KYBERNETIKA ČÍSLO 2, ROČNÍK 4/1968

Axiomy a-entropie zobecněného pravděpodobnostního schématu

IGOR VAJDA

V práci se předkládá nová axiomatická definice zobecněné varianty a-entropie Havrdy a Charváta.

1.

Budiž $(\Omega, \mathscr{B}, P)$ pravděpodobnostní prostor, tj. Ω budiž neprázdná množina elementárních jevů ω, $\mathscr{B}$ budiž σ-algebra podmnožin množiny Ω a P pravděpodobnostní míra na $\mathscr{B}$. Náhodnou veličinou budeme v práci rozumět $\mathscr{B}$-měřitelnou funkci $\xi = \xi(\omega)$ definovanou na Ω a nabývající konečný počet různých hodnot $x_1, x_2, \ldots, x_n$. Pravděpodobnostním schématem $\mathscr{P}$ příslušným náhodné veličině ξ budeme chápat vektor pravděpodobností $(p_1, p_2, \ldots, p_n)$, kde

$$p_i = P(\xi = x_i), \quad i = 1, 2, \ldots, n. \tag{1.1}$$

Pravděpodobnostní schéma $\mathscr{P}$ je tedy n-tice nezáporných čísel $p_1, p_2, \ldots, p_n$, kde $n = 1, 2, \ldots$, pro které

$$\sum_{i=1}^{n} p_i = 1. \tag{1.2}$$

Naopak, každému pravděpodobnostnímu schématu $\mathscr{P}$ přísluší nekonečně mnoho náhodných veličin ξ, jimž příslušné pravděpodobnostní schéma je $\mathscr{P}$.

Jestliže $\eta = \eta(\xi)$, pak střední hodnota náhodné veličiny η je definovaná vztahem

$$E\eta = \sum_{i=1}^{n} \eta(x_i)\, p_i. \tag{1.3}$$

Jestliže je zadáno pravděpodobnostní schéma $\mathscr{P}$, pak vzniká otázka, jak lze kvantitativně charakterizovat jistotu, s níž můžeme předpovědět realizaci náhodné veličiny ξ, která přísluší schématu $\mathscr{P}$. Přitom nepředpokládáme provádění žádných měření ani na x_i ani na realizaci kterékoliv jiné s ξ svázané veličiny, jestliže k realizaci x_i došlo před naším rozhodnutím. Je zřejmé, že adekvátní kvantitativní míra nesmí záviset na hodnotách $x_1, x_2, \ldots, x_n$ veličiny ξ, nýbrž jen na příslušných pravděpodobnostech $p_1, p_2, \ldots, p_n$. Vzhledem k tomu, že první úspěšnou kvantitativní mírou neurčitosti při predikci náhodných veličin byla termodynamická entropie, je všeobecně rozšířené nazývat uvažované míry entropiemi.

Při studiu minimální pravděpodobnosti chyby při testování konečně nebo spočetně mnoha hypotéz byla autorem s výhodou použita následující tzv. h-entropie (viz např. [1]), definovaná vztahem

$$h(\mathscr{P}) = 1 - E\eta\,, \tag{1.4}$$

kde náhodná veličina $\eta = \eta(\xi)$ je pro ξ příslušnou $\mathscr{P}$ definována podmínkou

$$\eta(x_i) = P(\xi = x_i)\,, \quad i = 1, 2, \ldots, n\,. \tag{1.5}$$

h-entropie je speciálním případem tzv. a-entropie, kterou zavedli J. Havrda a F. Charvát v [2]. a-entropie $h_a(\mathscr{P})$ pravděpodobnostního schématu $\mathscr{P}$ je podle [2] definovaná pro $a > 0$, $a \neq 1$ vztahem

$$h_a(\mathscr{P}) = c(a)\, E(1 - \eta^{a-1})\,, \tag{1.6}$$

kde η je dáno podmínkou (1.5) a kde $c(a)$ je normovací konstanta,

$$c(a) = \frac{2^{a-1}}{2^{a-1} - 1}\,. \tag{1.7}$$

Protože se ukázalo, že lim $h_a(\mathscr{P})$ pro $a \to 1$ existuje a je rovna*

$$H(\mathscr{P}) = E \log \frac{1}{\eta}\,, \tag{1.8}$$

což je termodynamická, resp. shannonovská entropie, bylo rozumné dodefinovat $h_a(\mathscr{P})$ pro $a = 1$ vztahem

$$h_1(\mathscr{P}) = H(\mathscr{P})\,. \tag{1.9}$$

a-entropii lze pro každé $\mathscr{P}$ postulovat následujícím způsobem:

