Aplikace matematiky

Albert Pérez

Matematická teorie informace. Část II

Aplikace matematiky, Vol. 3 (1958), No. 2, 81–105

Persistent URL: `http://dml.cz/dmlcz/102607`

ČLÁNKY

MATEMATICKÁ TEORIE INFORMACE

Část II.[1])

ALBERT PEREZ

(Došlo dne 24. června 1957.) DT: 621.39.001:51

5. Mohutnost rozlišitelnosti a jemnost pozorování. Úloha pojmu informace. Pojem kapacity kanálu

Nechť $(X_0, \mathfrak{X}_0, \mu_0)$ je zdroj informace připojený přímo na vstup kanálu $(\mathfrak{X}_0, \nu_{x_0}, \mathfrak{Y}_0)$ a nechť $(X_0 \times Y_0, \mathfrak{X}_0 \times \mathfrak{Y}_0, \omega_0)$ je příslušný dvojitý zdroj a $(Y_0, \mathfrak{Y}_0, \nu_0)$ příslušný výstupní zdroj. Budeme dále předpokládat, že $\omega_0 \ll \mu_0 \times \nu_0$ $(\mathfrak{X}_0 \times \mathfrak{Y}_0)$ a příslušnou hustotu pravděpodobnosti označíme f.

Měřitelný prostor $(X_0, \mathfrak{X}_0)$ budeme nazývat *vstupní abecedou* a měřitelný prostor $(Y_0, \mathfrak{Y}_0)$ *abecedou výstupní*. „Písmena" vstupní abecedy se pozorují kanálem jako písmena výstupní abecedy.

Nechť nyní $(\mathfrak{X}_0^n, \nu_{x^n}, \mathfrak{Y}_0^n)$ je kanál, který po písmenech přiřazuje každému „textu" x^n o n *nezávislých* písmenech vstupní abecedy „text" y^n rovněž o n *nezávislých* písmenech abecedy výstupní. Odpovídající vstupní a výstupní zdroje jsou tedy $(X_0^n, \mathfrak{X}_0^n, \mu_0^n)$, $(Y_0^n, \mathfrak{Y}_0^n, \nu_0^n)$ a dvojitý zdroj je $(X_0^n \times Y_0^n, \mathfrak{X}_0^n \times \mathfrak{Y}_0^n, \omega_0^n)$.

Necháme-li n vzrůstat bez omezení, pak nechť $(\mathfrak{X}, \nu_x, \mathfrak{Y})$ značí kanál (*stacionární* a *nezávislý*, vytvořený kanálem $(\mathfrak{X}_0, \nu_{x_0}, \mathfrak{Y}_0)$), který takto obdržíme a $(X, \mathfrak{X}, \mu)$, $(Y, \mathfrak{Y}, \nu)$ a $(X \times Y, \mathfrak{X} \times \mathfrak{Y}, \omega)$ příslušné zdroje vstupní, výstupní a dvojitý (zdroje *stacionární* a *nezávislé* s abecedami $(X_0, \mathfrak{X}_0)$, $(Y_0, \mathfrak{Y}_0)$, $(X_0 \times Y_0, \mathfrak{X}_0 \times \mathfrak{Y}_0)$). Prvky množin X resp. Y jsou „nekonečně dlouhé texty" t. j. nekonečné posloupnosti písmen z X_0 resp. Y_0 a prvky z $X \times Y$ dvojice takových textů. $\mathfrak{X}$, $\mathfrak{Y}$ a $\mathfrak{X} \times \mathfrak{Y}$ jsou σ-algebry konstruované obvyklým způsobem.

Nyní jde o to, *jak stoupá v závislosti na n mohutnost ε-rozlišitelnosti* ($\varepsilon > 0$ pevné libovolně malé) kanálu $(\mathfrak{X}_0^n, \nu_{x^n}, \mathfrak{Y}_0^n)$ nebo kanálu $(\mathfrak{X}_0^n, \nu_x, \mathfrak{Y}_0^n)$, což je

[1]) První část tohoto článku je otištěna v 1. čísle 1958 tohoto časopisu, str. 1—21.

zřejmě totéž. Jinak řečeno, jak přispívá vzrůstající rozsah pozorování (t. j. delší „přijímané texty") na vzrůst mohutnosti ε-rozlišitelnosti, kterou budeme značit $N_n(\varepsilon)$.

