DECISION RULES, BASED ON THE DISTANCE, FOR THE PROBLEMS OF INDEPENDENCE, INVARIANCE AND TWO SAMPLES*

By KAMEO MATUSITA AND HIROTUGU AKAIKE

(Received Oct. 25, 1955)

1. Introduction

The purpose of this paper is to deal with the problems of independence, invariance (symmetry) and two samples along the line of the previous paper [1]. The characteristic feature of our treatment is, as in [1], that the risk function of our decision rule is uniformly bounded by a preassigned value. Here we consider only the risk caused by the error in inference, because the cost of experiment can easily be taken into account with it (see [2]). Throughout this paper, we are concerned with finite discrete distributions, and employ the definition of distance given in [2]. As will be seen below, this very definition of distance makes our decision rule applicable to the problems of independence, invariance, and two samples, and reduces the problems to those of finding the maximum eigenvalue of a symmetric, positive semi-definite matrix or the Euclidean length of a vector. Our idea is very straightforward and the methods are simple (see [4], [5], [6], [7]).

First we shall state the two fundamental theorems and the general decision rule. One of the theorems is new, and ~~will~~ in some class of applications ~~be~~ more precise than the other. Second we shall show how our decision rule applies to the problems of independence, invariance (symmetry), and two samples. The last problem has been treated in [1], but the present treatment can be applied when we also want to know at the same time the independence of two concerned random variables. Finally we shall give some examples. In the appendix, we shall illustrate the relation between our affinity and the correlation coefficient concerning two-variate Gaussian distributions, which will serve for understanding the move of affinity.

---

* The results of this paper were partly announced in October, 1954 at the annual meeting of the Japan Statistical Society.

2. Fundamental theorems and decision rule.

Let $E_1, E_2, \cdots, E_k$ be the events which can be observed in our experiment. In this section we consider only distributions on these events, and probabilities of these events are represented like $p_1, p_2, \cdots, p_k$, for instance. The empirical distribution based on n observations, i.e., $\left(\frac{n_1}{n}, \frac{n_2}{n}, \cdots, \frac{n_k}{n}\right)$, where n_i denotes an observed frequency of E_i ($i=1, 2, \cdots, k$), will generally be denoted by S_n. The distance between two distributions $F=\{p_i\}$ and $G=\{q_i\}$ is given by

$$||F-G||=\sqrt{\sum_{i=1}^{k}(\sqrt{p_i}-\sqrt{q_i})^2}$$

THEOREM I. *Let F_0 be the distribution of the random variable on which observation is made, and S_n its empirical distribution. Then, for any set of distributions ω and for any positive number η we have*

$$\Pr\{\inf_{F\in\omega}||F-S_n||\leqq\eta\}\geqq 1-\frac{k-1}{n\eta^2} \quad \textit{when } F_0\in\omega$$

and

$$\Pr\{\inf_{F\in\omega}||F-S_n||>\eta\}\geqq 1-\frac{k-1}{n(\varepsilon-\eta)^2} \quad \textit{when } \inf_{F\in\omega}||F-F_0||\geqq\varepsilon>\eta\,.$$

This is a slight modification of Theorem I in [1].

PROOF. When $F_0\in\omega$, we have clearly

$$\inf_{F\in\omega}||F-S_n||\leqq||F_0-S_n||$$

therefore, according to theorem I in [1],

$$\Pr\{\inf_{F\in\omega}||F-S_n||\leqq\eta\}\geqq\Pr\{||F_0-S_n||\leqq\eta\}\geqq 1-\frac{k-1}{n\eta^2}\,.$$

When $\inf_{F\in\omega}||F-F_0||\geqq\varepsilon>\eta$, we have

$$\inf_{F\in\omega}||F-S_n||\geqq\inf_{F\in\omega}||F-F_0||-||F_0-S_n||\geqq\varepsilon-||F_0-S_n||$$

therefore,

$$\Pr\{\inf_{F\in\omega}||F-S_n||>\eta\}\geqq\Pr\{||F_0-S_n||\leqq\varepsilon-\eta\}\geqq 1-\frac{k-1}{n(\varepsilon-\eta)^2}\,.$$

THEOREM II.[1), 2)] *For any set of distributions ω and for any positive*

1) Thanks are due to Professor J. Wolfowitz and Mr. M. Motoo for their suggestions to this theorem.

