Igor Vajda

On the statistical decision problems with finite parameter space

Kybernetika, Vol. 3 (1967), No. 5, (451)--466

Persistent URL: http://dml.cz/dmlcz/124363

KYBERNETIKA ČÍSLO 5, ROČNÍK 3/1967

On the Statistical Decision Problems with Finite Parameter Space

IGOR VAJDA

In the paper there are obtained some both-sides estimates of the Bayes risk and information in a sample concerning a parameter in terms of total variations of pairs of conditional distributions in the framework of the classical model of statistical decision with finite parameter space. On the base of these relations the rate of convergence of the risk and of information to their limit values is studied, for the case if the sample size tends to infinity.

1. INTRODUCTION AND PRELIMINARIES

In this paper we shall deal with the classical model of statistical decision with a finite parameter probability space $(X, \mathscr{X}, \mu)$ ($\mathscr{X}$ is assumed to be the class of all subsets of the set X and the prior probability $\mu(x)$ is assumed to be positive for every parameter value $x \in X$), with measurable sample space $(Y, \mathscr{Y})$, set $\{\nu_x\}$, $x \in X$, of conditional distributions of the variable $y \in Y$ defined on the σ-algebra $\mathscr{Y}$, decision measurable space $(X, \mathscr{X})$, and with a non-negative loss function w defined on $X \otimes X$.

It is clear that a most important numerical characteristics of the model just described is the so called Bayes risk, i.e. the quantity defined as

$$r = \inf \sum_{x \in X} \mu(x) \int_Y w(x, \varrho(y)) \, \mathrm{d}\nu_x(y) \,, \tag{1.1}$$

where the infimum is extended over the set of all $\mathscr{Y}$-measurable decision functions $\varrho : Y \to X$ (for a more particular discussion of the decision model, see [3]–[6]). Another characteristics of great importance is the average amount of information I contained in the sample y concerning the parameter x defined as

$$I = \int_{X \otimes Y} \log f(x, y) \, \mathrm{d}\omega(x, y) \,, \tag{1.2}$$

where f is the Radon-Nikodym density of the joint probability distribution

$$\omega(E) = \sum_{x \in X} \mu(x)\, \nu_x(\{y \in Y : (x, y) \in E\}), \quad E \in \mathscr{X} \otimes \mathscr{Y},$$

with respect to the Cartesian product distribution $\mu \otimes \tilde{\omega} \gg \omega$ on $\mathscr{X} \otimes \mathscr{Y}$, where by $\tilde{\omega}$ we denote the marginal distribution of ω on $\mathscr{Y}$. (Let us note that all logarithms in this paper are taken to the base e.)

From the intuitive point of view it is clear that, though the information I does not depend on the loss function w, for a sufficiently wide class of loss functions there exists a relation between r and I. The data reduction theory developed by A. Perez [3], [4], shows that the indicated relation plays a growing role in solutions of certain class of decision problems.

The purpose of this paper is to estimate the Bayes risk and information in the framework of the decision model with finite parameter space and to investigate the indicated relation between them. This general questions are studied in Sec. 2; the results of Sec. 2 are then used in Sec. 3 devoted to the study of the rate of convergence of information and Bayes risk in some classes of decision models as it is more precisely described below.

Let the measurable sample space of the model we have considered be of the form

$$(Y^n, \mathscr{Y}^n) = \bigotimes_{i=1}^{n} (Y_i, \mathscr{Y}_i), \quad n = 1, 2, \ldots, \infty, \tag{1.3}$$

(i.e. suppose that the samples are of size n, $y = (y_1, y_2, \ldots, y_n)$), where Y_i is the set of all y_i's with a given σ-algebra $\mathscr{Y}_i$, and denote the joint probability distribution of the n-vector $(y_1, \ldots, y_n)$ under the condition that $x \in X$ is the realized value of the parameter by ν_x^n. It is clear that, for every $x \in X$, ν_x^n is the restriction of ν_x^∞ on the σ-algebra $\mathscr{Y}^n \subset \mathscr{Y}^\infty$, where the latter inclusion (as well as the inclusions $\mathscr{Y}_i \subset \mathscr{Y}^n$, $i = 1, 2, \ldots, n$, $n = 1, 2, \ldots, \infty$, that will be used below) is written in accordance with a well-known identification convention for product σ-algebras.

We shall denote by I_n or r_n the information or the Bayes risk respectively corresponding to the measurable space (1.3). It is known that I_n or r_n, $n = 1, 2, \ldots$, is a non-decreasing or non-increasing sequence respectively and that $\lim I_n = I_\infty$, $\lim r_n = r_\infty$ as n tends to infinity (cf. [3], [4]). From the point of view of applications it is important to ask which is the rate of convergence above. This question is studied in Sec. 3 under the assumption that, for every realized value $x \in X$, the sequence $y_1, y_2, \ldots$, of samples is independent random sequence, i.e. that

$$\nu_x^\infty = \bigotimes_{i=1}^{\infty} \nu_{xi} \quad \text{for every} \quad x \in X, \tag{1.4}$$

where ν_{xi}, $i = 1, 2, \ldots, \infty$, denotes over all the paper the restriction of ν_x^∞ on the sub-σ-algebra $\mathscr{Y}_i \subset \mathscr{Y}^\infty$, and that it is a $*$-mixing random sequence, i.e. that there exists a positive integer N and a non-increasing real valued function φ defined for

all integers $k \geqq N$ with $\lim \varphi(k) = 0$ for $k \to \infty$ such that, for every $k \geqq N$, $l \leqq m$, $m + k \leqq n$ and $x \in X$, the following inequality takes place