Nechť $f_n(x, y)$ je hustota ω vzhledem k $\mu \times \nu$ měřitelná na σ-algebře $\mathfrak{X}_0^n \times \mathfrak{Y}_0^n$. Zcela analogicky jako v důkaze formule (3.2) v § 3 lze dokázat, že *podle ω-středu*

$$\lim_{n\to\infty} \frac{1}{n} \log f_n(x, y) = R = I(\omega_0, \mathfrak{X}_0 \times \mathfrak{Y}_0) = \int \log f \, d\omega_0 \,, \tag{5.1}$$

při čemž výraz pro zobecněnou informaci zavedený v § 4 se tu objeví automaticky. Z (5.1) dostaneme jednak formuli

$$R = \lim_{n\to\infty} \frac{1}{n} \int \log f_n \, d\omega = \lim_{n\to\infty} \frac{1}{n} I(\omega, \mathfrak{X}_0^n, \mathfrak{Y}_0^n) \,, \tag{5.2}$$

která je analogická s (3.4), jednak *konvergenci posloupnosti* $\left\{\frac{1}{n} \log f_n(x, y)\right\}$ *k* R *podle ω-pravděpodobnosti* (*zobecněná vlastnost* E).

Právě na základě této poslední vlastnosti můžeme (viz [12]) dokázat existenci množiny ε-rozlišitelné kanálem $(\mathfrak{X}_0^n, \nu_{xn}, \mathfrak{Y}_0^n)$, která pro dostatečně velká n má mohutnost větší než $e^{n(R-\varepsilon)}$, kde e značí vždy použitou bázi logaritmů.

Podle definice mohutnosti ε-rozlišitelnosti jako suprema mohutností všech množin ε-rozlišitelných uvažovaným kanálem, dostáváme nerovnost

$$N_n(\varepsilon) > e^{n(C-2\varepsilon)} \,, \tag{5.3}$$

kde

$$C = \sup_{\mu} R \tag{5.4}$$

a toto supremum bereme přes všechny nezávislé stacionární zdroje $(X, \mathfrak{X}, \mu)$, které lze přímo připojit na vstup kanálu $(\mathfrak{X}, \nu_x, \mathfrak{Y})$ a pro které platí $\omega \ll \mu \times \nu$ $(\mathfrak{X}_0^n \times \mathfrak{Y}_0^n)$, $n = 1, 2, 3, \ldots$.

Číslo C nazýváme *kapacitou* stacionárního nezávislého kanálu $(\mathfrak{X}, \nu_x, \mathfrak{Y})$. Jde tu o zobecnění pojmu kapacity kanálu zaváděné obvykle pro případ konečné vstupní a výstupní abecedy (viz [1] a [12]). Jak uvidíme v následujícím paragrafu, předešlé výsledky zůstávají v platnosti za mnohem obecnějších podmínek. Nezávislost není nutná pro zajištění konvergence podle středu vyjádřené ve formulích (3.2) a (5.1) a tím spíše není nutná pro konvergenci podle pravděpodobnosti (zobecněná vlastnost E), která, jak jsme viděli, má základní důležitost ještě pro celou řadu dalších důkazů.

6. Limitní věta McMillanova a její zobecnění pro případ libovolné abecedy. Rychlost entropie a informace. Redundance

V předešlém paragrafu jsme naznačili, že konvergence podle středu ve formulích (3.2) a (5.1), které tvoří základ pro důkazy celé řady vět z teorie informace, platí za mnohem obecnějších podmínek.

Abychom tyto věty mohli formulovat (jejich důkazy viz [12] a v případě konečné abecedy také [1] a [9]), je třeba definovat *pojem stacionárního* a *ergodického zdroje.*

Nechť $(X_0, \mathfrak{X}_0)$ je libovolná abeceda a nechť

$$(X, \mathfrak{X}) = (X_0^I, \mathfrak{X}_0^I) = \mathop{\mathrm{X}}_{i=-\infty}^{\infty} (X_i, \mathfrak{X}_i), (X_i, \mathfrak{X}_i) = (X_0, \mathfrak{X}_0) ,$$
$$i = \ldots, -1, 0, 1, \ldots, \tag{6.1}$$

je měřitelný prostor nekonečných posloupností $x = (\ldots, x_{-1}, x_0, x_1, \ldots)$ „písmen" této abecedy.

Nechť T je *transformace posunutí* tohoto prostoru do sebe definovaná vztahem $T(\ldots, x_{-1}, x_0, x_1, \ldots) = (\ldots, x'_{-1}, x'_0, x'_1, \ldots)$, kde $x'_i = x_{i+1}$ pro $i = \ldots, -1, 0, 1, \ldots$.

Nechť μ je pravděpodobnostní míra v $\mathfrak{X}$. Zdroj informace $(X, \mathfrak{X}, \mu)$ nazýváme *stacionárním,* když transformace T zachovává míru každé množiny $E \in \mathfrak{X}$, t. j. $\mu(T(E)) = \mu(E)$.

Zdroj informace $(X, \mathfrak{X}, \mu)$ nazýváme *ergodickým,* když je stacionární a když pro každou množinu $E \in \mathfrak{X}$ invariantní vzhledem k transformaci T, t. j. takovou, že $T(E) = E$, platí $\mu(E) = 1$ nebo 0. Stacionární nezávislý zdroj je ergodický, jak se snadno dokáže.

Limitní věta McMillanova [1], [9] se týká zdroje informace $(X, \mathfrak{X}, \mu)$, který je stacionární, ergodický a má *konečnou* abecedu (předpokládá se, že jednobodové množiny patří do $\mathfrak{X}$).