2) We have obtained another precise theorem, which is to appear in the forthcoming issue of these Annuals.

number η we have

$$\Pr\{\inf_{F\in\omega} ||F-S_n||\leqq\eta\}\geqq 1-2ke^{-n\eta^4/2k^2} \quad when\ F_0\in\omega$$

and

$$\Pr\{\inf_{F\in\omega} ||F-S_n||\geqq\eta\}\geqq 1-2ke^{-n(\varepsilon-\eta)^4/2k^2} \quad when\ ||F-F_0||\geqq\varepsilon>\eta\ .$$

PROOF. From

$$\left(\sqrt{\frac{n_i}{n}}-\sqrt{p_i}\right)^2>\eta^2$$

it follows that

$$\text{(i)} \qquad \sqrt{\frac{n_i}{n}}-\sqrt{p_i}>\eta$$

or

$$\text{(ii)} \qquad \sqrt{p_i}-\sqrt{\frac{n_i}{n}}>\eta$$

When (i) holds, we have

$$\sqrt{\frac{n_i}{n}}>\eta+\sqrt{p_i}$$

hence

$$\frac{n_i}{n}>\eta^2+p_i$$

that is

$$\frac{n_i}{n}-p_i>\eta^2 .$$

When (ii) holds, we have

$$\frac{p_i-\dfrac{n_i}{n}}{\sqrt{p_i}+\sqrt{\dfrac{n_i}{n}}}>\eta$$

and

$$\sqrt{p_i}>\eta$$

from which it follows that

$$p_i-\frac{n_i}{n}>\sqrt{p_i\eta^2}>\eta^2 .$$

Thus we get always

$$\left|\frac{n_i}{n}-p_i\right|>\eta^2$$

from

$$\left(\sqrt{\frac{n_i}{n}}-\sqrt{p_i}\right)^2>\eta^2.$$

Therefore, when $F_0=\{p_1^0, p_2^0, \cdots, p_k^0\}$ we obtain

$$Pr\left(\left(\sqrt{\frac{n_i}{n}}-\sqrt{p_i^0}\right)^2>\eta^2\right)\leqq Pr\left(\left|\frac{n_i}{n}-p_i^0\right|>\eta^2\right)<2e^{-n/2\eta^4}$$

according to the inequality in Uspensky's book [9], p. 102. Using this inequality we have

$$Pr(\|F_0-S_n\|>\eta)=Pr\left(\sum_{i=1}^k\left(\sqrt{\frac{n_i}{n}}-\sqrt{p_i}\right)^2>\eta^2\right)\leqq\sum_{i=1}^k Pr\left(\left(\sqrt{\frac{n_i}{n}}-\sqrt{p_i}\right)^2>\frac{\eta^2}{k}\right)$$
$$\leqq 2ke^{-n\eta^4/2k^2}$$

from which the inequalities in the theorem follow.

By these theorems we obtain the following

DECISION RULE:

When $\inf_{F\in\omega}\|F-S_n\|\leqq\eta$, *decide* $F_0\in\omega$

and when $\inf_{F\in\omega}\|F-S_n\|>\eta$, *decide* $\inf_{F\in\omega}\|F-F_0\|\geqq\varepsilon\,(>\eta)$.

Then, the risk function of this decision rule is uniformly bounded by $\frac{k-1}{n\eta^2}$ *or* $\frac{k-1}{n(\varepsilon-\eta)^2}$, *or by* $2ke^{-n\eta^4/2k^2}$ *or* $2ke^{-n(\varepsilon-\eta)^4/2k^2}$, *provided that the weight function takes the value* 0 *when the decision is correct, and the value smaller than* 1 *when the decision is wrong.*

As is easily seen, Theorem I gives more advantage than Theorem II for some values of k when the risk is required to be ~~less~~ greater than 0.02 and Theorem II is ~~almost always~~ preferable for large n.

If each distribution F under consideration admits χ^2, based on n observations of the random variable with F, to have asymptotically the chi-square distribution with $k-1$ degrees of freedom, the inequalities in the above theorems can be made more precise and the risk is made smaller than $\Pr\{\chi^2_{(k-1)}\geqq 4n\eta^2\}$ or $\Pr\{\chi^2_{(k-1)}\geqq 4n(\varepsilon-\eta)^2\}$, where $\chi^2_{(k-1)}$ is a random variable having the χ^2 distribution with $k-1$ degrees of freedom (see Theorem II in [1]). A set of distributions for which the above condition holds was called a class (C_n) in [1].