(I) $h_a(p, 1 - p)$ je spojitou funkcí proměnné p na intervalu $0 \leqq p \leqq 1$, přičemž

$$h_a(\tfrac{1}{2}, \tfrac{1}{2}) = 1.$$

(II) $h_a(\mathscr{P})$ je symetrickou funkcí komponent schématu $\mathscr{P}$.

(III) Pro každé schéma $\mathscr{P}$ a pro každé $0 \leqq t \leqq 1$ platí

$$h_a(tp_1, (1 - t)\, p_1, p_2, \ldots, p_n) = h_a(p_1, p_2, \ldots, p_n) + p_1^a h_a(t, 1 - t)\,.$$

* $\log = \log_2$.

Důkaz tohoto tvrzení se příliš neliší od důkazu analogického tvrzení v [2] a proto ho zde uvádět nebudeme.

Všimněme si, že tyto naše postuláty obsahují jako speciální případ (při $a = 1$) známé Fadějevovy postuláty shannonovské entropie (viz [3]).

2.

Níže ukážeme jiný systém postulátů definující a-entropii na poněkud obecnější třídě pravděpodobnostních schémat $\mathscr{P}$. Instuitivní význam funkcionálů h_a přitom zůstává týž: kvantitativní charakterizace neurčitosti. Nejdříve budeme věnovat pozornost jistému zobecnění pojmu náhodné veličiny a pravděpodobnostního schématu.

$\mathscr{B}$-měřitelnou funkci $\xi = \xi(\omega)$ definovanou pro $\omega \in \Omega_1$, kde $P(\Omega_1) > 0$, a nabývající konečně mnoha různých hodnot $x_1, x_2, \ldots, x_n$, budeme nazývat zobecněná náhodná veličina. Příslušný vektor $\mathscr{P} = (p_1, p_2, \ldots, p_n)$ pravděpodobností definovaných vztahem (1.1) budeme nazývat zobecněné pravděpodobnostní schéma. Je zřejmé, že v tomto případě vztah (1.2) splněn být nemusí. Jestliže

$$W(\mathscr{P}) = \sum_{i=1}^{n} p_i < 1\,, \tag{2.1}$$

pak náhodnou veličinu ξ, resp. schéma $\mathscr{P}$ nazýváme neúplnou. $W(\mathscr{P})$ je samozřejmě vždy kladné a (2.1) platí tehdy a jen tehdy, když $P(\Omega_1) < 1$. V opačném případě mluvíme ve shodě s výše zavedenou terminologií o náhodné veličině a pravděpodobnostním schématu. Intuitivně lze neúplné náhodné veličiny interpretovat jako veličiny popisující výsledek experimentu, který závisí na náhodě a který je přístupný pozorování pouze s pravděpodobností $P(\Omega_1) > 0$.

Střední hodnotu funkce $\eta(\xi)$ zobecněné náhodné veličiny ξ definujeme vztahem (1.3), ve kterém zobecněné schéma $\mathscr{P} = (p_1, p_2, \ldots, p_n)$, příslušné veličině ξ, zaměníme pravděpodobnostním schématem $\mathscr{P}' = (p'_1, p'_2, \ldots, p'_n)$, kde

$$p'_i = \frac{p_i}{W(\mathscr{P})}\,,$$

tj.

$$E\eta = \frac{\sum_{i=1}^{n} \eta(x_i)\, p_i}{W(\mathscr{P})}\,. \tag{2.2}$$

Tato úmluva nám umožní rozšířit platnost definice (1.6) na zobecněná pravděpodobnostní schémata. Jinými slovy, pomocí (2.2) a (1.6) máme definovanou a-entropii