Nechť μ_n značí pravděpodobnostní míru indukovanou pravděpodobnostní mírou μ v σ-algebře

$$\mathfrak{X}_0^n = \mathop{\mathrm{X}}_{i=0}^{n-1} \mathfrak{X}_i , \tag{6.2}$$

při čemž v dalším označení této σ-algebry zachováme a stejně jako dříve ji nerozlišujeme od σ-algebry odpovídajících válců z $\mathfrak{X}$, což není na závadu, neboť obě uvedené σ-algebry jsou isomorfní. Pak platí

limitní věta McMillanova. *Když zdroj informace* $(X, \mathfrak{X}, \mu)$ (s konečnou abecedou) *je stacionární, pak*

$$\lim_{n\to\infty} -\frac{1}{n} \log \mu_n(\{x\}) = h(x) \quad \textit{podle } \mu\textit{-středu.} \tag{6.3}$$

Když zdroj je ještě navíc ergodický, pak funkce $h(x)$ je skoro jistě $[\mu]$ rovna rychlosti entropie

$$H = \lim_{n\to\infty} -\frac{1}{n} \int_X \log \mu_n(\{x\}) \, d\mu(x) , \tag{6.4}$$

která pro stacionární zdroje vždy existuje.

Tato věta je základem současné diskretní teorie informace (s konečnou abecedou) a obsahuje t. zv. „*vlastnost E*". Speciálně na základě této věty se dokazuje t. zv. „Feinsteinovo fundamentální lemma" [1].

Obsah vlastnosti E lze vyjádřit takto: Nechť $(X, \mathfrak{X}, \mu)$ je stacionární a ergodický zdroj informace s konečnou abecedou a s rychlostí entropie H. Prostor X_0^n posloupností x^n o n písmenech lze rozložit ve dvě množiny E a $E' = X_0^n - E$, z nichž první se nazývá „*množina* (řetězů x^n) *o velké pravděpodobnosti*" a její počet bodů je mezi $e^{n(H-\varepsilon)}$ a $e^{n(H+\varepsilon)}$ a druhá se nazývá „*množina s malou pravděpodobností*", při čemž její pravděpodobnost je menší než η, kde $\varepsilon > 0$ a $\eta > 0$ mohou být libovolně malá, když n je dostatečně velké.

Tato vlastnost zdroje informace má velikou důležitost pro *kodování*. Když s je počet písmen abecedy, pak zdroj informace má maximální rychlost entropie, když je nezávislý a všechna písmena mají stejnou pravděpodobnost $\left(\text{rovnou } \frac{1}{s}\right)$. Tato maximální rychlost entropie je rovna $\log s$. Všechny řetězy x^n, jejichž počet je s^n, jsou pak stejně pravděpodobné. Když tedy $H < \log s$, pak zřejmě *kodováním* (jednoznačnou a měřitelnou transformací) *dlouhých řetězů daného zdroje na řetězy zdroje s rychlostí entropie* $\log s$ *lze dosáhnout průměrného zkrácení původního textu přibližně v poměru $H/\log s$*, jak ve speciálním případě dokázal Chinčin v [22] a kromě toho dokázal, že *většího zkrácení dosáhnout nelze.*

Možnost takového zkrácení je konec konců založena na využití *vnitřní závislosti* ve struktuře textu (*redundance, nadbytečnost*), který kodováním přejde v text se vzájemně nezávislými písmeny, t. j. s redundancí nulovou.

Redundance není vždycky zbytečná. Opatření proti rušicím vlivům, t. j. *šumu* (který představuje pozorovací kanál, jenž nemusí být vždy kanálem přímého pozorování) je právě založeno na různých způsobech využití redundance. Opakování textu je speciálním případem, avšak to nemusí být vždycky ten nejúčinnější prostředek. Účelem teorie informace je právě řešit základní problémy sdělování zpráv co nejúčinnějším způsobem.

Zastavme se nyní u zajímavého faktu, který lze dalekosáhle zobecnit. *Rychlost entropie H* definovaná pomocí (6.4), kde integrál druhého členu není nic jiného než obyčejná entropie odpovídající prostoru X_0^n řetězů x^n o n písmenech, *se shoduje se střední podmíněnou entropií jednoho písmene za předpokladu, že celý předešlý text je dán.*

Přejděme nyní k případu stacionárního zdroje $(X_0^I, \mathfrak{X}_0^I, \mu)$ s *libovolnou* (ne nutně konečnou) abecedou $(X_0, \mathfrak{X}_0)$. Model bude analogický modelu v § 3, avšak nezávislá stacionární míra μ_1 je tu nahrazena mírou μ, která je jenom stacionární (ne nutně nezávislá). Nechť μ_0 (odpovídající míře ν_1 v § 3) je pravděpodobnostní míra indukovaná pravděpodobnostní mírou μ na $\mathfrak{X}_0$ (odpovídající souřadnici x_0) a μ' pravděpodobnostní míra indukovaná mírou μ na σ-algebře $\mathfrak{X}' = \underset{i=-\infty}{\overset{-1}{\times}} \mathfrak{X}_i$ (odpovídající souřadnicím $\ldots, x_{-2}, x_{-1}$).