Now, we have

$$\| F-G \|^2 = 2\{1-\rho(F, G)\}$$

where $\rho(F, G)$ is the affinity between F and G, i.e., $\sum_{i=1}^{k} \sqrt{p_i q_i}$ (see [2]). Hence, the above decision rule is rewritten in terms of affinity as follows:

When $\sup_{F \in \omega} \rho(F, S_n) \geqq 1 - \frac{\eta^2}{2}$, *decide* $F_0 \in \omega$ *and when* $\sup_{F \in \omega} \rho(F, S_n) < 1 - \frac{\eta^2}{2}$, *decide* $\inf_{F \in \omega} \| F - F_0 \| \geqq \varepsilon \ (> \eta)$.

These forms of the decision rule are applicable whenever $\inf_{F \in \omega} \| F - S_n \|$ or $\sup_{F \in \omega} \rho(F, S_n)$ is calculated from S_n and the set ω, itself. For example, when ω consists of a single distribution, this can be done, and the problem then is that of goodness of fit, which we have already treated in [1]. In the following sections it will be shown that our decision rule is applicable to the problems of independence and invariance. The two-sample problem will also be referred to, though it was already treated in [1].

3. Problem of independence

Let X, Y be two random variables taking $E_1, E_2, \cdots, E_k$, and $E'_1, E'_2, \cdots, E'_{k'}$, as the values, respectively. Denote by F_0 the joint distribution of (X, Y) and by ω the set of all distributions on the joint events (E_i, E'_j) such that the probabilities of (E_i, E'_j) determined by them can always be written in the form $p_i q_j$ with $\sum_{i=1}^{k} p_i = 1$, $\sum_{i=1}^{k'} q_i = 1$, $p_i \geqq 0$, $q_i \geqq 0$. Then, the problem we treat here is to decide whether $F_0 \in \omega$ or $\inf_{F \in \omega} \| F - F_0 \| > \varepsilon$ where ε is a preassigned positive number.

To apply our decision rule to this problem, we have only to show that $\sup_{F \in \omega} \rho(F, S_n)$ can actually be calculated. Let $F = \{p_i q_i\}$ and $S_n = \{p_{ij}\}$, and put $x_i = \sqrt{p_i}$, $y_j = \sqrt{q_j}$, $a_{ij} = \sqrt{p_{ij}}$. Then we have

$$\rho(F, S_n) = \sum_{i=1}^{k} \sum_{j=1}^{k'} \sqrt{p_i} \sqrt{q_j} \sqrt{p_{ij}} = \sum_{i, j} a_{ij} x_i y_j$$

When we denote the matrix (a_{ij}) and the vectors $\begin{pmatrix} x_1 \\ \vdots \\ x_{k'} \end{pmatrix}, \begin{pmatrix} y_1 \\ \vdots \\ y_{k'} \end{pmatrix}$ by A and $\mathfrak{x}, \mathfrak{y}$, respectively, $\rho(F, S_n)$ is represented as the inner product of two vectors:

$$\rho(F, S_n)=(A'\mathfrak{x}, \mathfrak{y}) \quad \text{or} \quad (\mathfrak{x}, A\mathfrak{y})$$

where A' is the transposed matrix of A. Since

$$|\mathfrak{x}|=\sqrt{\sum_{i=1}^{k} x_i^2}=1, \quad |\mathfrak{y}|=\sqrt{\sum_{i=1}^{k'} y_i^2}=1,$$

it holds that

$$(A'\mathfrak{x}, \mathfrak{y})\equiv(\mathfrak{x}, A\mathfrak{y})\leq|A'\mathfrak{x}| \quad \text{or} \quad |A\mathfrak{y}|$$

and the expression $(A'\mathfrak{x}, \mathfrak{y})$ or $(\mathfrak{x}, A\mathfrak{y})$ attains its maximum value when and only when the direction of $A'\mathfrak{x}$ (or $\mathfrak{y}$) coincides with that of $\mathfrak{x}$ (or $A\mathfrak{y}$). Therefore, we have

$$\max_{F\in\omega} \rho(F, S_n)=\max_{\substack{|\mathfrak{x}|=1 \\ x_i\geqq 0}} |A'\mathfrak{x}| \qquad (=\max_{\substack{|\mathfrak{y}|=1 \\ y_i\geqq 0}} |A\mathfrak{y}|).$$