$$\left|v_x^n(E \cap F) - v_x^n(E)\, v_x^n(F)\right| \leqq \varphi(k)\, v_x^n(E)\, v(F) \tag{1.5}$$

for every

$$E = \{(y_1, \ldots, y_n) : (y_l, y_{l+1}, \ldots, y_m) \in \tilde{E}\}$$

and

$$F = \{(y_1, \ldots, y_n) : (y_{m+k}, y_{m+k+1}, \ldots, y_n) \in \tilde{F}\}\,,$$

where

$$\tilde{E} \in \bigotimes_{i=l}^{m} \mathscr{Y}_i$$

and

$$\tilde{F} \in \bigotimes_{i=m+k}^{n} \mathscr{Y}_i\,.$$

We shall show that in both these cases the exponential rate of convergence takes place.

In the remainder of this section we list some properties of a Δ-divergence of two probability distributions for references later. The concept of Δ-divergence plays an important role over all this paper.

If ν_1, ν_2 are probability measures on a measurable space $(Y, \mathscr{Y})$, then Δ-divergence $\Delta(\nu_1, \nu_2)$ is defined by

$$\Delta(\nu_1, \nu_2) = \frac{1}{2}\int_Y |\xi_1 - \xi_2|\, \mathrm{d}(\nu_1 + \nu_2)\,, \tag{1.6}$$

where ξ_i is the $\mathscr{Y}$-measurable version of the Radon-Nikodym density $\mathrm{d}\nu_i/\mathrm{d}(\nu_1 + \nu_2)$ for $i = 1, 2$. Δ is clearly reflexive, symmetric, and the triangle inequality satisfying distance measure in the space of all probability distributions on $(Y, \mathscr{Y})$. It is clear that $2\,\Delta(\nu_1, \nu_2) = |\nu_1 - \nu_2|\,(Y)$, where $|\nu_1 - \nu_2|$ denotes the total variation of the signed measure $\nu_1 - \nu_2$. This implies in particular that Δ takes values between 0 and 1, and that $\Delta(\nu_1, \nu_2) = 0$ if and only if $\nu_1 = \nu_2$ (on $\mathscr{Y}$) and $\Delta(\nu_1, \nu_2) = 1$ if and only if $\nu_1 \perp \nu_2$, and also the following very useful statement:

(i) There exists a set $F \in \mathscr{Y}$ such that

$$\nu_1(E) - \nu_2(E) \leqq \nu_1(F) - \nu_2(F) = \Delta(\nu_1, \nu_2) \quad \text{for every} \quad E \in \mathscr{Y}\,.$$

The following two assertions follow from the theory of semi-martingales ([2], Th. 4.1s, Chap. VII), as $|\xi|$ is convex function of ξ (cf. the definition of Δ-divergence).

(ii) If $\mathscr{Y}^{(n)}$, $n = 1, 2, \ldots$, is a non-decreasing sequence of sub-σ-algebras of the σ-algebra $\mathscr{Y}$ such that $\mathscr{Y}$ is generated by

$$\bigcup_{n=1}^{\infty} \mathscr{Y}^{(n)}$$

and if we denote by $\nu_1^{(n)}$, $\nu_2^{(n)}$ the restriction of the distributions ν_1, ν_2 on $\mathscr{Y}^{(n)}$, then $\Delta(\nu_1^{(n)}, \nu_2^{(n)})$ is for $n = 1, 2, \ldots$ a non-decreasing sequence and

$$\lim_{n\to\infty} \Delta(\nu_1^{(n)}, \nu_2^{(n)}) = \Delta(\nu_1, \nu_2) .$$

(iii) If $(Z, \mathscr{Z})$ is a measurable space and if T is a measurable transformation of $(Y, \mathscr{Y})$ to $(Z, \mathscr{Z})$, then

$$\Delta(\nu_1 T^{-1}, \nu_2 T^{-1}) \leqq \Delta(\nu_1, \nu_2) .$$

(iv) If there exist two numbers $a, b \geqq 0$ and probability measures ν_{ij}, $i, j = 1, 2$, such that

$$\nu_1 = a\nu_{11} + b\nu_{12} ,$$

$$\nu_2 = a\nu_{21} + b\nu_{22} ,$$

where $\nu_{i1} \perp \nu_{j2}$, $i, j = 1, 2$, then

$$\Delta(\nu_1, \nu_2) = a\,\Delta(\nu_{11}, \nu_{21}) + b\,\Delta(\nu_{12}, \nu_{22}) .$$

Proof. The proof of this equality can be based on (i) and on the assumption of singularity. Details are omitted here.

(v) If $B(.\,; p, n)$, $B(.\,; q, n)$ are binomial distributions with $p \neq q$, then there exist numbers $A > 0$, $0 < \lambda < 1$ such that

$$\Delta(B(.\,; p, n), B(.\,; q, n)) > 1 - A\lambda^n \quad \text{for every} \quad n = 1, 2, \ldots$$

Proof. According to (i), if $p < q$, then

$$\Delta(B(.\,; p, n), B(.\,; q, n)) = 1 - \min_{k=0,1,\ldots,n} (b_1(n, k) + b_2(n, k)) ,$$

where

$$b_1(n, k) = \sum_{i=k+1}^{n} \binom{n}{i} p^i (1 - p)^{n-i}$$

$$b_2(n, k) = \sum_{i=0}^{k} \binom{n}{i} q^i (1 - q)^{n-i} .$$

If one of the numbers p, q is equal 0 or 1, (v) is trivial. Let, for $0 < p < q < 1$, k_n be the least integer greater than or equal to ϱn, where $0 < \varrho < 1$ is the unique solution of the equation

$$\left(\frac{p}{q}\right)^{\varrho} = \left(\frac{1 - q}{1 - p}\right)^{1-\varrho} .$$

Using Stirling's formula and some elementary properties of the numbers $B(.\,;.\,,n)$ it can be shown that there is $A > 0$ and $0 < \lambda < 1$ such that for every $n = 1, 2, \ldots$

$$b_1(n, k_n) + b_2(n, k_n) < A\lambda^n\,;$$

the remainder of the proof is now clear.

