Czechoslovak Mathematical Journal

Igor Vajda

Rate of convergence of the information in a sample concerning a parameter

Czechoslovak Mathematical Journal, Vol. 17 (1967), No. 2, 225–231

Persistent URL: http://dml.cz/dmlcz/100771

Czechoslovak Mathematical Journal, 17 (92) 1967, Praha

RATE OF CONVERGENCE OF THE INFORMATION IN A SAMPLE CONCERNING A PARAMETER

IGOR VAJDA, Praha

(Received January 27, 1966)

Let us consider, for $n = 1, 2, \ldots, \infty$, the classical statistical decision problem with a finite parameter probability space $(X, \mathcal{X}, \mu)$, an abstract sample space

$$(Y^n, \mathcal{Y}^n) = \bigotimes_{i=1}^{n} (Y_i, \mathcal{Y}_i),$$

a set of probability measures

$$\nu^n(. \mid x) = \bigotimes_{i=1}^{n} \nu_i(. \mid x), \quad x \in X,$$

on $\mathcal{Y}^n$, a decision space $(X, \mathcal{X})$, and a weight function w. We shall assume without loss of generality that $\mathcal{X}$ contains all subsets of X and that $\mu(x) > 0$ for every $x \in X$. If we define a probability measure ω^n on $\mathcal{X} \otimes \mathcal{Y}^n$ by

$$\omega^n(E) = \sum_{x \in X} \mu(x)\, \nu^n(\{y^n : (x, y^n) \in E\}), \quad E \in \mathcal{X} \otimes \mathcal{Y}^n$$

and if we denote by $\tilde{\omega}^n$ the marginal measure induced by ω^n on $\mathcal{Y}^n$, then the average information I_n in a sample $y^n \in Y^n$ concerning the parameter x can be defined as follows (cf. [2], [3], [4]):

$$(1) \qquad I_n = \int \log f \, d\omega^n \geqq 0,$$

where f is the Radon-Nikodym density of the joint probability measure ω^n with respect to the product measure $\mu \otimes \tilde{\omega}^n$ (note that $\omega^n \ll \mu \otimes \tilde{\omega}^n$ holds). According to Theorem 11 in [2], I_n, $n = 1, 2, \ldots$, is a non-decreasing sequence and $\lim I_n = I_\infty$.

It has recently become clear that there is a relation between the Bayes risk r_n of the problem we have considered and I_n. For example the results of the data reduction theory, developed by Perez [3], yield in our case

$$(2) \qquad 0 \leqq r_n - r_\infty < \sqrt{(2 w_0 r_n (I_\infty - I_n))}$$

where w_0 is a constant defined by $w \leqq w_0$. That is why the evaluation of I_n plays an important role in the statistical decision theory.

This paper deals with the rate of convergence of I_n to I_∞ and with the value of I_∞.

In the sequel we shall use a distance measure Δ of two probability measures, say η_1, η_2, defined on a measurable space $(Y, \mathscr{Y})$:

$$\Delta(\eta_1, \eta_2) = \tfrac{1}{2}|\eta_1 - \eta_2| (Y)$$

where $|\eta_1 - \eta_2|$ denotes the total variation of the signed measure $\eta_1 - \eta_2$. It is clear that Δ is a metric taking values between 0 and 1, and, in view of the Jordan decomposition theorem, there is $F_0 \in \mathscr{Y}$ such that

$$\Delta(\eta_1, \eta_2) = \eta_1(F_0) - \eta_2(F_0) = \sup_{F \in \mathscr{Y}} \{\eta_1(F) - \eta_2(F)\}\,. \tag{3}$$

Let us point out that $\Delta(\eta_1, \eta_2)$ is a measure of divergence of η_1 and η_2, $\Delta(\eta_1, \eta_2) = 0$ if and only if $\eta_1 = \eta_2$, $\Delta(\eta_1, \eta_2) = 1$ if and only if $\eta_1 \perp \eta_2$.