Now,

$$|A'\mathfrak{x}|^2=(A'\mathfrak{x}, A'\mathfrak{x})=(AA'\mathfrak{x}, \mathfrak{x})$$

and AA' is a symmetric, positive semi-definite matrix, so AA' can be transformed into a diagonal form by an orthogonal matrix $U\,(UU'=E)$, i.e.,

$$UAA'U'=\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_k \end{pmatrix}.$$

The eigenvalues $\lambda_1, \cdots, \lambda_k$ are all non-negative, and their sum is 1, for $\sum_{i=1}^{k} \lambda_i=\text{trace}\,(AA')=\sum_{i,j} a_{ij}a_{ij}=1$. (From this fact it can also be seen that $\max_i \lambda_i=1$ means $\text{rank}\,(AA')=\text{rank}\,(A)=1$, that is, $S\in\omega$.) Let λ_{S_n} be the maximum value among them. Then we have

$$\max_{|\mathfrak{x}|=1} |A'\mathfrak{x}|^2=\lambda_{S_n}\,.$$

As is well known, this maximum eigenvalue and the corresponding eigenvector of AA' can eassily be obtained by the iterative method and this procedure assures us that, since $a_{ij}\geqq 0$, starting with some vector with non negative components we can reach a desired eigenvector with non-negative components*. Consequently, we have

$$\max_{F\in\omega} \rho(F, S_n)=\sqrt{\lambda_{S_n}} \text{ (=maximum eigenvalue of } AA') .$$

Similarly, we have

$$\max_{F\in\omega} \rho(F, S_n)=\sqrt{\lambda'_{S_n}} \text{ (=maximum eigenvalue of } A'A)$$

from which it follows also that the maximum eigenvalue of AA' is generally equal to that of $A'A$. Therefore, we can use the one with fewer degrees of AA' and AA' to obtain $\max_{F\in\omega} \rho(F, S_n)$. An approximate value of the maximum eigenvalue of a symmetric matrix is easily obtained.* Thus, the problem of independence can be treated by our decision rule.

4. Problem of invariance

Suppose, there are given k events $E_1, E_2, \cdots, E_k$, and a random variable X takes the joint event $(E_{i_1}, \cdots, E_{i_r})$ as the value where each E_{i_j} is one of $E_1, \cdots, E_k$. Let F_0 be its distribution, and for a given partition of $(1, 2, \cdots, r)$ into subsets $M_1=(1, 2, \cdots, r_1)$, $M_2=(r_1+1, \cdots, r_1+r_2), \cdots, M_s=\left(\sum_{i=1}^{s-1} r_i+1, \cdots, r\left(=\sum_{i=1}^{s} r_i\right)\right)$, represent by π the set of all permutations of $(1, 2, \cdots, r)$ which leave each M_i unchanged. When $F=\{p_{i_1,\cdots,i_r}\}$ is a distribution on $(E_{i_1}, \cdots, E_{i_r})$, we call F invariant under π, if for any permutation T in π it always holds that

$$p_{i_1,\cdots,i_r}=p_{T(i_1,\cdots,i_r)}$$

where $T(i_1, \cdots, i_r)$ means $(i_{j_1}, \cdots, i_{j_r})$ with $(j_1, \cdots, j_r)=T(1, \cdots, r)$. Let ω be the set of all distributions which are invariant under π. Our problem is then to decide whether $F_0\in\omega$ or $\inf_{F\in\omega}\|F-F_0\|>\varepsilon\ (>0)$. Denoting the empirical distribution based on n observations by $S_n=\{q_{i_1,\cdots,i_r}\}$, we have

$$\sup_{F\in\omega} \rho(F, S_n)=\sup_{F\in\omega} \sum \sqrt{p_{i_1,\cdots,i_r}}\sqrt{q_{i_1,\cdots,i_r}}$$

where $F=\{p_{i_1,\cdots,i_r}\}$, and the summation runs over all combinations $(i_1, \cdots, i_r)$. Now, putting

$$x_{i_1,\cdots,i_r}=\sqrt{p_{i_1,\cdots,i_r}}$$

* See, for instance, [8].