In what follows we shall denote by card (X) the cardinal of X and by $H(\mu)$, in accordance with [6], the entropy of the finite probability space (X, μ).

2. GENERAL INEQUALITIES

Lemma 1. *There exists a disjoint system of sets* $\{E_x\}$, $E_x \in \mathscr{Y}$, $x \in X$, *such that*

$$v_x(Y - E_x) < \operatorname{card}(X)\,(1 - \min_{x' \neq x} \Delta(v_x, v_{x'}))\,, \tag{2.1}$$

$$v_x(E_{x'}) < 1 - \Delta(v_x, v_{x'}) \quad \textit{for every} \quad x \neq x'\,. \tag{2.2}$$

Proof. According to (i), there exists for every $x, x' \in X$ a set $E_{xx'}$ such that

$$\Delta(v_x, v_{x'}) = v_x(E_{xx'}) - v_{x'}(E_{xx'})\,; \tag{2.3}$$

hence

$$\Delta(v_x, v_{x'}) \leqq v_x(E_{xx'})$$

or

$$v_x(Y - E_{xx'}) \leqq 1 - \Delta(v_x, v_{x'}) \tag{2.4}$$

and consequently

$$v_x\Big(\bigcup_{x' \neq x} (Y - E_{xx'})\Big) < \operatorname{card}(X)\,(1 - \min_{x' \neq x} \Delta(v_x, v_{x'}))\,.$$

Since

$$\bigcup_{x' \neq x} (Y - E_{xx'}) = Y - \bigcap_{x' \neq x} E_{xx'}\,,$$

it remains to put in the latter inequality

$$E_x = \bigcap_{x' \neq x} E_{xx'} \tag{2.5}$$

and (2.1) is proved. From (2.3) we can easily obtain the following relation

$$E_{xx'} = Y - E_{x'x} \quad \text{for all} \quad x, x' \in X\,. \tag{2.6}$$

Since

$$v_x(E_{x'}) < v_x(E_{x'x}) = v_x(Y - E_{xx'})\,,$$

to prove (2.2) it remains to use (2.4). In order to prove that the system of measurable sets $\{E_x\}$, $x \in X$, defined by (2.5) is disjoint let us point out the following inclusions

$$E_x = \bigcap_{x'' \neq x} E_{xx''} \subset E_{xx'},$$

$$E_{x'} = \bigcap_{x'' \neq x'} E_{x'x''} \subset E_{x'x}$$

and then let us use (2.6).

Theorem 1. *If we denote*

$$A = \sqrt{\operatorname{card}(X)\,(1 + 2\sum_x \sqrt{[\mu(x)\,(1 - \mu(x))]})}, \tag{2.7}$$

$$\mu = \min_{x \in X} \mu(x) > 0, \tag{2.8}$$

then

$$\mu \log 2(1 - \min_{x' \neq x} \Delta(v_x, v_{x'}) \leqq H(\mu) - I \leqq A \sqrt{(1 - \min_{x' \neq x} \Delta(v_x, v_{x'}))}. \tag{2.9}$$

Proof. Since, for every $x, x' \in X$ the following inequality holds

$$\frac{\mu}{2} \leqq \frac{\mu(x)\,\mu(x')}{\mu(x) + \mu(x')},$$

the left inequality in (2.9) follows from Th. 3 in [6]. We next prove the right inequality. It follows from Lemma 2 in [6] that, for the class $\{E_x\}$, $E_x \in \mathscr{Y}$, $x \in X$, defined in Lemma 1 above,

$$H(\mu) - I \leqq \sum_{x,x' \in X} \sqrt{[\mu(x)\, v_x(E_{x'}) \sum_{x'' \neq x} (\mu(x'')\, v_{x''}(E_{x'})]} + \\ + \sum_{x \in X} \sqrt{[\mu(x)\, v_x(E_0) \sum_{x'' \neq x} (\mu(x'')\, v_{x''}(E_0)]},$$

where

$$E_0 = \bigcap_{x \in X} (Y - E_x),$$

and consequently we can write

$$H(\mu) - I \leqq \sum_{x \in X} \sqrt{[\mu(x)\, v_x(E_x) \sum_{x'' \neq x} \mu(x'')\, v_{x''}(E_x)]} + \\ + \sum_{x \in X} \sum_{x' \neq x} \sqrt{[\mu(x)\, v_x(E_{x'}) \sum_{x'' \neq x} \mu(x'')\, v_{x''}(E_{x'})]} + \\ + \sum_{x \in X} \sqrt{[\mu(x)\, v_x(Y - E_x) \sum_{x'' \neq x} \mu(x'')]}.$$

If we denote the terms on the right side subsequently by (I), (II), and (III), then by (2.1)

$$(2.10) \qquad (III) \leqq \sqrt{[\operatorname{card}(X)(1 - \min_{x \neq x'} \Delta(v_x, v_{x'})]} \sum_{x \in X} \sqrt{[\mu(x)(1 - \mu(x))]} .$$