If we denote by $H(\mu)$ the entropy of $(X, \mathscr{X}, \mu)$ i.e.

$$H(\mu) = -\sum_{x \in X} \mu(x) \log \mu(x)$$

then the results of the paper may be summarized as follows.

Theorem. *If*

$$\inf_{n=1,2,\dots} \frac{1}{n} \sum_{i=1}^{n} \Delta(\nu_i(.\mid x'), \nu_i(.\mid x'')) > 0 \tag{4}$$

for every $x', x'' \in X$, $x' \neq x''$, *then there exist numbers* $A > 0$ *and* $0 < \lambda < 1$ *such that*

$$0 \leqq I_\infty - I_n < A\lambda^n\,, \tag{5}$$

where

$$I_\infty = H(\mu)\,. \tag{6}$$

The inequality

$$0 \leqq I_\infty \leqq H(\mu) \tag{7}$$

allways holds.

Remark. If $\nu_i(.\mid x)$, $x \in X$, are mutually different for every $i = 1, 2, \dots$ and if there is a disjoint decomposition $\{1, 2, \dots\} = N_1 \cup N_2 \cup \dots \cup N_k$ such that

$$(Y_i, \mathscr{Y}_i, \nu_i(.\mid x)) = (Y_j, \mathscr{Y}_j, \nu_j(.\mid x))$$

for every $i, j \in N_m$, $m = 1, 2, \ldots, k$ and $x \in X$, then condition (4) is satisfied. Therefore the Theorem contains as a special case the results of RÉNYI [4] assuming that the sequence of samples is a stationary finite-state process.

Assertions (5) and (6) are based on the following property of independent processes:

Lemma 1. *If*

$$(8) \qquad \inf_{n=1,2,\ldots} \frac{1}{n} \sum_{i=1}^{n} \Delta(\nu_i(.\mid x'), \nu_i(.\mid x'')) = \alpha > 0 \quad \textit{for some} \quad x', x'',$$

then there is a number $0 < \beta \leqq \exp(-\alpha/4)$ *such that* $\Delta(\nu^n(.\mid x'), \nu^n(.\mid x'')) > > 1 - 4\beta^n$.

Proof. If $\alpha = 1$, then $\nu^n(.\mid x') \perp \nu^n(.\mid x'')$ for every n and Lemma 1 holds for every $\beta > 0$. In the case $\alpha < 1$ we proceed in the following manner. In view of (3), there is $F_i \in \mathcal{Y}_i$ such that

$$\nu_i(F_i \mid x') - \nu_i(F_i \mid x'') = \Delta(\nu_i(.\mid x'), \nu_i(.\mid x'')) \quad \text{for every} \quad i = 1, 2, \ldots$$

so that, in view of (8),

$$(9) \qquad \frac{1}{n} \sum_{i=1}^{n} \nu_i(F_i \mid x') \geqq \frac{1}{n} \sum_{i=1}^{n} \nu_i(F_i \mid x'') + \alpha, \quad n = 1, 2, \ldots$$

Define on $(Y^n, \mathcal{Y}^n)$ a sequence of measurable functions $f_1, f_2, \ldots, f_n$ by

$$f_i(y^n) = \chi_{F_i}((y^n)_i), \quad i = 1, 2, \ldots, n,$$

where $(y^n)_i$ denotes the i-th coordinate of the n-vector y^n and χ is the characteristic function. It can be seen that, for every measure $\nu^n(.\mid x)$ on $\mathcal{Y}^n$, f_i are independent random variables, $0 \leqq f_i \leqq 1$, with expectations $\nu_i(F_i \mid x)$ and with variances bounded from above by $\frac{1}{4}$. Therefore, using the inequality § 18.1. A in [1], Chapter V, we obtain for every $0 < \tau < \frac{1}{4}$ and $n = 1, 2, \ldots$

$$(10) \qquad \nu^n(Y^n - E_n(x, \tau) \mid x) < 2 \exp(-n\tau),$$

where

$$E_n(x, \tau) = \left\{ y^n : \frac{1}{n} \left| \sum_{i=1}^{n} (f_i(y^n) - \nu_i(F_i \mid x)) \right| \leqq \tau \right\}.$$

