous.

Until now we made no use of axiom (iii) which asserts "perfect distinguishability". We shall now formulate the main theorem which will serve as a foundation for our classification scheme. In the sequel, q^* will denote the number defined by (3), i.e. the lower bound of q's such that the condition of axiom (iii) holds.

THEOREM 5. *Let $\{b_j\}$ be any sequence in $L(h, a^*)$ with $q^* < h < 1$. For every $x \in B$, if $0 < p(x, b_m) < 1$ for some m, then $p(x, b_{m+k}) = 1$ and $p(b_{m-k}, x) = 1$ for all $k \geqslant 4$.*

Proof. Assume first that $\frac{1}{2} \leqslant p(x, b_m) < 1$. Then, by (ii), we have

$$p(x, b_{m+1}) \geqslant \max[p(x, b_m), p(b_m, b_{m+1})] \geqslant h > q^*.$$

Since $p(b_{m+1}, b_{m+2}) \geqslant h > q^*$, using (iii) we obtain $p(x, b_{m+2}) = 1$. Repeated application of (ii) shows now that $p(x, b_{m+k}) = 1$ for $k \geqslant 2$, which implies the first assertion of the theorem.

Next, we must have $p(b_{m-2}, x) > 1/2$, since in the opposite case we would have $p(x, b_m) = 1$ by the assertion already proved, contrary to the assumption. Thus, by (ii), we get

$$p(b_{m-3}, x) \geqslant \max[p(b_{m-3}, b_{m-2}), p(b_{m-2}, x)] \geqslant h > q^*,$$

and since $p(b_{m-4}, b_{m-3}) > q^*$, by (iii) we obtain $p(b_{m-4}, x) = 1$. Repeated application of (ii) shows that $p(b_{m-k}, x) = 1$ for all $k \geqslant 4$, which completes the proof for the case $\frac{1}{2} \leqslant p(b_m, x) < 1$. The proof for the case $0 < p(b_m, x) \leqslant 1/2$ is analogous.

Let now $\mathcal{N}$ be the class of all increasing sequences $\ldots < n_{-1} < n_0 < n_1 < n_2 < \ldots$ of integers such that $n_0 = 0$ and

$$\min_j (n_{j+1} - n_j) \geqslant 4.$$

Let $\{n_k\} \in \mathcal{N}$ and $\{b_j\} \in L(h, a^*)$ with $q^* < h < 1$. Consider the subsequence $\ldots, b_{n_{-2}}, b_{n_{-1}}, b_{n_0}, b_{n_1}, b_{n_2}, \ldots$ of sequence $\{b_j\}$ and define the function $t(x) = t(x; \{b_j\}, \{n_k\}), x \in B$, as follows:

if $b_{n_{-k}}$ is the last term of the sequence with negative index and $p(b_{n_{-k}}, x) < 1$, put $t(x) = -k-1$. Otherwise, put $t(x) = \max\{k: p(b_{n_k}, x) = 1\}$.

We prove

PROPOSITION 3. *Function $t(x)$ is well defined and finite for every $x \in B$.*

Proof. Note first that since $p(b_j, b_{j+1}) \geqslant h > q^*$ and $n_{k+1} - n_k \geqslant 4$, repeated application of (ii) and (iii) shows that $p(b_{n_j}, b_{n_k}) = 1$ for all j, k with $j < k$. Thus, if $p(b_{n_j}, x) = 1$ for some j, we have $p(b_{n_k}, x) = 1$ for all $k < j$. Similarly, if $p(x, b_{n_j}) = 1$ for some j, then $p(x, b_{n_k}) = 1$ for all $k > j$. In other words, the sequence of probabilities $\{p(b_{n_j}, x), j = \ldots \ldots, -1, 0, 1, \ldots\}$ is non-increasing. Theorem 5 implies that at most one term of this sequence may be different from 0 or 1.