If we apply on the sum (I) the Schwarz's inequality, we obtain

$$(2.11) \qquad (I) \leqq \sqrt{\sum_{x'' \in X} \mu(x'') v_{x''}(Y - E_{x''})} \leqq \sqrt{[\operatorname{card}(X)(1 - \min_{x' \neq x} \Delta(v_x, v_{x'}))]} .$$

One more application of Schwarz's inequality yields

$$\begin{aligned} (II) &\leqq \sum_{x \in X} \sqrt{[\mu(x) v_x(\bigcup_{x' \neq x} E_{x'}) \sum_{x'' \neq x} \mu(x'') v_{x''}(\bigcup_{x' \neq x} E_{x'})]} \leqq \\ &\leqq \sum_{x \in X} \sqrt{[\mu(x) v_x(Y - E_x) \sum_{x'' \neq x} \mu(x'')]} \leqq \\ &\leqq \sqrt{[\operatorname{card}(X)(1 - \min_{x' \neq x} \Delta(v_x, v_{x'}))]} \sum_{x \in X} \sqrt{[\mu(x)(1 - \mu(x))]} . \end{aligned}$$

Using this together with (2.10) and (2.11) we obtain the desired result.

Theorem 2. *If the loss function w is bounded from above by w_0 then*

$$(2.12) \qquad \frac{\gamma\mu}{2}(1 - \min_{x' \neq x} \Delta(v_x, v_{x''})) \leqq r \leqq w_0 \operatorname{card}(X)(1 - \min_{x \neq x'} \Delta(v_x, v_{x'})) ,$$

$$(2.13) \qquad \frac{\gamma\mu}{2A}(H(\mu) - I)^2 \leqq r \leqq \frac{w_0}{2 \log 2}(H(\mu) - I) ,$$

where A is defined by (2.7), *μ by* (2.8), *and γ by*

$$\gamma = \min_{\substack{x, x' \in X \\ x' \neq x}} w(x, x') .$$

(*Left inequalities remain true also without restriction $w \leqq w_0$.*)

Proof. It is clear that for every measurable disjoint decomposition $\{E_x\}$, $x \in X$, of Y the following inequality holds

$$r \leqq w_0 \sum_{x \in X} \mu(x)(1 - v_x(E_x))$$

and in order to prove the right inequality in (2.12) it remains to use Lemma 1. The left inequality immediately follows from Th. 1 in [6]. The right inequality (2.13) was proved in [6], Th. 2, and the left inequality follows from Th. 1 and from (2.12).

We shall say that a sequence a_n, $n = 1, 2, \ldots$ of numbers converges exponentially to a number a if there exist numbers $A > 0$ and $0 < \lambda < 1$ such that

$$|a_n - a| < A\lambda^n \quad \text{for every} \quad n = 1, 2, \ldots$$

The loss function w will be said reflexive if $w(x, x') > 0$ for every $x \neq x'$.

An immediate consequence of Theorem 2 is the following Corollary that will be very useful later.

Corollary. Let $\mathscr{Y}^{(n)}$, $v_x^{(n)}$, $x \in X$, $n = 1, 2, \ldots$ be defined as in (ii) and let $r_{(n)}$ and $I_{(n)}$ denote the corresponding Bayes risk and information respectively. Then $I_{(n)}$ converges to $H(\mu)$ exponentially if and only if $\Delta(v_x^{(n)}, v_{x'}^{(n)})$ converges to 1 exponentially for every $x \neq x'$. If the loss function is bounded, then this condition is sufficient in order that $r_{(n)}$ converges to zero exponentially. If, moreover, the loss function is reflexive, then this condition is necessary and sufficient for the exponential rate of convergence of $r_{(n)}$ to zero.

Remark. This corollary need not be true in case of an infinite parameter space. In order to prove this we proceed in the following manner. For ease of writing let us assume that $X = \{1, 2, \ldots\}$ and let $w(i, j) = 0$ or 1 depending on whether $i = j$ or $i \neq j$. Let $\mathscr{Y}^{(n)}$ be the σ-algebra of Lebesgue measurable sets in the n-dimensional Euclidian space and let $v_i^{(n)}$ be the n-dimensional Cartesian product of uniform distribution on the interval $\langle 2^{-i} - 1, 2^{-i} \rangle$ for every $i \in X$. Under this assumptions one can show that

$$\Delta(v_i^{(n)}, v_j^{(n)}) = 1 - (1 - |2^{-i} - 2^{-j}|)^n \quad \text{for every} \quad n = 1, 2, \ldots \quad \text{and} \quad i, j \in X .$$

Hence on the hand it is clear that $\Delta(v_i^{(n)}, v_j^{(n)})$ converges for every $i \neq j$ exponentially to 1 as $n \to \infty$ and on the other hand one may show on the base of Th. 1 and Th. 2 in [6] that

$$r_{(n)} \geqq \frac{\mu(i)\,\mu(j)}{\mu(i) + \mu(j)} (1 - |2^{-i} - 2^{-j}|)^n , \quad n = 1, 2, \ldots .$$

$$H(\mu) - I_{(n)} \geqq 2 \log 2 \frac{\mu(i)\,\mu(j)}{\mu(i) + \mu(j)} (1 - |2^{-i} - 2^{-j}|)^n , \quad n = 1, 2, \ldots,$$

for every $i \neq j$ so that, for every $0 < \lambda < 1$, there exist numbers $\tilde{A} > 0$ and $\lambda < \tilde{\lambda} < 1$ such that

$$r_{(n)} \geqq \tilde{A}\tilde{\lambda}^n ,$$

$$H(\mu) - I_{(n)} \geqq \tilde{A}\tilde{\lambda}^n \quad \text{for every} \quad n = 1, 2, \ldots$$

We are now in a position to prove by a simple contradiction that $r_{(n)}$ does not converge to zero as well as $I_{(n)}$ to $H(\mu)$ exponentially for $n \to \infty$.