Let us put $E_n(x') = E_n(x', \tau)$, $E_n(x'') = E_n(x'', \tau)$ for $\tau = \frac{1}{4}\alpha$. As $0 < \alpha < 1$, the condition $0 < \tau < \frac{1}{4}$ is satisfied and using (10), we obtain

$$\nu^n(E_n(x') \mid x') > 1 - 2\beta^n, \quad \nu^n(Y^n - E_n(x'') \mid x'') < 2\beta^n$$

for $\beta = \exp(-\frac{1}{4}\alpha)$. Since in view of (9), $E_n(x')$ and $E_n(x'')$ are disjoint, we have

$$\nu^n(E_n(x') \mid x') - \nu^n(E_n(x') \mid x'') > 1 - 4\beta^n ,$$

which, according to (3), completes the proof.

On the base of Lemma 1 we can immediately prove (6). Namely, Lemma 1 implies that the measures $\nu^\infty(. \mid x)$, $x \in X$ are mutually singular and, consequently, there is a disjoint decomposition $Y^\infty = \bigcup_{x \in X} G_x$ where $G_x \in \mathscr{Y}^\infty$, $\nu^\infty(G_x \mid x) = 1$ for every $x \in X$. Define $\mathscr{X} \otimes \mathscr{Y}^\infty$-measurable function f by

$$f(x, y^\infty) = \frac{1}{\mu(x)} \chi_{G_x}(y^\infty) \quad \text{for every} \quad (x, y^\infty) \in X \otimes Y^\infty .$$

It is easily proved that for every $E \in \mathscr{X} \otimes \mathscr{Y}^\infty$

$$\int_E f \, d(\mu \otimes \tilde{\omega}^\infty) = \sum_{x \in X} \frac{1}{\mu(x)} \int_{\{x\} \otimes (E_x \cap G_x)} d(\mu \otimes \tilde{\omega}^\infty) = \sum_{x \in X} \mu(x) \, \nu^\infty(E_x \cap G_x \mid x) = \omega^\infty(E)$$

where

$$E_x = \{y^\infty : (x, y^\infty) \in E\} ;$$

hence f is the Radon-Nikodym density of ω^∞ with respect to $\mu \otimes \tilde{\omega}^\infty$ and we can write

$$I_\infty = \sum_{x \in X} \int_{\{x\} \otimes G_x} \log f \, d\omega^\infty = \sum_{x \in X} \omega^\infty(\{x\} \otimes G_x) \log \frac{1}{\mu(x)} .$$

The desired result follows from the equality $\omega^\infty(\{x\} \otimes G_x) = \mu(x)$.

The proof of (5) is based on the following

Lemma 2. *If $Y_1, Y_2, \ldots$ are finite sets and if (4) holds, then there exist $A > 0$ and $0 < \lambda < 1$ such that (5) is valid.*

Proof. We may clearly suppose that $\mathscr{Y}_i$ contains all subsets of Y_i, $i = 1, 2, \ldots$ In the sequel we shall use the following convention: By writing $a_n < \Theta(n)$ for a sequence $a_n \geqq 0$, $n = 1, 2, \ldots$, we shall always mean that there is $A > 0$ and $0 < \lambda < 1$ such that $a_n < A\lambda^n$, for every $n = 1, 2, \ldots$

A routine verification (using (6) and the expression

$$f(x, y^n) = \frac{\nu^n(y^n \mid x)}{\sum_{x' \in X} \mu(x') \, \nu^n(y^n \mid x')}$$

for the Radon-Nikodym density $d\omega^n / d(\mu \otimes \tilde{\omega}^n)$ provided Y^n is finite) gives

$$(11) \qquad I_\infty - I_n = H(\mu) - I_n = \sum_{y^n \in Y^n} \sum_{x \in X} \psi(x, y^n) ,$$

where

$$\psi(x, y^n) = \mu(x)\, v^n(y^n \mid x) \log \left(\frac{\sum_{x' \in X} \mu(x')\, v^n(y^n \mid x')}{\mu(x)\, v^n(y^n \mid x)} \right) \geqq 0\,. \tag{12}$$