Consider first sequences $\{b_{n_j}\}$ which contain the last term with negative index. If $\{b_{n_j}\}$ contains also the last term with positive index, the above reasoning implies that $t(x)$ is defined unambigously and is finite. If $\{b_{n_j}\}$ has the last negative term, say $b_{n_{-k}}$, and $p(b_{n_{-k}}, x) = 1$, to prove the finiteness of $t(x)$ is suffices to show that the sequence $\{p(b_{n_j}, x), j = -k, -k+1, \ldots\}$ cannot consist of terms all equal to one. This property, however, is assured by axiom (v), part I; in fact, the sequence $b_{n_{-k}}, b_{n_{-k}+1}, \ldots$ satisfies the assumptions of part I of (v), as $p(b_j, b_{j+1}) \geqslant h > 1/2$; thus, there exists an $m = m(x)$ such that $p(x, b_m) \geqslant 1/2$ or $p(b_m, x) \leqslant 1/2 < 1$. Clearly, $t(x) \leqslant k^*$, where $k^* = 1+\min\{k\colon n_k \geqslant m\}$, which completes the proof for the case of sequences finite from the left. Now, if the sequence $\{b_{n_j}\}$ contains no last term with negative index, by similar reasoning based on part II of axiom (v) we exclude the possibility that the sequence $\{p(b_{n_j}, x), j = \ldots, -1, 0, 1, \ldots\}$ consists of terms equal to zero only. Thus, the proof of Proposition 3 is complete.

Let now $t'(x)$ be defined as above for sequence $\{b_j'\} \epsilon L(h, a^*)$ and $\{n_k'\} = \{\ldots, -4, 0, 4, \ldots\}$. We prove

THEOREM 6. *For any $\{b_j\} \epsilon L(h, a^*)$ with $q^* < h < 1$ and $\{n_k\} \epsilon \mathcal{N}$ the function $t(x) = t(x; \{b_j\}, \{n_k\})$ satisfies the inequality $|t(x)| \leqslant |t'(x)|$ for all $x \epsilon B$.*

Proof. Suppose that $t(x) = k \geqslant 0$; thus, $p(b_{n_k}, x) = 1$. It follows from Theorem 4 that b'_{n_k} is defined and that $p(b'_{n_k}, b_{n_k}) \geqslant 1/2$. Next, $n_k = (n_1 - n_0) + (n_2 - n_1) + \ldots + (n_k - n_{k-1}) \geqslant 4k = n_k'$, hence $p(b'_{n_k'}, b'_{n_k}) \geqslant 1/2$. Repeated application of (ii) yields $p(b'_{n_k'}, x) = 1$, hence $t'(x) \geqslant k$. The proof for x such that $t(x) < 0$ is analogous.

The above constructions were based on probabilities $p(a, b)$ describing statistical properties of choices $T_s^{(i)}(a, b)$, and not on the choices themselves. We shall now define a family of integer-valued random variables which will lead directly to the construction of a classification scheme.

Let us take sequences $\{b_j\} \epsilon L(h, a^*)$ with $q^* < h < 1$ and $\{n_k\} \epsilon \mathcal{N}$. Given the family of choices

$$\{T_s^{(i)}(a, b), (a, b) \epsilon B \times B, s \epsilon S, i = 1, 2, \ldots\}$$

satisfying the properties stated at the beginning of this section, define the family of random variables

$$\{K_s^{(i)}(x) = K_s^{(i)}(x; \{b_j\}, \{n_k\}), x \epsilon B, s \epsilon S, i = 1, 2, \ldots\}$$

by putting

(a) $K_s^{(i)}(x) = k \geqslant 0$ if $T_s^{(i)}(b_{n_k}, x) = b_{n_k}$ and either b_{n_k} is the last element of the sequence $\{b_{n_j}\}$ with positive index or $T_s^{(i)}(b_{n_{k+1}}, x) = x$;

(b) $K_s^{(i)}(x) = k < 0$ if $T_s^{(i)}(b_{n_{k+1}}, x) = x$ and either $b_{n_{k+1}}$ is the last term of the sequence $\{b_{n_j}\}$ with negative index or $T_s^{(i)}(b_{n_k}, x) = b_{n_k}$.