V dalším budeme vždycky předpokládat, že $\mu \ll \mu' \times \mu_0\,(\bar{\mathfrak{X}})$, *kde* $\bar{\mathfrak{X}} = \mathfrak{X}' \times \mathfrak{X}_0$ (odpovídající souřadnicím $\ldots, x_{-2}, x_{-1}, x_0$).

Poznamenejme, že tento předpoklad je automaticky splněn v případě konečné abecedy. V obecném případě je tento předpoklad splněn za vhodných obecnějších podmínek (viz [12]). Budeme tento předpoklad krátce nazývat *vlastností F* uvažovaného zdroje informace.

Nechť w_0 (nahrazující ν_2 v § 3) je míra na $\mathfrak{X}_0$ *dominující* μ_0, t. j. $\mu_0 \ll w_0\,(\mathfrak{X}_0)$. Nechť w je *nezávislá* míra indukovaná mírou w_0 definovaná na $\mathfrak{X}_0^I$ (zrovna tak jako μ_2 byla indukována mírou ν_2 v § 3). Pak $\mu \ll w\,(\mathfrak{X}_0^n)$ pro $n = 1, 2, 3, \ldots$. Když $f_n(x)$ je příslušná hustota (nahrazující $f_n(y)$ v § 3), pak

$$H_w(\mu, \mathfrak{X}_0^n) = -\int \log f_n \,\mathrm{d}\mu, \quad n = 1, 2, \ldots \tag{6.5}$$

Když existuje limita

$$\lim_{n\to\infty} \frac{1}{n} H_w(\mu, \mathfrak{X}_0^n) = H_w\,, \tag{6.6}$$

pak ji nazýváme (*zobecněnou*) *rychlostí entropie* (stacionárního) zdroje informace $(X_0^I, \mathfrak{X}_0^I, \mu)$ vzhledem k w.

Věta 6.1. *Nechť* $(X_0^I, \mathfrak{X}_0^I, \mu)$ *je stacionární zdroj informace s libovolnou abecedou a nechť*

1) $\mu \ll \mu' \times \mu_0\,(\bar{\mathfrak{X}})$, *kde* $\bar{\mathfrak{X}} = \mathfrak{X}' \times \mathfrak{X}_0$,

2) $\mu_0 \ll w_0$, *tedy* $\mu \ll w\,(\mathfrak{X}_0^n)$ *pro* $n = 1, 2, 3, \ldots$.

Pak existuje rychlost entropie (6.6) *a platí*

$$H_w = \lim_{n\to\infty} \frac{1}{n} H_w(\mu, \mathfrak{X}_0^n) = H_{\mu' \times w_0}(\mu, \bar{\mathfrak{X}})\,. \tag{6.7}$$

Zobecněná entropie $H_{\mu' \times w_0}(\mu, \bar{\mathfrak{X}})$ není nic jiného než *střední podmíněná entropie jednoho písmene za předpokladu, že známe celý předešlý text* (vzhledem k dominující míře w_0) odpovídající uvažovanému zdroji informace. Tento výsledek je tedy zobecněním analogického výsledku pro případ konečné abecedy. Přechod od zobecněné rychlosti entropie k obyčejné rychlosti entropie je přesně stejný jako přechod od zobecněné entropie k obyčejné entropii (viz § 3). Jde o vhodnou volbu dominující míry w.

Je vhodné upozornit na analogii mezi formulemi (3.4), (6.4) a (6.7), z nichž poslední obsahuje obě předcházející jako zvláštní případy. Naskýtá se zcela přirozeně otázka, jestli snad také formule (3.2) a (6.3) nejsou speciální případy obecnější formule, totiž $\lim\limits_{n\to\infty} -\frac{1}{n} \log f_n = H_w$ podle μ-středu. Následující věta udává na tuto otázku kladnou odpověď a nazýváme ji proto *zobecněnou versí limitní věty McMillanovy*.

Věta 6.2. *Nechť* $(X_0^I, \mathfrak{X}_0^I, \mu)$ *je stacionární zdroj informace s libovolnou abecedou a předpokládejme, že*

1) $\mu \ll \mu' \times \mu_0\ (\bar{\mathfrak{X}})$, *kde* $\bar{\mathfrak{X}} = \mathfrak{X}' \times \mathfrak{X}_0$,

2) $\mu_0 \ll w_0$ *tedy* $\mu \ll w\ (\mathfrak{X}_0^n)$ *pro* $n = 1, 2, 3, \ldots$, *a* f_n *je příslušná hustota,*