The left inequality in (5) follows from (11) and (12). Further, in view of Lemma 1 and (3), there exist sequences $E_n(x) \in \mathscr{Y}^n$ such that

$$v^n(E_n(x) \mid x') < \Theta(n) \tag{13}$$

$$v^n(Y^n - E_n(x') \mid x') < \Theta(n) \quad \text{for every} \quad x, x' \in X\,, \quad x \neq x'\,. \tag{14}$$

In view of the fact that $a_n^{(i)} < \Theta(n)$ for $i = 1, 2, \ldots, k$ implies $\sum_{i=1}^{k} a_n^{(i)} < \Theta(n)$, it remains to prove that, for every $x, x^* \in X$,

$$\sum_{y^n \in E_n(x^*)} \psi(x, y^n) < \Theta(n)\,, \tag{15}$$

$$\sum_{y^n \in Y^n - \bigcup_{x^* \in X} E_n(x^*)} \psi(x, y^n) < \Theta(n)\,. \tag{16}$$

To prove (16) we use the following easily verified inequality:

$$\psi(x, y^n) \leqq \sum_{x' \neq x} \mu(x')\, v^n(y^n \mid x')\,. \tag{17}$$

In view of (14), (17), and in view of the inclusion

$$Y^n - \bigcup_{x^* \in X} E_n(x^*) \subset Y^n - E_n(x')\,,$$

(16) is valid.

To prove (15) under $x = x^*$ we use (17) obtaining

$$\sum_{E_n(x)} \psi(x, y^n) \leqq \sum_{x' \neq x} \mu(x)\, v^n(E_n(x) \mid x')$$

and then apply (13).

Suppose now that $x \neq x^*$. Since $\log(1 + z) \leqq \sqrt{z}$ holds for every real $z > 0$, the following inequality holds

$$\psi(x, y^n) \leqq \sqrt{(\mu(x)\, v^n(y^n \mid x))}\, \sqrt{[\sum_{x' \neq x} \mu(x')\, v^n(y^n \mid x')]}$$

(cf. (12)) and hence, using Schwarz's inequality, we can write

$$\sum_{E_n(x^*)} \psi(x, y^n) \leqq \sqrt{[\sum_{E_n(x^*)} \mu(x)\, v^n(y^n \mid x)]}\, \sqrt{[\sum_{E_n(x^*)} \sum_{x' \neq x} \mu(x')\, v^n(y^n \mid x'')]} \leqq$$
$$\leqq \sqrt{[\mu(x)\, v^n(E_n(x^*) \mid x)]} < \Theta(n)$$

(cf. (13)) and the proof of the Lemma is complete.

To prove (5) we proceed in the following manner. According to (3) and (4), there exist $F_i(x', x'') \in \mathscr{Y}_i$, $i = 1, 2, \ldots$ such that

$$\inf_{n=1,2,\ldots} \frac{1}{n} \sum_{i=1}^{n} (v_i(F_i \mid x') - v_i(F_i \mid x'')) > 0 \quad \text{for every} \quad x' \neq x'' .$$

If we denote by $\mathscr{Y}_i^*$ the σ-algebra generated by the class of all $F_i(x', x'')$, $x', x'' \in X$ then $\mathscr{Y}_1^*, \mathscr{Y}_2^*, \ldots$ are finite sets and (4) holds for

$$\Delta^*(v_i(. \mid x'), v_i(. \mid x'')) = \sup_{F \in \mathscr{Y}_i^*} \{v_i(F \mid x') - v_i(F \mid x'')\} .$$