We have

PROPOSITION 4. *Let* $\{b_j\} \in L(h, a^*)$ *with* $q^* < h < 1$ *and* $\{n_k\} \in \mathcal{N}$ *and let* $t(x) = t(x; \{b_j\}, \{n_k\})$. *Then the random variables* $K_s^{(i)}(x)$ *are defined unambigously with probability* 1, *and satisfy, with probability* 1, *the inequality*

$$t(x) \leqslant K_s^{(i)}(x) \leqslant t(x)+1$$

for all $x \in B$, $s \in S$ *and* $i = 1, 2, \ldots$

Proof. $K_s^{(i)}(x)$ is defined unambigously since, as noted in the proof of Proposition 3, the sequence of probabilities $\{p(b_{n_j}, x), j = \ldots, -1, 0, 1, \ldots\}$ is non-increasing and at most one term is strictly between zero and one. Now, suppose that $t(x) = k$, i.e. (for infinite sequences) $p(b_{n_k}, x) = 1$, $p(b_{n_{k+1}}, x) < 1$. Then $p(b_{n_{k+2}}, x) = 0$ and with probability 1 we must have $T_s^{(i)}(x, b_{n_k}) = b_{n_k}$ and $T_s^{(i)}(x, b_{n_{k+2}}) = x$. Thus $K_s^{(i)}(x)$ equals k or $k+1$ depending on whether $T_s^{(i)}(x, b_{n_{k+1}})$ equals x or $b_{n_{k+1}}$, respectively. The proof for bounded sequences requires obvious modifications.

From the assumptions about the random variables $T_s^{(i)}(a, b)$ it follows that $(i, s, x) \neq (i', s', x')$ implies independence of $K_s^{(i)}(x)$ and $K_{s'}^{(i')}(x')$. Moreover, for every $s \in S$ and $x \in B$ the random variables $K_s^{(i)}(x)$, $i = 1, 2, \ldots$, have the same distribution.

We can now define the classification scheme of the set B with respect to categories $\mathscr{C} = \{\ldots, C_{-1}, C_0, C_1, \ldots\}$ by putting simply

$$\{\xi_s^{(i)}(a) = C_{K_s^{(i)}(a)}, a \in B, s \in S, i = 1, 2, \ldots\}. \tag{4}$$

We summarize all our results as follows:

Suppose that the family of choices $T_s^{(i)}(a, b)$ *satisfies conditions* 1° *and* 2° *stated on p.* 14, *and that the induced probabilities* $p(a, b)$ *satisfy axioms* (i)-(v). *Fix* $a^* \in B$ *and* h *such that* $q^* < h < 1$, *choose sequences* $\{b_j\} \in L(h, a^*)$ *and* $\{n_k\} \in \mathcal{N}$ *and define random variables* $K_s^{(i)}(x)$ *in the manner described above. Then the random variables defined by* (4) *constitute a classification scheme, i.e. they satisfy the requirements* 1° *and* 2° *stated in Section* 2 *on p.* 10. *Let us call every classification scheme obtained in this manner an* (h, a^*)-*-scheme.*

For every (h, a^*)*-scheme, the vector of probabilities*

$$p_{s,a}(j) = P\{\xi_s^{(i)}(a) = C_j\}$$

does not depend on s, *and from Proposition* 4 *it follows that at most two of its components do not vanish. Consequently, the function* $u_{s_1,s_2}(x)$ *defined by* (2) *is at least equal to* 1/2 *for every* $x \in B$ *and does not depend on* s_1, s_2. *Finally, the* (h, a^*)*-scheme based on sequences* $\{b_k'\}$ *and* $\{n_k'\} = \{4k\}$ *is optimal* *in the class of all (h, a^*)-schemes, in the sense of "maximal number of categories", as explained in Theorem* 4.

Some natural questions arise in connection with the set of axioms (i)-(v). Define $a \prec b$ if and only if $p(a, b) > 1/2$; then relation $\prec$ is non--reflexive, antisymmetric and transitive in B. Moreover, for any $a, b \in B$ we have $a \prec b$, $a \sim b$ or $b \prec a$, hence we have a linear ordering of B. The question arises whether there exists a metric in B which can be defined in terms of $p(a, b)$ alone. The set of axioms (i)-(v) is, however, too weak and the answer is negative. Indeed, for such a construction it would be necessary to impose conditions on probabilities for "end-points" of intervals consisting of two adjoining intervals stronger than those stated in (ii) and (iii). Moreover, an obvious prerequisite for such a metric would be the requirement that any two points situated at the same distance from a given point and lying on the same side of it should be equivalent; however, from axioms (i)-(v) it does not follow that $p(a, b) = p(a, c) > 1/2$ implies $p(b, c) = 1/2$.