3) *rychlost entropie* (6.7) (která podle věty 6.1 vždy existuje) *je konečná.*

Pak platí

$$\lim_{n\to\infty} \frac{1}{n} \log f_n(x) = h(x) \quad \textit{podle } \mu\textit{-středu}, \tag{6.8}$$

a odtud plyne

$$\lim_{n\to\infty} \frac{1}{n} \int \log f_n \,\mathrm{d}\mu = \int h \,\mathrm{d}\mu = -H_w\,. \tag{6.9}$$

Když zdroj je ještě navíc ergodický, pak $h(x) = -H_w$ *skoro jistě* $[\mu]$, *takže*

$$\lim_{n\to\infty} \frac{1}{n} \log f_n(x) = -H_w \quad \textit{podle } \mu\textit{-středu}. \tag{6.10}$$

Důkazy nalezne čtenář v [12].

Právě uvedená věta má pro obecnou teorii informace (s libovolnou abecedou) stejnou základní důležitost jako její speciální případ, t. j. limitní věta McMillanova, pro diskretní teorii informace (s konečnou abecedou).

Následující věta je jistým přizpůsobením věty 6.2 podmínkám na dvojitý zdroj $(X \times Y,\ \mathfrak{X} \times \mathfrak{Y},\ \omega)$ indukovaný stacionárním zdrojem $(X, \mathfrak{X}, \mu)$ a *stacionárním kanálem* $(\mathfrak{X}, \nu_x, \mathfrak{Y})$, t. j. takovým kanálem, že pro každé $x \in X$ a každou $F \in \mathfrak{Y}$ platí

$$\nu_{T(x)}(T(F)) = \nu_x(F)\,,$$

kde T je transformace posunutí (která byla definovaná na začátku tohoto paragrafu), použitá na prostory X resp. Y. Snadno se dokáže (viz [12]), že *příslušný dvojitý zdroj je také stacionární právě tak jako výstupní zdroj* $(Y, \mathfrak{Y}, \nu)$. Účelem této věty je zobecnit vztahy (5.1) a (5.2) předešlého paragrafu.

Věta 6.3. *Nechť* $\omega \ll \mu \times \nu\ (\mathfrak{X}_0^n \times \mathfrak{Y}_0^n)$ *pro* $n = 1, 2, 3, \ldots$, *a nechť* $f_n(x, y)$ *je příslušná hustota. Když dvojitý zdroj má vlastnost* F, *pak limita*

$$R = \lim_{n\to\infty} \frac{1}{n} \int \log f_n \,\mathrm{d}\omega = \lim_{n\to\infty} \frac{1}{n} I(\omega, \mathfrak{X}_0^n \times \mathfrak{Y}_0^n)\,, \tag{6.12}$$

která se nazývá zobecněnou rychlostí informace, vždy existuje. Když $R < \infty$, *pak*

$$\lim_{n\to\infty} \frac{1}{n} \log f_n(x, y) = h(x, y) \quad \textit{podle } \omega\textit{-středu} \tag{6.13}$$

a když dvojitý zdroj je ještě navíc ergodický, pak

$$\lim_{n\to\infty} \frac{1}{n} \log f_n(x, y) = R \quad \textit{podle } \omega\textit{-středu}. \tag{6.14}$$

7. Zobecněný tvar základního Feinsteinova lemmatu. Pojem regulárního kanálu a kapacity

Nechť $(\mathfrak{X}, \nu_x, \mathfrak{Y})$ je kanál a $\{\mathfrak{Y}_n\}$ *stoupající* posloupnost pod-σ-algeber výstupní σ-algebry $\mathfrak{Y}$. Posloupnost mohutností ε-rozlišitelnosti odpovídající posloupnosti kanálů $(\mathfrak{X}, \nu_x, \mathfrak{Y}_n)$ je pak *neklesající.*

Otázka co se stane z hlediska rozlišitelnosti, když uvažujeme *stoupající* posloupnost $\{\mathfrak{X}_n\}$ pod-σ-algeber σ-algebry $\mathfrak{X}$ vstupu kanálu $(\mathfrak{X}, \nu_x, \mathfrak{Y})$ má obecně smysl, když na vstup kanálu je přímo připojen zdroj informace $(X, \mathfrak{X}, \mu)$. Nyní má smysl mluvit o podmíněné pravděpodobnosti p_{nx} na $\mathfrak{Y}_n$ pro dané x. Tato podmíněná pravděpodobnost jako funkce x je měřitelná vzhledem k σ-algebře $\mathfrak{X}_n$ pro $n = 1, 2, 3, \ldots$. Když tato podmíněná pravděpodobnost je pravděpodobnostní míra, pak má smysl mluvit o posloupnosti kanálů $(\mathfrak{X}_n, p_{nx}, \mathfrak{Y}_n)$, která, jak vidíme, obecně závisí na zdroji informace připojeném přímo na vstup původního kanálu, t. j. *tato posloupnost kanálů obecně závisí na pravděpodobnostní míře* μ. Toliko ve výjimečných případech tato posloupnost nezávisí na μ, nýbrž jenom na původním kanálu. Takový případ je v § 5, kde kanál $(\mathfrak{X}, \nu_x, \mathfrak{Y})$ je stacionární a nezávislý a posloupnost $\{(\mathfrak{X}_n, p_{nx}, \mathfrak{Y}_n)\}$ se shoduje s posloupností kanálů $(\mathfrak{X}_0^n, \nu_x, \mathfrak{Y}_0^n)$.