If we put in Lemma 2: $Y_i = \tilde{\mathscr{Y}}_i$, $\mathscr{Y}_i = \mathscr{Y}_i^*$ where $\tilde{\mathscr{Y}}_i \subset \mathscr{Y}_i^*$ is a disjoint decomposition of Y_i such that the σ-algebra generated by itself is $\mathscr{Y}_i^*$, then we obtain positive numbers A and $\lambda < 1$ such that $0 \leqq I_\infty^* - I_n^* < A\lambda^n$, where I_n^*, $n = 1, 2, \ldots, \infty$, is the information obtained by replacing $\mathscr{Y}_i$ by $\mathscr{Y}_i^*$, $i = 1, 2, \ldots$ Since in view of $\mathscr{Y}_i^* \subset \mathscr{Y}_i$ we have $I_n^* \leqq I_n$, $n = 1, 2, \ldots$ (cf. [2]), and since we have, according to (6), $I_\infty^* = I_\infty$, the right inequality in (5) holds for the given A and λ. To prove the left inequality we refer again to [2].

It remains to prove (7). Using the notation employed above we have, according to (11) and (12), $H(\mu) \geqq I_n^*$, $n = 1, 2, \ldots$, for every sequence $\mathscr{Y}_i^*$, $i = 1, 2, \ldots$ of finite sub-σ-algebras of $\mathscr{Y}_i$'s. Hence, by Theorem 13 in [2], the following inequality holds $H(\mu) \geqq I_n$, $n = 1, 2, \ldots$; considering the limit for $n \to \infty$ we obtain the desired result (7) and the proof of the Theorem is complete.

Let us end the paper by the evaluation of the Bayes risk r_n in the statistical decision problem we have considered under the assumption that (4) holds. An easy verification (using Lemma 1) gives $r_\infty = 0$. Using (2) we obtain $0 \leqq r_n < 2w_0(I_\infty - I_n)$ so that, in view of the Theorem, there exists $A > 0$ and $0 < \lambda < 1$ such that

$$0 \leqq r_n < A\lambda^n . \tag{20}$$

References

[1] *M. Loéve:* Probability Theory, 3-rd edition, Princeton, N. J.

[2] *A. Perez:* Notions géneralisées d'incertitude, d'entropie et d'information du point de vue de la theorie de martingales, Trans. of First Prague Conf. on Inf. Theory, Stat. Dec. Functions, Random Processes, 1957.

[3] *A. Perez:* Information theory methods in reducing complex decision problems, Trans. of Fourth Prague Conf. on Inf. Theory, Stat. Dec. Functions, Random Processes, 1965.

[4] *A. Rényi:* On the amount of information in a sample concerning a parameter, Publications of the Math. Inst. of the Hungarian Acad. Sci. (in print).

Author's address: Praha 2, Vyšehradská 49, ČSSR (Ústav teorie informace a automatizace).

Резюме

СКОРОСТЬ СХОДИМОСТИ ИНФОРМАЦИИ В ВЫБОРКЕ ОТНОСИТЕЛЬНО ПАРАМЕТРА

ИГОР ВАЙДА (Igor Vajda), Прага

В работе рассматривается средная информация I_n содержащаяся в выборке $(y_1, y_2, \ldots, y_n) \in \bigotimes_{i=1}^{n} Y_i$, $n = 1, 2, \ldots, \infty$ (где Y_i, $i = 1, 2, \ldots$ абстрактные пространства) относительно параметра x принимающего значения из конечного множества X. Показывается, что всегда имеет место неравенство (7), где $H(\mu)$ обозначает энтропию параметрового пространства X при распределении вероятностей μ. Если случайная последовательность $y_1, y_2, \ldots$ независима для каждого значения параметра x и если выполняется условие (4), то для некоторых A и λ имеет место (5), где $A > 0$ и $0 < \lambda < 1$. Этот результат представляет собой обобщение ранее полученного результата Рени [4], предполагающего конечность пространств Y_i, $i = 1, 2, \ldots$ и стационарность последовательности $y_1, y_2, \ldots$ для каждого значения параметра x.