4. Applications. We shall now sketch briefly the results of application of the above construction of classification and methods of evaluating it to the empirical problem of classification of banknotes outlined in the introduction.

The experiments were of a preliminary nature and their aim was to gather data which would allow to assess the practical value of the suggested theory rather than provide the final version of the classification scheme.

A large sum in 20 zł banknotes was borrowed from the bank; these banknotes covered the whole "range" of various degrees of usage. After preliminary inspection, all banknotes which were "not typical", e.g. had distinct stains or were torn, were eliminated from the experimental material. Next, one banknote was selected as the initial banknote b_0 by the experimenters. The initial banknote was neither too new nor too much used; roughly speaking, it came from the "middle". Then a package of banknotes was selected by the experimenters: this package consisted of banknotes for which there could be difference of opinions whether they are "more" or "less" used up than the initial banknote. This package was then sorted in two groups: the ones judged as "better" than the initial banknote, and the ones judged as "worse" than it (ties were not allowed). The total of 24 partitions were performed by 12 students from the Institute of Mathematics of the Polish Academy of Sciences, two partitions by each of them. The results of each partition was recorded, and the package was shuffled by the experimenters before next presentation for partition. The individuals performing the partitions were not informed about the results of previous partitions.

The object was to find in this package two banknotes which evoked respectively 25% and 75% of judgments "better" than the initial banknote. Thus, it was assumed on a priori grounds that $q = 0.75$ is a number sufficient to ensure the assertion of axiom (iii) to hold.

After selecting two banknotes (b_{-1} and b_1), the procedure was repeated with packages selected by experimenters for banknotes b_{-1} and b_1; here the object was to find one banknote (b_{-2}) which evoked 25% of judgments "better than b_{-1}" and one banknote (b_2) which evoked 75% judgments "better than b_1". The procedure was repeated, thus expanding the sequence of banknotes in two directions. Every two second banknotes of the sequence were presented to the subjects: they were asked to point that one which was "better" than the other. Out of 10 persons asked, usually 10 or 9 pointed that banknote which should have been indicated if axiom (iii) were true. The fact that the decisions were not always unanimous may have been due to two factors: 1° the probability $q = 0.75$ may be somewhat too small for assertion of axiom (iii) to hold, and 2° selections of banknotes which evoked 18 out of 24 (and later, 15 out of 20) decisions may occasionally lead to pairs of banknotes a, b for which $p(a, b)$ is considerably smaller than 0.75.

In any case, the repetition of the above procedure supplied us with a sequence of 17 banknotes; presumably, the sequence could have been extended still further in both directions, as the best banknote in the sequence, though quite new and good looking, was still distincly different from a brand new one, and the "worst" banknote of the sequence was distinctly better than some banknotes one encounters in everyday life.

The 17 banknotes were then used for the purpose of classification: by taking every fourth of them, five banknotes were selected, these banknotes constituting the "boundaries" between six successive categories.

A package of banknotes was then chosen for experiments with classifications: after preliminary elimination of "non-typical" banknotes, there remained 100 banknotes to be classified into six categories of degree of waste.

The total of 13 classifications of this set of 100 banknotes was observed: five persons performed two classifications each and three persons one classification each.

According to theory, the classifications of every banknote should fall either within one category or into two neighbouring categories. This turned out to be true for 55 banknotes; for 43 banknotes the classifications fell into three successive categories, and for 2 banknotes — into four successive categories. However, in 37 cases out of the above 43 the banknotes were put into the "outlying" category only once, and all remaining 12 classifications fell into the other two categories.

The investigation of stability and inter-individual consistency for five persons who performed two classifications are presented on the following table:

	s_1	s_2	s_3	s_4	s_5
s_1	0.86	0.75 *0.04*	0.57 *0.06*	0.69 *0.05*	0.54 *0.05*
s_2		0.80	0.53 *0.08*	0.64 *−0.01*	0.52 *0.01*
s_3			0.67	0.50 *0.05*	0.47 *0.04*
s_4				0.59	0.55 *0.03*
s_5					0.47

In this table the upper numbers give values of the estimates of

$$m_{s_i,s_j}(Q) = \frac{1}{100}\sum_{k=1}^{100} u_{s_i,s_j}(b_k),$$

where $Q = \{b_1, \ldots, b_{100}\}$ are banknotes of the sample, and $u_{s_i,s_j}(b_k)$ is the probability that classification of banknote b_k by individuals s_i and s_j will coincide.