$$a_{i_1,\cdots,i_r}=\frac{1}{r_1!\cdots r_s!}\sum_{T\in\pi}\sqrt{q_{T(i_1,\cdots,i_r)}}$$
$$=\frac{l_1^{(1)}!\,l_2^{(2)}!\cdots l_k^{(1)}!}{r_1!}\cdots\frac{l_1^{(s)}!\,l_2^{(s)}!\cdots l_k^{(s)}!}{r_s!}\sum_{T\in\pi}{}'\sqrt{q_{T(i_1,\cdots,i_r)}}$$

where $l_1^{(u)},\cdots,l_k^{(u)}$ are the numbers of $E_1,\cdots,E_k$ in E_{i_j} for $j\in M_u$ $(u=1,\cdots,s)$ and the summation $\sum'$ runs over the set of different $T(i_1,\cdots,i_r)$, and representing the vectors $(a_{i_1,\cdots,i_r})$, $(x_{i_1,\cdots,i_r})$, whose components correspond to the different events $(E_{i_1},\cdots,E_{i_r})$, by $\mathfrak{a}$, $\mathfrak{x}$, respectively, we obtain

$$\max_{F\in\omega}\rho(F,S_n)=\max_{\substack{|\mathfrak{x}|=1\\ x_i\geqq 0}}(\mathfrak{x},\mathfrak{a})$$

$\mathfrak{x}$ is clearly invariant under π, and $(\mathfrak{x},\mathfrak{a})$ attains its maximum value when and only when $\mathfrak{x}$ is parallel to $\mathfrak{a}$ in the positive sense. The maximum value of $(\mathfrak{x},\mathfrak{a})$, is, therefore, $|\mathfrak{a}|=\sqrt{\sum a^2_{i_1,\cdots,i_r}}$, from which it also follows that

$$\max_{F\in\omega}\rho(F,S_n)=|\mathfrak{a}|\,.$$

By means of this equation our decision rule can be applied to the problem of invariance (or symmetry).

The problem of invariance in general is similarly treated. First to state the problem, for a given partition of $(1,\cdots,k)$ into subsets $M_1'=(1,\cdots,r_1),\cdots,M_s'=\left(\sum_{i=1}^{s-1}r_i+1,\cdots,k\left(=\sum_{i=1}^{s}r_i\right)\right)$, denote by ω the set of all distributions on $E_1,\cdots,E_k$ such that $p_1=\cdots=p_{r_1}$, $p_{r_1+1}=\cdots=p_{r_1+r_2},\cdots,p_{\sum_{i=1}^{s-1}r_i+1}=\cdots=p_k$. The problem of invariance is then to decide whether the random variable which we observe has a distribution in ω or a distribution at a distance more than ε (>0) from ω.

The treatment is quite similar to the above. When $S_n=\{q_i\}$, we have

$$\sup_{F\omega}\rho(F,S_n)=\sup_{F\in\omega}\sum_{i=1}^{k}\sqrt{p_i}\sqrt{q_i}=\sup_{F\in\omega}\sum_{i=1}^{k}\frac{\sum_k^{(i)}\sqrt{q_h}}{r_j(i)}\sqrt{p_i}$$

where the summation $\sum_k^{(i)}$ runs over $M_j'(i)$ which contains i. Put $a_i=\dfrac{\sum_k^{(i)}\sqrt{q_h}}{r_j(i)}$ and $\mathfrak{a}=\begin{pmatrix}a_1\\ \vdots\\ a_n\end{pmatrix}$. Then $a_1=\cdots=a_{r_1}$, $a_{r_1+1}=\cdots=a_{r_1+r_1},\cdots,$ $a_{\sum_{i=1}^{s-1}r_i+1}=\cdots=a_k$ and we obtain $\max_{F\in\omega}\rho(F,S_n)=|\mathfrak{a}|$.

5. Two-sample problem

Though this problem was treated in [1], we shall in this section give another treatment which is similar to that in section 3. The former treatment is much simpler than the present one, but the present one can apply to the case where the two random variables concerned are not known beforehand to be independent of each other and we want to know their independence at the same time.