Obecně kanál $(\mathfrak{X}, \nu_x, \mathfrak{Y})$ budeme nazývat *regulárním* vzhledem ke stoupajícím posloupnostem pod-σ-algeber $\{\mathfrak{X}_n\}$ a $\{\mathfrak{Y}_n\}$ σ-algeber $\mathfrak{X}$ resp. $\mathfrak{Y}$, když pro každý zdroj informace připojený přímo na vstup kanálu posloupnost kanálů $(\mathfrak{X}_n, p_{nx}, \mathfrak{Y}_n)$ se shoduje s posloupností kanálů $(\mathfrak{X}_n, \nu_x, \mathfrak{Y}_n)$ aspoň pro jednu versi podmíněné pravděpodobnosti p_{nx}, jinak řečeno, když $\nu_x(F_n)$ jako funkce x je měřitelná vzhledem k $\mathfrak{X}_n$ pro každou množinu $F_n \in \mathfrak{Y}_n$, $n = 1, 2, 3, \ldots$.

Regulární kanál jsme zavedli proto, abychom mohli rozšířit metodu použitou v § 5 na stacionární nezávislé kanály pro odhad mohutnosti rozlišitelnosti na obecnější případ kanálů stacionárních, avšak ne nutně nezávislých. Bez použití pojmu regularity kanálu nevíme jakým způsobem zavést vhodně kapacitu kanálu, která by kanál charakterisovala nezávisle na zdrojích připojovaných na jeho vstup.

Nechť $\mathfrak{E}$ je množina všech uvažovaných stacionárních a ergodických zdrojů informace $(X, \mathfrak{X}, \mu)$, které po přímém připojení na vstup stacionárního kanálu $(\mathfrak{X}, \nu_x, \mathfrak{Y})$, kde $\mathfrak{X} = \mathfrak{X}_0^I$, $\mathfrak{Y} = \mathfrak{Y}_0^I$, indukují dvojitý zdroj $(X \times Y, \mathfrak{X} \times \mathfrak{Y}, \omega)$, který je nejen stacionární (jak jsme dříve už zjistili) *nýbrž i ergodický, takže platí* (6.14) (viz větu 6.3).

Předpokládáme-li, že kanál $(\mathfrak{X}, \nu_x, \mathfrak{Y})$ je *regulární* vzhledem k posloupnostem vstupních a výstupních σ-algeber $\{\mathfrak{X}_0^n\}$ a $\{\mathfrak{Y}_0^n\}$, pak *kapacitu* kanálu C_1 definujeme jako

$$C_1 = \sup_{\mathfrak{E}} R \,, \tag{7.1}$$

kde supremum bereme přes všechny rychlosti informace R (viz (6.12)), určené daným kanálem a zdroji z množiny $\mathfrak{C}$.

Následující věta je zobecněním výsledků § 5 a obsahuje jako zvláštní případ *Feinsteinovo fundamentální lemma* pro konečné abecedy (viz [1] a [10]).

Věta 7.1. *Když stacionární kanál* $(\mathfrak{X}_0^I, \nu_x, \mathfrak{Y}_0^I)$ *je regulární vzhledem k posloupnostem σ-algeber* $\{\mathfrak{X}_{(n+m)}\}$ *a* $\{\mathfrak{Y}_0^n\}$, *kde první je určena souřadnicemi* $x_{-m}, \ldots, \ldots, x_0, \ldots, x_{n-1}$ *(m je pevné přirozené číslo nazývané „pamětí kanálu") a druhá jako obvykle souřadnicemi* $y_0, \ldots, y_{n-1}$, *pak pro každé* $0 < \varepsilon < \frac{1}{2}$ *libovolně malé existuje* $n_0(\varepsilon)$ *tak, že pro* $n > n_0(\varepsilon)$ *mohutnost* $N_n(\varepsilon)$ *ε-rozlišitelnosti kanálem* $(\mathfrak{X}_{(n+m)}, \nu_x, \mathfrak{Y}_0^n)$ *splňuje nerovnost*

$$N_n(\varepsilon) > e^{n(C_1 - 2\varepsilon)} . \tag{7.2}$$

Důkaz této věty, který je v [12], se provádí na základě vztahu (6.14) resp. na základě konvergence podle pravděpodobnosti, která z něho plyne (*zobecněná vlastnost E*). Platnost vztahu (6.14) je zajištěna (viz definici kapacity C_1) pro všechny zdroje informace z $\mathfrak{C}$ připojované přímo na vstup uvažovaného kanálu. *Když tedy R je rychlost informace odpovídající zdroji z* $\mathfrak{C}$, *pak lze ukázat* (přesně tak jako v § 5), *že existuje množina ε-rozlišitelná kanálem* $(\mathfrak{X}_{(n+m)}, \nu_x, \mathfrak{Y}_0^n)$, *která pro dostatečně velké n má mohutnost větší než* $e^{n(R-\varepsilon)}$.