The numbers in italics (lower parts of the entries) give values of estimates of

$$\sigma^2_{s_i,s_j}(Q) = \frac{1}{100}\sum_{k=1}^{100} [u_{s_i,s_j}(b_k) - m_{s_i,s_j}(Q)]^2.$$

These values could not have been computed for entries on the diagonal, as nobody performed four classifications.

The variance of estimates whose values are given on the diagonal are bounded by 1/400. The variance of estimates whose values are given in upper parts of the off-diagonal entries are bounded by 1/800, and the variance of estimates whose values are given in italics are bounded by 3/400.

At first sight, these results may appear not too satisfactory. However, they seem to indicate the possibility of definite improvement of the situation from the point of view of the bank. The preliminary survey of the present situation in sorting banknotes in the bank revealed that the present state of affairs is rather far from satisfactory, as may be judged from the following data. A sample of 1000 banknotes was put through the normal present procedure of counting with elimination of destructs by subjects in two branches of the bank; each subject performed the process of counting and elimination twice on two different days, without being aware that she was being tested. In one case, a subject eliminated on different days from the sample of 1000 banknotes 280 and 470 banknotes as destructs, while the other eliminated (from the same sample) only 118 and 82 banknotes. These are extreme cases, to be sure, but in the remaining cases the stability and inter-individual consistency was also low. The tests in the Institute of Mathematics, reported above, were made under experimental conditions, and the subtests were trying to do their best. The fact remains, however, that they reached reasonably high standards of stability and interindividual consistency in performing a considerably more difficult task of classifying into six, and not only two categories. It appears therefore that the method employed, namely the use of suitably selected "standard banknotes" for comparison, may provide the bank with a method of improving the present situation in the process of elimination of destructs. The results of classifications show also that one might hope to develop reasonably objective methods of assessing the overall quality of the population of banknotes. This, in turn, may lead to a search for choice of such a "level of destructs" which would give some optimum with respect to suitable criteria involving the number of banknotes rejected and the resulting overall quality of the population of banknotes.

References

[1] R. Bartoszyński, *A note on subjective classifications*, Rev. Internat. Statist. Inst., in press.

[2] T. Dalenius and O. Frank, *Control of classification*, ibidem 36 (1968), No. 2.

INSTITUTE OF MATHEMATICS OF THE POLISH ACADEMY OF SCIENCES

Received on 9. 6. 1969

R. BARTOSZYŃSKI (Warszawa)

O KONSTRUKCJI I OCENIE SUBIEKTYWNYCH KLASYFIKACJI

STRESZCZENIE

W pracy podana jest definicja klasyfikacji subiektywnej i metody oceny jakości takiej klasyfikacji. Z grubsza biorąc, klasyfikacją *subiektywną* nazywa się przyporządkowanie każdemu obiektowi danego zbioru i każdej z klasyfikujących osób pewnej zmiennej losowej o wartości ze zbioru, którego elementy nazywają się *kategoriami*

klasyfikacji. Jakość klasyfikacji (w odniesieniu do danego obiektu i pary osób) wyrażona została przez prawdopodobieństwo, że obiekt ten zostanie przez te osoby zaklasyfikowany jednakowo. Podano estymatory wartości średniej i „wariancji" takich prawdopodobieństw dla zbiorów obiektów.

W drugiej części pracy podano metody konstrukcji klasyfikacji subiektywnych zbioru obiektów ze względu na cechę, której wartości w zbiorze klasyfikowanych obiektów nie mogą być bezpośrednio mierzone. Można je jedynie oceniać subiektywnie; dokładniej, zakłada się, że z każdą osobą i parą obiektów, powiedzmy $\langle a, b\rangle$, związana jest zmienna losowa o wartościach w zbiorze $\{a, b\}$ reprezentująca wybór tego obiektu z pary, który ma (w odczuciu wybierającego) „mniejszą" wartość cechy. Podano układ założeń o takich zmiennych losowych, które pozwalają skonstruować klasyfikację opartą o ciąg obiektów-wzorców, stanowiących granice między kategoriami klasyfikacji.

Na zakończenie zilustrowano zaproponowane metody wynikami empirycznych badań nad klasyfikacją banknotów ze względu na ich zużycie.