Let X, Y be two random variables which take $E_1, \cdots, E_k$ as the values, and suppose, our observation is made on the joint events (E_i, E_j). The distribution F_0 of (X, Y) is then represented as $\{p_{ij}\}$. When ω denotes the set of all distributions $\{p_{ij}\}$ such that p_{ij} can be written as $p_i p_j$ with $\sum_{i=1}^{k} p_i = 1$, $p_i \geqq 0$, the two-sample problem is formulated as follows:

$$\text{decide whether } F_0 \in \omega \quad \text{or} \quad \inf_{F \in \omega} \| F - F_0 \| > \varepsilon \ (> 0) .$$

In this case, we have

$$\sup_{F\in\omega} \rho(F, S_n) = \sup_{F\in\omega} \sum_{i,j} \sqrt{p_i}\sqrt{p_j}\sqrt{p_{ij}} = \sup_{\substack{|\mathfrak{x}|=1 \\ x_i \geqq 0}} (A\mathfrak{x}, \mathfrak{x})$$

where $S_n = \{p_{ij}\}$, $F = \{p_i p_j\}$, $A = \left(\frac{1}{2}(\sqrt{p_{ij}} + \sqrt{p_{ji}})\right)$, $\mathfrak{x} = \begin{pmatrix} \sqrt{p_1} \\ \vdots \\ \sqrt{p_k} \end{pmatrix}$, therefore, as in section 3,

$$\sup_{F\in\omega} \rho(F, S_n) = \max_{|\mathfrak{x}|=1} (A\mathfrak{x}, \mathfrak{x}) = \text{maximum eigenvalue of } A.$$

6. Examples

1. The following table* shows a set of observations on (X, Y). Are X and Y independent of each other?

Y \ X	1	2	3
1	154	12	15
2	25	6	4
3	18	7	12

In this case

$$S_n = \begin{pmatrix} 154 & 12 & 15 \\ 25 & 6 & 4 \\ 18 & 7 & 12 \end{pmatrix}.$$

Therefore

* This table shows the actual distribution of the sizes of letters reporting the same political news in a certain two weeks in two Japanese big newspapers. We classified the sizes of letters into three period of classes.

$$\sqrt{n}\,A=\begin{pmatrix}\sqrt{154} & \sqrt{12} & \sqrt{15}\\ \sqrt{25} & \sqrt{6} & \sqrt{4}\\ \sqrt{18} & \sqrt{7} & \sqrt{12}\end{pmatrix}$$

and

$$nAA'=\begin{pmatrix}30.999 & 27.480 & 72.762\\ 27.480 & 24.998 & 66.460\\ 72.762 & 66.460 & 197.011\end{pmatrix} .$$

The eigenvector which corresponds to the maximum eigenvalue is given by

$$\mathfrak{x}_{S_n}=\begin{pmatrix}0.890\\ 0.307\\ 0.338\end{pmatrix}$$

and the maximum eigenvalue by

$$\lambda_{S_n}=0.9787 .$$

Acoording to this value we can make decision, when the values of ε and η, mentioned in section 2, are given. But, when such values are not given, we usually proceed as follows. First we assign a bound of possible risk, and then calculate

$$d=\sqrt{\inf_{F\in\omega}\|F-S_n\|^2}=\sqrt{2(1-\sqrt{\lambda_{S_n}})} \text{ and } \eta=\sqrt{\frac{kk'-1}{n}\frac{1}{\alpha_0}}$$

where ω is the set of all distributions of (X, Y) such that X, Y are independent. If η exceeds d, we decide that X, Y are independent of each other. Exactly speaking, we can state with the risk smaller than α_0 that the distribution of (X, Y) belongs to ω if η exceeds d, or it belongs to the class of alternative distributions each of which lies at a distance more than $2\sqrt{\frac{kk'-1}{n}}$ from ω. In the above case

$$d^2=2(1-\sqrt{\lambda_{S_n}})=0.0214$$

and for $\alpha_0=0.08$

$$\eta^2=\frac{8}{253}\cdot\frac{1}{0.08}=0.395 ,$$

so X, Y can be considered to be independent. When we restrict our consideration to a class $(C_{n,\,kk'-1})$ (see section 2 or [1]), the case where the ordinary χ^2 test applies, we can replace η by $\eta^*(>0)$ which satisfies

$$Pr\{\chi^2 \text{ with } kk'-1 \ d.f. \leqq 4n\eta^{*2}\} = \alpha_0 .$$

For $\alpha_0=0.05$, $kk'-1=8$ and $n=253$ we have then

$$\eta^{*2}=0.0153 ,$$

accordingly, we decide that X, Y are independent. In this case the class of alternative distributions consists of distributions with a distance more than $2\eta^*$ from ω, and the risk is α_0. (When applying the ordinary χ^2 test to the above example, we have $\chi^2=26.955$, and the degree of freedom is 4. Therefore, the hypothesis that X, Y are independent, is rejected even at level of significance 0.1%.)