Předešlá věta právě tak jako pomocné výsledky, které jsme zdůraznili, mají velkou důležitost pro teorii přenosu zpráv.

8. O pravděpodobnostní koncepci pojmu rozlišitelnosti (pokračování): problém možnosti přenosu zdroje informace sdělovacím kanálem. Kodování

V tomto paragrafu se dále rozvíjí pravděpodobnostní koncepce rozlišitelnosti, která byla zavedena v § 2. Týká se problému možnosti přenosu zdroje informace sdělovacím kanálem.

Sdělovací kanál musí obecně být odlišitelný od kanálu příjemce zpráv nejenom tím, že před sdělovací kanál je obvykle zařazeno *kodování* (měřitelná transformace měřitelného prostoru zdroje do měřitelného prostoru vstupu do kanálu), nýbrž také tím, že nejčastěji faktory nezávislé na příjemci zpráv a jeho ideálních možnostech mají tendenci udělat sdělovací kanál méně výhodný než kanál příjemce zpráv.

V dalším budeme předpokládat, že sdělovací kanál je vytvořen sdělovacím kanálem v pravém slova smyslu s kanálem příjemce zpráv v kaskádě. Za těchto okolností by bylo ilusorní žádat, aby sdělovací kanál zajistil výhodnější situaci pro pozorování zdroje informace než kanál příjemce zpráv. Bylo by nejen mnohem nákladnější, nýbrž také zcela neužitečné volit sdělovací kanál v pravém slova smyslu, který by měl příliš velikou rozlišovací schopnost.

Abychom mohli odpovědět na otázku, jaká je minimální mohutnost rozlišitelnosti, kterou musí mít sdělovací kanál, aby jím bylo možno „přenést" daný zdroj informace, je třeba napřed rozřešit *problém charakterisace zdroje informace s hlediska jeho přenosu*, při čemž na jedné straně je třeba vzít v úvahu příjemce zpráv a na druhé straně ekonomii přenosu.

Podle naší koncepce jde o charakterisaci zdroje informace $(X, \mathfrak{X}, \mu)$ pomocí množiny ε-rozlišitelné kanálem $(\mathfrak{X}, \nu_x, \mathfrak{Y})$ příjemce zpráv, tak aby její mohutnost za daných pozorovacích podmínek byla co nejmenší (ekonomické úvahy), avšak aby byla ve shodě s požadavky příjemce. Příjemce zpráv totiž nemůže připustit aby uvažovaná množina měla „mezery" v tom smyslu, že by jí bylo možno podstatně zvětšit, aniž by se porušila ε-rozlišitelnost.

Pozorovacími podmínkami rozumíme kromě kanálu příjemce zpráv také předpokládaný stupeň rozlišitelnosti (ε-rozlišitelnost) a po případě také horní hranice η pravděpodobnosti $\mu(E)$ množiny $E \in \mathfrak{X}$, kterou by příjemce eventuelně mohl zaměnit za množinu jednobodovou aniž by se snažil rozlišit body množiny E, když to ekonomické úvahy vyžadují. Takovou množinu nazýváme *výjimečnou*. Splňuje nerovnost $\mu(E) < \eta$ a může být také prázdná.

Zavedení pojmu výjimečné množiny dává smysl pojmu *ε-rozlišitelnosti vzhledem k podmnožině uvažovaného prostoru*, obsahujícího toliko prvky této podmnožiny.

Množinu ε-rozlišitelnou daným kanálem nazýváme *maximální*, když její mohutnost je rovna mohutnosti podmnožiny vzhledem k níž je definována, nebo když její rozšíření o libovolnou množinu dalších bodů právě uvedené množiny nemůže dát ε-rozlišitelnou množinu větší mohutnosti.

Je samozřejmé, že v případě, že supremum mohutností všech ε-rozlišitelných množin definovaných vzhledem k dané množině je konečné, pak tato množina obsahuje maximální ε-rozlišitelnou množinu. Protože ještě nevíme, je-li toto tvrzení správné ve všech případech, omezíme se v dalším toliko na výjimečné množiny, jejichž komplementy obsahují maximální ε-rozlišitelnou množinu.

Při daných podmínkách pozorování uvažujme pro každou výjimečnou množinu E míry ne větší než η maximální ε-rozlišitelnou množinu (definovanou vzhledem k $X - E$) o minimální mohutnosti a mezi nimi zvolme jednu, jejíž mohutnost je minimální vzhledem k výjimečné množině E. Tuto množinu (rozšířenou eventuelně o bod výjimečné množiny, když to neporuší její ε-rozlišitelnost) nazýváme *charakteristickou množinou* zdroje informace $(X, \mathfrak{X}, \mu)$ za daných podmínek pozorování a její mohutnost nazýváme *charakteristickou mohutností*.