 $h_a(\mathscr{P})$ na třídě Δ všech zobecněných schémat $\mathscr{P}$, přičemž

$$h_a(\mathscr{P}) = c(a)\left(1 - \frac{\sum_{i=1}^{n} p_i^a}{W(\mathscr{P})}\right), \quad a > 0,\, a \neq 1\,. \tag{2.3}$$

Jak se lze snadno přesvědčit výpočtem příslušné limity, platí i v zobecněném případě, že

$$\lim_{a \to 1} h_a(\mathscr{P}) = H(\mathscr{P})\,, \tag{2.4}$$

kde $H(\mathscr{P})$ je dáno vztahem (1.8) za předpokladu, že střední hodnotu chápeme ve smyslu (2.2), tj.

$$H(\mathscr{P}) = \frac{\sum_{i=1}^{n} p_i \log \frac{1}{p_i}}{W(\mathscr{P})}\,. \tag{2.5}$$

I v zobecněném případě můžeme tedy dodefinovat

$$h_1(\mathscr{P}) = H(\mathscr{P})\,. \tag{2.6}$$

Jestli $\mathscr{P} = (p_1, p_2, \ldots, p_n) \in \Delta$, $\mathscr{Q} = (q_1, q_2, \ldots, q_n) \in \Delta$, pak ve shodě s Rényim [4] definujeme direktní součin pravděpodobnostních schémat $\mathscr{P}$ a $\mathscr{Q}$ vztahem

$$\mathscr{P} \otimes \mathscr{Q} = (p_1 q_1, p_1 q_2, \ldots, p_1 q_n, p_2 q_1, \ldots, p_n q_n)\,.$$

Zřejmě $\mathscr{P} \otimes \mathscr{Q} \in \Delta$ přičemž $W(\mathscr{P} \otimes \mathscr{Q}) = W(\mathscr{P})\, W(\mathscr{Q})$. Z této rovnosti plyne, že je-li alespoň jedno ze schémat $\mathscr{P}$, $\mathscr{Q}$ neúplné, je i $\mathscr{P} \otimes \mathscr{Q}$ neúplné. Jestliže $W(\mathscr{P}) + W(\mathscr{Q}) \leqq 1$, pak definujeme součet schémat $\mathscr{P}$ a $\mathscr{Q}$ vztahem

$$\mathscr{P} \oplus \mathscr{Q} = (p_1, p_2, \ldots, p_n, q_1, q_2, \ldots, q_n)\,.$$

Zřejmě platí $\mathscr{P} \oplus \mathscr{Q} \in \Delta$ a $W(\mathscr{P} \oplus \mathscr{Q}) = W(\mathscr{P}) + W(\mathscr{Q})$.

Nyní můžeme přistoupit k formulaci hlavního výsledku.

3.

Jestliže reálná funkce $h_a(\mathscr{P})$ definovaná na Δ splňuje pro $a > 0$ pět níže uvedených postulátů, pak platí (2.3), tj. $h_a(\mathscr{P})$ je a-entropie schématu $\mathscr{P}$.

(I) Jestliže $\mathscr{P} = (p) \in \Delta$, pak existuje

$$\lim_{p \to 0^+} \frac{h_a(1 - p)}{p} = \alpha\,.$$

(II) $h_a(1/2) \quad 1.$

(III) $h_a(\mathscr{P})$ je symetrickou funkcí všech komponent p_i, $i = 1, 2, \ldots, n$ schématu $\mathscr{P}$.

(IV) Pro každé $\mathscr{P}$, $\mathscr{Q} \in \Delta$ platí

$$h_a(\mathscr{P} \otimes \mathscr{Q}) = h_a(\mathscr{P}) + h_a(\mathscr{Q}) + (2^{1-a} - 1)\, h_a(\mathscr{P})\, h_a(\mathscr{Q})\,.$$

(V) Jestliže $\mathscr{P}$, $\mathscr{Q} \in \Delta$, $W(\mathscr{P}) + W(\mathscr{Q}) \leqq 1$, pak

$$h_a(\mathscr{P} \oplus \mathscr{Q}) = \frac{W(\mathscr{P})\, h_a(\mathscr{P}) + W(\mathscr{Q})\, h_a(\mathscr{Q})}{W(\mathscr{P}) + W(\mathscr{Q})}\,.$$