3. DECISION MODELS WITH INDEPENDENT AND *-MIXING SAMPLES

In this section the classical model of statistical decision with a sample space $(Y^n, \mathscr{Y}^n)$ as it is described in Sec. 1 will be studied. We shall follow the terminology and notation employed above.

It was shown in [7] that in the independent case the condition

$$\inf_{n=1,2,\ldots} \frac{1}{n} \sum_{i=1}^{n} \Delta(\nu_{xi}, \nu_{x'i}) = \alpha > 0 \tag{3.1}$$

implies that

$$\Delta(\nu_x^n, \nu_{x'}^n) > 1 - 4\beta^n \quad \text{for every} \quad n = 1, 2, \ldots, \tag{3.2}$$

where β lies between 0 and $e^{-\alpha/4}$. In view of the Corollary in the preceding section it is clear that if, for every $x \in X$, the sequence of samples is independent and if the condition (3.1) holds for every $x \neq x'$, then $I_\infty = H(\mu)$ and I_n converges to I_∞ exponentially. If moreover the loss function is bounded, then also $r_\infty = 0$ and r_n converges to zero exponentially.

If, for every $x \in X$, the sequence of samples is moreover stationary, then the condition (3.1) is equivalent to the condition

$$\nu_x^1 \neq \nu_{x'}^1 . \tag{3.3}$$

It follows from Th. 1 and from the considerations above, that $I_\infty = H(\mu)$ if and only if (3.3) holds for every $x \neq x'$. If this condition is satisffied, then in view of (2.13) also $r_\infty = 0$ for every bounded loss function. If, moreover, the loss function is reflexive, then in order that $r_\infty = 0$ it is necessary and sufficient that (3.3) holds for every $x \neq x'$. Always when (3.3) holds $H(\mu) - I_n$ as well as r_n converges to zero exponentially. The short discussion of the independent stationary case we can conclude by a note that the exponential convergence rate for I_n to $H(\mu)$ under the condition that the sample space Y_1 is finite was first proved by A. Rényi [5], and then generalized by author in [7]. The exponential convergence rate for r_n under the same conditions was first proved by A. Perez [3].

In Sec. 1 we have defined the model of statistical decision with $*$-mixing samples. According to the definition, in this case the sequence of samples is $*$-mixing random sequence in the sense of [1] for every realized value of the parameter. The question is which is the class of all $*$-mixing sequences. It is clear that independent random sequence is $*$-mixing. It is easy to show that if a Markov chain possesses a long-run distribution then it is $*$-mixing. Especially if a finite Markov chain is geometrically ergodic (in the well known sense of Kendall), then there is $0 < \varrho < 1$ such that $\varphi(k) = \text{const}\, \varrho^k$ satisfies for $N = 1$ the requirements of the definition of $*$-mixing random sequence (cf. (1.5)). Hence the class of all $*$-mixing sequences is wider than the class of all independent sequences and consequently the $*$-mixing case must be studied separately.

Our considerations will be based on the following

Lemma 2. *If, for every $x \in X$, the sequence of samples is $*$-mixing and if, for $i = 1, 2, \ldots$,*

$$\Delta(\nu_{xi}, \nu_{x'i}) \geqq \alpha > 0 \quad \textit{for some} \quad x, x' \in X , \tag{3.4}$$

then there exists $0 < \lambda < 1$ *such that*

$$\Delta(v_x^n, v_{x'}^n) > 1 - 8\lambda^n \quad \textit{for every} \quad n = 1, 2, \ldots \tag{3.5}$$

Proof. It follows from (3.4) that

$$\frac{1}{m} \sum_{s=1}^{m} \Delta(v_{xi_s}, v_{x'i_s}) \geqq \alpha$$

uniformly with respect to increasing sequences $i_1, i_2, \ldots, i_m$ of positive integers and hence, in view of the Lemma of [7], there is such $0 < \beta < 1$ that

$$\Delta\Big(\bigotimes_{s=1}^{m} v_{xi_s}, \bigotimes_{s=1}^{m} v_{x'i_s}\Big) > 1 - 4\beta^m \quad \text{for every} \quad m = 1, 2, \ldots \tag{3.6}$$

uniformly in the sense given above. Let $k > N$ (cf. the definition of $*$-mixing random sequence) be positive integer such that

$$1 + \varphi(k) < 1/\beta \tag{3.7}$$

and let n be an arbitrary integer. Define integers m and r by $n = km - r$ where $m \geqq 1$, $0 \leqq r < k$, and let ψ be a mapping of Y^n to $\bigotimes_{i=1}^{m} Y_{ki-r}$ defined by

$$\psi((y_1, \ldots, y_n)) = (y_{k-r}, y_{2k-r}, \ldots, y_{mk-r}) .$$

If we denote by $\mathscr{Y}^{(n)}$ the σ-algebra generated by the class of all sets E of the form

$$E = \bigcap_{i=1}^{m} \{(y_1, \ldots, y_n) : y_i \in F_i\} \quad \text{where} \quad F_i \in \mathscr{Y}_{ki-r} ,$$

then the following assertions hold:

(a) $$Y^{(n)} \subset Y^n ,$$

(b) ψ is a measurable transformation and

$$\psi^{-1}\Big(\bigotimes_{i=1}^{m} Y_{ki-r} - F\Big) = \bigotimes_{i=1}^{n} Y_i - \psi^{-1}(F) \quad \text{for every} \quad F \in \bigotimes_{i=1}^{m} \mathscr{Y}_{ki-r} ,$$

(c) $$v_x^n(\psi^{-1}(F)) \leqq 1 + \varphi(k))^m \bigotimes_{i=1}^{m} v_{xki-r}(F) .$$

The assertions (a) and (b) are obvious. The inequality (c) follows immediately from the definition of $*$-mixing sequence when F is of the form

$$F = \bigotimes_{i=1}^{m} F_i , \quad F_i \in \mathscr{Y}_{ki-r}$$

and consequently also when F is a denumerable union of disjoint sets of this form. An application of a well-known approximation argument yields the general validity of (c).

Using (3.6) together with (i) we obtain that there exists a set

$$F \in \bigotimes_{i=1}^{m} \mathscr{Y}_{ki-r}$$

such that

$$\bigotimes_{i=1}^{m} \nu_{xki-r}\Big(\bigotimes_{i=1}^{m} Y_{ki-r} - F\Big) < 4\beta^m\,,$$

$$\bigotimes_{i=1}^{m} \nu_{x'ki-r}(F) < 4\beta^m$$

or, in view of (a), (b), and (c), that there exists a set $E \in Y^n$ such that

$$\nu_x^n(Y^n - E) < 4\big[(1 + \varphi(k))\,\beta\big]^m\,,$$

$$\nu_{x'}^n(E) < 4\big[(1 + \varphi(k))\,\beta\big]^m\,.$$

These inequalities imply in view of (i) that

$$\Delta(\nu_x^n, \nu_{x'}^n) > 1 - 8\big[(1 + \varphi(k))\,\beta\big]^m\,.$$

If we put $\lambda = \big[(1 + \varphi(k))\,\beta\big]^{1/k+1}$, then by (3.7) the proof of (3.5) is complete.

Theorem 3. *If, for every $x \in X$, the sequence of samples is $*$-mixing and if*

$$\Delta(\nu_{xi}, \nu_{x'i}) \geqq \alpha > 0 \quad \textit{for every} \quad i = 1, 2, \ldots \quad \textit{and} \quad x \neq x'\,, \tag{3.8}$$

then $I_\infty = H(\mu)$ and I_n converges to I_∞ exponentially. If the loss function is bounded, then $r_\infty = 0$ and r_n converges to zero exponentially.

Proof. It follows from the assumptions of the Theorem that the assumptions of the preceding Lemma are satisfied for every $x \neq x'$ and hence there exists $0 < \lambda < 1$ such that

$$1 - \min_{x \neq x'} \Delta(\nu_x^n, \nu_{x'}^n) < 8\lambda^n \quad \text{for every} \quad n = 1, 2, \ldots \tag{3.9}$$

and, in accordance with (ii), $\Delta(\nu_x^\infty, \nu_{x'}^\infty) = 1$ for every $x \neq x'$. Using this together with (2.9) or (2.12) we obtain $I_\infty = H(\mu)$ or $r_\infty = 0$ respectively. The desired exponential rate of convergence follows from (3.9), from the Corollary in Sec. 2, and from (2.12).

We shall say that a $*$-mixing sequence of random samples is stationary for realized value $x \in X$ of parameter if $(Y_i, \mathscr{Y}_i) = (Y_j, \mathscr{Y}_j)$ and $\nu_{xi} = \nu_{xj}$ for every $i, j = 1, 2, \ldots$ It can be easily verified that this stationarity is rather weaker than the usual stationarity in the strict sense.

It is clear that if the sequence of samples is $*$-mixing and stationary for every $x \in X$, then (3.8) is satisfied if and only if $\nu_x^1 \neq \nu_{x'}^1$ for every $x \neq x'$ and hence we have proved the following

Theorem 3s. *If the sequence of samples is $*$-mixing and stationary for every $x \in X$, and if*

$$\nu_x^1 \neq \nu_{x'}^1 \quad \text{for every} \quad x \neq x' ,$$

then $I_\infty = H(\mu)$ and I_n converges to I_∞ exponentially. If the loss function is bounded, then $r_\infty = 0$ and r_n converges to zero exponentially.

Remark 1. It is to be noted that the analogies between the independent and $*$-mixing case are not complete. Namely, a routine verification (using the Lemma of [7]) gives in the stationary independent case that the necessary and sufficient condition for the validity of $\Delta(\nu_x^\infty, \nu_{x'}^\infty) = 1$ is (3.3) and that if the latter condition is satisfied, then $\Delta(\nu_x^n, \nu_{x'}^n)$ converges to 1 exponentially as $n \to \infty$. An analogous assertion does not hold in the (stationary) $*$-mixing case. An example of statistical decision problem with stationary $*$-mixing samples given below shows that, even when $\nu_x^1 = \nu_{x'}^1$, holds, the equality $\Delta(\nu_x^\infty, \nu_{x'}^\infty) = 1$ as well as the exponential convergence rate of $\Delta(\nu_x^n, \nu_{x'}^n)$ to 1 is possible. Hence in the $*$-mixing case the condition (3.4) is sufficient but not necessary for the exponential rate of convergence of $\Delta(\nu_x^n, \nu_{x'}^n)$ to 1.