Now, we want to make a remark. Since our distance between two distributions can not exceed 2, it may happen that the class of alternative distributions is empty according to the values of α_0, η or η^*, n, k, k'. For example, let $\alpha_0=0.05$, $kk'-1=8$. Then, in general, the alternative class is empty for $n<320$, and when we restrict ourselves to a class $(C_{n,\ kk'-1})$, the alternative class is empty for $n \leqq 7$. If the alternative class is empty, the decision rule is trivial and meaningless. Therefore, for given k, k', we should determine α_0, n appropriately. That is, if we want to make α_0 small, we must make n large, and as n becomes large, the alternative class is enlarged. Conversely, if we want to keep n not so large, we must be satisfied with α_0 not small, and as α_0 becomes large, the alternative class is enlarged. In the above case, where $\alpha_0=0.05$, $kk'-1=8$, if each distribution under consideration belongs to $(C_{n,\ kk'-1})$, and if $n=300$, the alternative class consists of distributions which lie apart more than $2\eta^*=\sqrt{0.05169}=0.227$ from ω. It will be not without interest to compare here this situation with that of the case where two specified distributions lying at distance 0.227 from each other are concerned. In the latter case, to make decision with the risk smaller than 0.05, sample-size 23 suffices (see [2], [3]).

2. Problem of independence

$$n=497$$

$$S=\begin{pmatrix} 144 & 16 & 25 \\ 36 & 16 & 16 \\ 64 & 36 & 144 \end{pmatrix} .$$

In this case, we have

$$\sqrt{n}\,A=\begin{pmatrix}12 & 4 & 5\\ 6 & 4 & 4\\ 8 & 6 & 12\end{pmatrix}.$$

$$A'A=\begin{pmatrix}0.4909 & 0.2414 & 0.3622\\ 0.2414 & 0.1368 & 0.2173\\ 0.3622 & 0.2173 & 0.3722\end{pmatrix}.$$

$$\lambda_{S_n}=0.9313, \qquad d^2=2(1-\sqrt{\lambda_{S_n}})=0.0699, \qquad nd^2=34.7514$$

$$\frac{kk'-1}{nd^2}=0.2302\ .$$

Therefore, we generally decide by Theorem I with the risk smaller than 0.05 that X, Y are independent~~, but when we use theorem II we decide that X, Y are dependent~~ (~~and~~ when we restrict ourselves to a class $(C_{n,\ kk'-1})$, we ~~also~~ decide that X, Y are dependent since $Pr\{\chi^2_{d.f.8}\geqq 15.507\}=0.05$).

The following examples will be interpreted in the same way. ~~Thus, examples with symbols I*, II* are considered to be dependent or not symmetric with the risk 0.05 by Theorems I and II, respectively, and examples with symbol II$^\triangle$ are considered to be dependent or not symmetric with the risk 0.01 by Theorem II.~~ Computations needed for solving these examples are carried out in five or six stages of iterations starting with $\mathfrak{x}_0=\begin{pmatrix}1\\1\\1\end{pmatrix}$.

3. Problem of independence.

S_n	$\mathfrak{x}_{S_n}$	$\sqrt{\lambda_{S_n}}$	d^2	
$\begin{pmatrix}144 & 16 & 1\\ 9 & 25 & 1\\ 0 & 4 & 100\end{pmatrix}$,	$\begin{pmatrix}0.854\\ 0.431\\ 0.292\end{pmatrix}$,	$\sqrt{0.632}=0.795$,	0.410,	~~I*~~
$\begin{pmatrix}144 & 16 & 1\\ 1 & 9 & 25\\ 100 & 4 & 0\end{pmatrix}$,	$\begin{pmatrix}0.954\\ 0.290\\ 0.024\end{pmatrix}$,	$\sqrt{0.892}=0.944$,	0.112,	~~II*~~
$\begin{pmatrix}144 & 1 & 1\\ 0 & 0 & 25\\ 4 & 4 & 121\end{pmatrix}$,	$\begin{pmatrix}0.701\\ 0.151\\ 0.697\end{pmatrix}$,	$\sqrt{0.619}=0.787$,	0.426,	~~II*~~

Therefore, within a class $(C_{n,\ kk'-1})$, we can decide with risk 0.05

that the random variables concerned are dependent in each case since $Pr\{\chi^2_{d.f.8} \geqq 4 \times 300 \times 0.0129\} = 0.05$.