Důkaz. Nechť $\mathscr{P} = (p)$, $\mathscr{Q} = (q)$, $0 \leqq p$, $q \leqq 1$. Pak $\mathscr{P} \otimes \mathscr{Q} = (pq)$, takže podle (IV)

$$h_a(pq) = h_a(p) + h_a(q) + (2^{1-a} - 1)\, h_a(p)\, h_a(q)\,. \tag{3.1}$$

Pro $p = 1$ odtud obdržíme

$$h_a(1) = h_a(1)\,(1 - 2^{1-a})\, h_a(q)\,.$$

Protože pro $q = 1/2$ je $h_a(q) = 1$, platí pro všechna uvažovaná a podmínka $h_a(1) = 0$. Položme nyní v (3.1) $q = 1 - \Delta p/p$. Obdržíme rovnici

$$h_a(p) - h_a(p - \Delta p) = h_a\left(1 - \frac{\Delta p}{p}\right)\left[(1 - 2^{1-a})\, h_a(p) - 1\right].$$

Odtud a z (I) plyne, že pro všechna $0 < p \leqq 1$ existuje derivace $\mathrm{d}h_a(p)/\mathrm{d}p$, přičemž

$$\frac{\mathrm{d}h_a(p)}{\mathrm{d}p} = \frac{\alpha}{p}\left[(1 - 2^{1-a})\, h_a(p) - 1\right].$$

Lze se snadno přesvědčit, že řešením této diferenciální rovnice při okrajových podmínkách $h_a(1) = 0$ a $h_a(1/2) = 1$ (viz (II)) je funkce

$$h_a(p) = \frac{2^{a-1}}{2^{a-1} - 1}\,(1 - p^{a-1})\,. \tag{3.2}$$

Nechť nyní $\mathscr{P} = (p_1, p_2, \ldots, p_n) \in \Delta$ je libovolné schéma. Pak lze psát

$$\mathscr{P} = \bigoplus_{i=1}^{n} \mathscr{P}_i\,, \quad \text{kde} \quad \mathscr{P}_i = (p_i)\,, \quad i = 1, 2, \ldots, n\,. \tag{3.3}$$

Odtud a z (V) obdržíme indukcí

$$h_a(\mathscr{P}) = \frac{\sum_{i=1}^{n} W(\mathscr{P}_i)\, h_a(\mathscr{P}_i)}{\sum_{i=1}^{n} W(\mathscr{P}_i)}\,.$$

Podle (3.2), (3.3) a (1.7) můžeme tento výraz přepsat do tvaru

$$h_a(\mathscr{P}) = c(a) \frac{\sum_{i=1}^{n} p_i(1 - p_i^{a-1})}{W(\mathscr{P})} = c(a) \left(1 - \frac{\sum_{i=1}^{n} p_i^a}{W(\mathscr{P})}\right),$$

tj. (2.3) platí, což bylo dokázat.

Závěrem poznamenejme, že podmínku (I) lze zaměnit požadavkem spojitosti $h_a(p)$ pro $0 < p \leqq 1$. Za tohoto předpokladu lze k rovnosti (3.2) dospět takto: Definujme na intervalu $0 < p \leqq 1$ funkci $\Phi_a(p) = 1 + (2^{1-a} - 1)\, h_a(p)$. Ze spojitosti h_a plyne zřejmě spojitost Φ_a. Jestliže vynásobíme obě strany v (3.1) výrazem $(2^{1-a} - 1)$ a přičteme jedničku, získáme pro Φ_a funkcionální rovnici $\Phi_a(pq) = \Phi_a(p)\, \Phi_a(q)$. Z analýzy je známé, že tato funkcionální rovnice má ve třídě spojitých funkcí jediné řešení $\Phi_a(p) = p^\beta$, kde β je libovolné reálné číslo. Z (II) plyne podmínka $\Phi_a(1/2) = 2^{1-a}$, takže $\beta = a - 1$. Odtud plyne (3.2).

Pět postulátů, které získáme takovouto záměnou prvního postulátu, obsahuje jako speciální případ (při $a = 1$) Rényiho axiomy shannonovské entropie $H(\mathscr{P})$ (viz [4]).*

(Došlo dne 4. září 1967.)