Example 1. Let $X = \{x', x''\}$, $Y_i = \{1, 2\}$ for every $i = 1, 2, \ldots$ and let, for every x under consideration, the sequence $y_1, y_2, \ldots$ of samples be a homogeneous Markov chain determined by an initial distribution ν_x^1 on $\{1, 2\}$ and by a matrix $W(x)$ of transition probabilities, whose element $w_{ij}(x)$, lying at the intersection of the i-th row and the j-th column, is given by

$$w_{ij}(x) = \mathbf{P}[y_{n+1} = j \mid x, y_n = i] \quad \text{for every} \quad n = 1, 2, \ldots \text{ and } i, j = 1, 2 .$$

Let us put $\nu_{x'}^1(1) = \nu_{x'}^1(2) = \nu_{x''}^1(1) = \nu_{x''}^1(2) = \frac{1}{2}$,

$$W(x') = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix},$$

$$W(x'') = \begin{pmatrix} q & 1-q \\ 1-q & q \end{pmatrix},$$

where p and q are assumed to be different. It is easily proved that in this case the sequence of samples in stationary for every $x \in X$. As the n-step transition probability matrices $W^n(x')$ or $W^n(x'')$ are given by

$$W^n(x') = \begin{pmatrix} \frac{1}{2} + (2p-1)^n & \frac{1}{2} - (2p-1)^n \\ \frac{1}{2} - (2p-1)^n & \frac{1}{2} + (2p-1)^n \end{pmatrix},$$

$$W^n(x'') = \begin{pmatrix} \frac{1}{2} + (2q-1)^n & \frac{1}{2} - (2q-1)^n \\ \frac{1}{2} - (2q-1)^n & \frac{1}{2} + (2q-1)^n \end{pmatrix}$$

respectively, it is obvious that both the homogeneous Markov chains are geometrically ergodic and consequently that the condition (1.5) is satisfied for $N = 1$ and

$$\varphi(k) = \text{const}\,(\max\{|2p-1|, |2q-1|\})^k\,, \quad k = 1, 2, \ldots\,;$$

this proves that the sequence of samples is $*$-mixing for every x. On the other hand it is obvious that $\nu^1_{x'} = \nu^1_{x''}$. It remains to prove that $\Delta(\nu^n_{x'}, \nu^n_{x''})$ converges to 1 exponentially. Let us denote for every $n = 1, 2, \ldots$

$$Y^n_* = \bigotimes_{i=1}^{n} \{0, 1\}_i \quad \text{where} \quad \{0, 1\}_i = \{0, 1\}\,, \quad i = 1, 2, \ldots$$

and let us define a mapping T_n of the space Y^{n+1} to the space Y^n_* by

$$T_n(y_1, \ldots, y_{n+1}) = (y^*_1, \ldots, y^*_n)\,,$$

where

$$y^*_i = y_i \otimes y_{i+1} \quad \text{for every} \quad i = 1, 2, \ldots, n\,,$$

and where $y_i \otimes y_{i+1} = 1$ or 0 depending on whether $y_i = y_{i+1}$ or $y_i \neq y_{i+1}$. Let us denote by $\tilde{\nu}^n_x$ a probability measure induced by T_n on Y^n_*, i.e. let

$$\tilde{\nu}^n_x(E) = \nu^{n+1}_x(T_n^{-1}E) \quad \text{for every} \quad E \subset Y^n_* \quad \text{and} \quad x \in \{x', x''\}\,.$$

A routine verification gives that

$$\tilde{\nu}^n_{x'}(y^*_1, \ldots, y^*_n) = p^{\Pi_n}(1-p)^{n-\Pi_n}\,,$$
$$\tilde{\nu}^n_{x''}(y^*_1, \ldots, y^*_n) = q^{\Pi_n}(1-q)^{n-\Pi_n}\,,$$

where

$$\Pi_n = \sum_{i=1}^{n} y^*_i\,. \tag{3.10}$$

According to (iii) we get

$$\Delta(\nu^{n+1}_{x'}, \nu^{n+1}_{x''}) \geqq \Delta(\tilde{\nu}^n_{x'}, \tilde{\nu}^n_{x''}) \geqq \Delta(B(.\,; p, n), B(.\,; q, n))\,,$$

since the distributions induced by $\tilde{\nu}^n_{x'}$ or $\tilde{\nu}^n_{x''}$ on the real line are $B(.\,; p, n)$ or $B(.\,; q, n)$ respectively. In view of the latter inequality, (iii), and (v), the exponential convergence rate of $\Delta(\nu^n_{x'}, \nu^n_{x''})$ to 1 holds.

Remark 2. The question arises if the $*$-mixing condition is also necessary for the exponential convergence rate of $\Delta(\nu^n_{x'}, \nu^n_{x''})$ to 1 (as well as of $H(\mu) - I_n$ or r_n to zero). In the sequel we shall show that this assumption is not true. In order to achieve this we shall give the following

Example 2. Let $X = \{x', x''\}$, $Y_i = \{1, 2, 3, 4\}$, for every $i = 1, 2, \ldots$ and let similarly as in the example above,

$$W(x') = \begin{pmatrix} p & 1-p & 0 & 0 \\ 1-p & p & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$W(x'') = \begin{pmatrix} q & 1-q & 0 & 0 \\ 1-q & q & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

be matrices of transition probabilities and let $v^1_{x'}(3) = v^1_{x''}(4) = \frac{1}{2}$, $v^1_{x'}(1) = v^1_{x'}(2) = v^1_{x''}(1) = v^1_{x''}(2) = \frac{1}{4}$ be initial distributions determining, for every $x \in \{x', x''\}$, a homogeneous Markov chain which will be referred to as a sequence of samples. It is clear that none of these random sequences is *-mixing as it is not in this case satisfied the following necessary condition

$$\lim_{n\to\infty} \left| \frac{v^1_x(i)\, w^n_{ij}(x) - v^1_x(i)\, v_{xn}(j)}{v^1_x(i)\, v_{xn}(j)} \right| = 0$$