4. Problem of independence.

S_n	$\sqrt{\lambda_{S_n}}$	d^2	
$\begin{pmatrix} 1 & 16 & 144 \\ 101 & 29 & 9 \end{pmatrix}$,	$\sqrt{0.728}=0.853$,	0.294,	~~II*, II△~~
$\begin{pmatrix} 1 & 16 & 144 \\ 25 & 13 & 101 \end{pmatrix}$,	$\sqrt{0.968}=0.984$,	0.032,	~~II*~~
$\begin{pmatrix} 1 & 1 & 144 \\ 146 & 4 & 4 \end{pmatrix}$,	$\sqrt{0.628}=0.792$,	0.416,	~~I*, II*, II△~~

5. Problem of symmetry.

S_n	$\lvert a_{S_n} \rvert$	d^2	
$\begin{pmatrix} 1 & 16 & 144 \\ 1 & 9 & 25 \\ 0 & 4 & 100 \end{pmatrix}$,	0.852,	0.296,	~~II*, II△~~
$\begin{pmatrix} 1 & 16 & 1 \\ 25 & 9 & 144 \\ 0 & 4 & 100 \end{pmatrix}$,	0.911,	0.178,	~~II*, II△~~
$\begin{pmatrix} 1 & 1 & 16 \\ 25 & 0 & 144 \\ 100 & 9 & 4 \end{pmatrix}$,	0.881,	0.238,	~~II*, II△~~
$\begin{pmatrix} 1 & 11 & 144 \\ 1 & 9 & 100 \\ 0 & 4 & 25 \end{pmatrix}$,	0.799,	0.402,	~~II*, II△~~

Appendix

Tables of the affinity and correlation coefficient.

Let $Z=(X, Y)$ and $Z_0=(X_0, Y_0)$ be two random variables whose distributions are the Gaussian with density functions

$$\frac{1}{2(1-r^2)} e^{-\frac{x^2-2rxy+y^2}{2(1-r^2)}}, \quad \frac{1}{2} e^{-\frac{1}{2}(x^2+y^2)},$$

respectively. On the right, numerical values of the affinity ρ between Z and Z_0 and the correlation coefficient r between X and Y are given for illustration.

r	ρ
0.0	1.
0.1	0.9987
0.2	0.9948
0.3	0.9879
0.4	0.9771
0.5	0.9611
0.6	0.9376
0.7	0.9021
0.8	0.8452
0.9	0.7393
1.0	0.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Matusita, Kameo: Decision rules based on the distance to the problems of fit, two samples and estimation. *Ann. Math. Stat.* Vol. 26 (1955), pp. 631-640.

[2] Matusita, Kameo: On the theory of decision functions, *Ann. Inst. Stat. Math.* Vol. III (1951), pp. 17–35.

[3] Matusita, Kameo and Hirotugu Akaike: Note on the decision problem, *Ann. Inst. Stat. Math.* Vol. IV (1952), pp. 11–14.

[4] Cramér, H: *Mathematical Methods of Statistics*, Princeton Univ. Press, Princeton, 1946.

[5] Hoeffding, Wassily: A non-parametric test of independence, *Ann. Math. Stat.* Vol. 19 (1948), pp. 546–557.

[6] Hoeffding, Wassily: "Optimum" non-parametric tests, *Proc. Second Berkeley Symposium on Math. Stat. and Prob.* edited by J. Neyman, University of California Press (1951), pp. 83–92.

[7] Lehmann, E.L.: On the theory of some non-parametric hypotheses, *Ann. Math. Stat.* Vol. 20 (1949), pp. 28–45.

[8] Hotelling, Harold: Some new methods in matrix calculation, *Ann. Math. Stat.* Vol. 14 (1943), pp. 1–34.

[9] Uspensky, J.V.: *Introduction to Mathematical Probability*, McGraw-Hill Book Co., Inc., New York, 1937.

ERRATA

These Annals Vol. VI, No. 3

P. 233, 1st line: read "Sense of Strangulation" instead of "Sence of Strangulation ⋯ "

Vol. VII, No. 2

P. 70, 7th line from bottom: read "when the risk is allowed to be greater than 0.02 and Theorem II is preferable for large n." instead of "when the risk is required to be ⋯ for large n."