LITERATURA

[1] И. Вайда: Оценки минимальной вероятности ошибки при проверке конечного или счетного числа гипотез. Проблемы передачи информации, *4* (1968), 1.

[2] J. Havrda, F. Charvát: Quantification method of classification processes. The concept of structural a-entropy. Kybernetika 3 (1967), 1, 30—35.

[3] Д. К. Фадеев: К понятию энтропии конечной вероятностной схемы. Успехи математических наук, 11 (1956), 1.

[4] A. Rényi: On measures of entropy and information. Proc. of 4-th Berkeley Symp. on Math. Stat. and Probability, Vol. I, 547—561.

* Dr. Perez si povšiml existence explicitního vztahu mezi a-entropií a Rényiho α-entropií. Ve světle tohoto vztahu je systém postulátů, které obdržíme uvažovanou záměnou prvního postulátu, ekvivalentní Rényiho axiomům α-entropie uvedeným v [4].

Axioms for a-entropy of a Generalized Probability Scheme

IGOR VAJDA

If $\mathscr{P} = (p_1, p_2, \ldots, p_n)$ is a complete or incomplete probability scheme, i.e. if $p_i \geqq 0$, $i = 1, 2, \ldots, n$, and

$$W(\mathscr{P}) = \sum_{i=1}^{n} p_i$$

is equal to 1 or $0 < W(\mathscr{P}) < 1$ respectively, then the amount of uncertainty concerning the outcome of an experiment, the observable results of which have the probabilities $p_1, p_2, \ldots, p_n$, can be measured by the so called a-entropy $h_a(\mathscr{P})$ defined by (2.3). In view of (2.4) (where $H(\mathscr{P})$ is the Shannon's entropy of the scheme $\mathscr{P}$ defined by (2.5)), we define $h_1(\mathscr{P})$ in accordance with (2.6). The a entropy of a complete scheme was first introduced by Havrda and Charvát in [2]. A special version of the a-entropy was introduced in studying an elementary statistical problem by the author of the present paper (cf. [1]).

The a-entropy of a complete probability scheme can be characterized by the following three postulates.

(I) $h_a(p, 1 - p)$ is a continuous function of p in the interval $0 \leqq p \leqq 1$ and $h_a(1/2, 1/2) = 1$.

(II) $h_a(\mathscr{P})$ is a symetric function of the elements of $\mathscr{P}$, for every complete $\mathscr{P}$.

(III) If $\mathscr{P}$ is a complete scheme and $0 \leqq t \leqq 1$, then

$$h_a(tp_1, (1 - t)\, p_1, p_2, \ldots, p_n) = h_a(p_1, p_2, \ldots, p_n) + p_1^a h_a(t, 1 - t)\,.$$

This postulates yields for $a = 1$ the Fadeev's postulates of the Shannon's entropy (cf. [3]).

If we consider the a-entropy as defined on the class Δ of all complete and incomplete probability schemes, then it can be characterized by the following five postulates.

(I) If $\mathscr{P} = (p) \in \Delta$, then

$$\lim_{p \to 0+} \frac{h_a(1-p)}{p}$$

exists.

(II) $h_a(1/2) = 1$.

(III) $h_a(\mathscr{P})$ is a symetric function of the elements of $\mathscr{P}$ for every $\mathscr{P} \in \Delta$.

(IV) If $\mathscr{P}, \mathscr{Q} \in \Delta$, then

$$h_a(\mathscr{P} \otimes \mathscr{Q}) = h_a(\mathscr{P}) + h_a(\mathscr{Q}) + (2^{1-a} - 1)\, h_a(\mathscr{P})\, h_a(\mathscr{Q})\,.$$

(V) If $\mathscr{P}, \mathscr{Q} \in \Delta$ and $W(\mathscr{P}) + W(\mathscr{Q}) \leqq 1$, then

$$h_a(\mathscr{P} \oplus \mathscr{Q}) = \frac{W(\mathscr{P})\, h_a(\mathscr{P}) + W(\mathscr{Q})\, h_a(\mathscr{Q})}{W(\mathscr{P}) + W(\mathscr{Q})}\,.$$

($\mathscr{P} \otimes \mathscr{Q}$ and $\mathscr{P} \oplus \mathscr{Q}$ are algebraic operations on Δ introduced by Rényi in [4].)

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