(cf. (1.5)), where $v_{xn}(j)$, in accordance with the notation employed above, is given by

$$v_{xn}(j) = \mathbf{P}[y_n = j \mid x], \quad n = 1, 2, \ldots,$$

and where $w^n_{ij}(x)$ is an element of the n-step transition probability matrix $W^n(x)$, lying at the intersection of the i-th row and j-th column. In order to prove this it suffices to put, for $x = x'$ or x'', $i = 2$, $j = 3$ or $i = 2, j = 4$ respectively, and then to use explicit expressions for the corresponding n-step transition probability matrices.

On the other hand we shall prove that in this case $\Delta(v^n_{x'}, v^n_{x''})$ converges to 1 exponentially (the condition (3.8) is however in this case satisfied). It follows from the definition of the joint probability distributions v^n_x, $x \in X$, and from the concrete definition of the sequence $y_1, y_2, \ldots$ of random samples above that in this special case the following equalities take place:

$$v^n_{x'} = \tfrac{1}{2}(v^{(n)}_{11} + v^{(n)}_{12}),$$

$$v^n_{x''} = \tfrac{1}{2}(v^{(n)}_{21} + v^{(n)}_{22}),$$

where $v^{(n)}_{11}(3, 3, \ldots, 3) = 1$ and $v^{(n)}_{21}(4, 4, \ldots, 4) = 1$, and where $v^{(n)}_{12}$ or $v^{(n)}_{22}$ coincides with $v^n_{x'}$ or $v^n_{x''}$ considered in Example 1. Consequently the assumptions of (iv) in Sec. 1 are satisfied and $\Delta(v^{(n)}_{12}, v^{(n)}_{22})$ converges to 1 exponentially. Since it is clear that $\Delta(v^{(n)}_{11}, v^{(n)}_{21}) = 1$, by using (iv) we conclude the proof of the desired result.

(Received December 8th, 1966.)

[1] J. R. Blum, D. L. Hanson, L. H. Koopmans: On the strong law of large numbers for a class of stochastic processes. Zeitschr. Wahrsch. *2* (1963), 1.
[2] J. L. Doob: Stochastic processes. J. Wiley, New York 1953.
[3] A. Perez: Information theory methods in reducing complex decision problems. Transactions of Fourth Prague Conf. on Inf. Theory, Stat. Dec. Functions, Random Processes. Academia, Praha 1967.
[4] A. Perez: Information, ε-sufficiency and data reduction problems. Kybernetika *1* (1965), 4.
[5] A. Rényi: On the amount of information in a sample concerning a parameter. Publ. of Math. Inst. of Hungarian Acad. Sci. *9* (1964), 617—625.
[6] I. Vajda: On the statistical decision problems with discrete parameter space. Kybernetika *3* (1967), 2.
[7] I. Vajda: Rate of convergence of the information in a sample concerning a parameter. Czechosl. Mathem. Journal *17* (1967), 2.

VÝTAH

O statistických rozhodovacích problémech s konečným parametrovým prostorem

IGOR VAJDA

Tato práce navazuje na [6], kde byl studován klasický model statistického rozhodování s abstraktním výběrovým prostorem a s nejvýše spočetným parametrovým prostorem. Ukazuje se, že odhady základních charakteristik uvažovaného modelu, a to Bayesova rizika a střední informace o parametru obsažené ve výběru pomocí některých jednodušších veličin uvedené v [6] dávají zajímavé výsledky zejména v případě konečného parametrového prostoru.

V první části práce je definován základní model, jemu příslušné Bayesovo riziko a informace a dva speciální modely odpovídající rozhodování na základě opakovaných výběrů, které jsou v práci vyšetřeny podrobněji. V prvním z těchto modelů se předpokládá, že jednotlivé výběry jsou statisticky nezávislé, kdežto v druhém se předpokládá jistý typ slabé závislosti. V této části je dále uvedena definice Δ-divergence dvou pravděpodobnostních distribucí a některé její vlastnosti.

V druhé části práce jsou nalezeny vztahy mezi informací resp. Bayesovým rizikem a mezi Δ-divergencemi dvojic podmíněných pravděpodobností modelu (věty 1 a 2). Z těchto výsledků plyne, že asymptotické chování Bayesova rizika a informace (při zjemňování σ-algebry výběrového prostoru, speciálně při zvětšování rozsahu výběru) je dáno asymptotickým chováním příslušných Δ-divergencí. V téže části práce je ovšem na příkladu ukázáno, že tyto závěry platí jen tehdy, když parametrový prostor je konečný.

Ve třetí části se studuje asymptotické chování Δ-divergence podmíněných pravděpodobnostní při velkých rozsazích výběru v obou speciálních modelech definovaných výše. Ukazuje se (viz diskusi na začátku třetí části a věty 3 a 3s), že rychlost konvergence Δ-divergence (a tedy i Bayesova rizika a informace) k jejich limitním hodnotám je v obou těchto případech za dosti obecných podmínek exponenciální.

Ing. Igor Vajda, Ústav teorie informace a automatizace ČSAV, Praha 2, Vyšehradská 49.