Jak jsme viděli v § 2 můžeme z hlediska pozorovatele (příjemce) zaměnit charakteristickou množinu množinou odpovídající systému náhodných jevů E_γ definovaných pomocí (2.3), což v dalším budeme provádět.

Podle definice, *zdroj informace lze přenést v užším smyslu* (v širším smyslu viz § 10) sdělovacím kanálem, když po vhodném kodování jeho charakteristická množina, rozlišitelná kanálem příjemce zpráv s daným stupněm rozlišitelnosti, zůstane rozlišitelná po průchodu uvažovaným kanálem se stejným stupněm rozlišitelnosti. (Pokud jde o další podrobnosti týkající se kodování odkazujeme čtenáře na [12]).

Toto hledisko má dostatečně universální charakter a umožňuje nám, jak v dalším uvidíme, uvažovat s jednotného hlediska jak diskretní, tak také nediskretní případ. To bylo umožněno zvláště zavedením pojmu kanálu příjemce zpráv.

Následující věta je skoro samozřejmá:

Věta 8.1. *Když pro daný stupeň rozlišitelnosti charakteristická mohutnost zdroje informace je menší než mohutnost rozlišitelnosti sdělovacího kanálu, pak zdroj lze přenést tímto kanálem. Když naproti tomu charakteristická mohutnost je větší než mohutnost rozlišitelnosti kanálu, pak zdroj informace nelze tímto kanálem přenést.*

9. Odhad charakteristické mohutnosti ve stacionárním případě

Je zvláště důležité mít k disposici co nejlepší odhad charakteristické mohutnosti zdroje informace, abychom mohli v co největší míře uplatnit ekonomické hledisko při stanovení nutné mohutnosti rozlišitelnosti pro sdělovací kanál, jak plyne z věty 8.1.

Následující věta se opírá o větu 6.3.

Věta 9.1. *Nechť* $(X_0^I, \mathfrak{X}_0^I, \mu)$ *je stacionární a ergodický zdroj informace, který při pozorování kanálem příjemce zpráv* $(\mathfrak{X}_0^I, \nu_x, \mathfrak{Y}_0^I)$, *o němž předpokládáme, že je stacionární, indukuje dvojitý zdroj, který je ergodický (má zobecněnou vlastnost E). Nechť* $R < \infty$ *je odpovídající rychlost informace* (viz (6.12)). *Dále nechť* $0 < \varepsilon < \frac{1}{2}$ *(kde* ε *pevně určuje stupeň rozlišitelnosti) a* $\eta > 0$ *(kde* η *shora omezuje pravděpodobnost výjimečné množiny). Pak existuje* $n_0(\varepsilon, \eta)$ *tak, že pro* $n > n_0(\varepsilon, \eta)$ *charakteristická mohutnost* $N_n(\varepsilon, \eta)$ *zdroje* $(X_0^n, \mathfrak{X}_0^n, \mu)$ *za podmínek pozorování daných kanálem* $(\mathfrak{X}_0^n, p_{nx}, \mathfrak{Y}_0^n)$ *a čísly* ε *a* η *splňuje nerovnost*

$$e^{n(R-2\varepsilon)} < N_n(\varepsilon, \eta) < e^{n(R+2\varepsilon)}, \tag{9.1}$$

při čemž první z nerovností platí pro $0 < \eta < 1$.

Poznamenejme, že právě uvedená věta dokázaná v [12], která je jistým tvarem zobecněné vlastnosti E, obsahuje, jak se dalo očekávat, vlastnost E, jak o ní byla zmínka v § 6 v souvislosti s limitní větou McMillanovou pro stacionární a ergodický zdroj s konečnou abecedou a rychlostí entropie H. Abychom ukázali, že tomu tak skutečně je, stačí ukázat, že množina bodů x^n

z „množiny o velké pravděpodobnosti" zvětšená o bod z „množiny o malé pravděpodobnosti", která je v tomto případě výjimečnou množinou, tvoří charakteristickou množinu, když jako kanál příjemce zpráv uvažujeme kanál přímého pozorování každého z měřitelných prostorů $(X_0^n, \mathfrak{X}_0^n)$ pro dostatečně velké n.

Skutečně za těchto podmínek jednak R se shoduje s H, jednak když s značí počet „písmen" abecedy X_0, procento z celkového počtu s^n bodů prostoru X_0^n patřících do výjimečné množiny („množiny o malé pravděpodobnosti") konverguje k 100%, když $n \to \infty$ pro $H < \log s$. Toto procento je totiž nejméně rovno $100\,(1 - \mathrm{e}^{\,n(\log s - H - \varepsilon)})$ a konverguje k 100%, když $n \to \infty$. Když na druhé straně H nabyde svého maxima, t. j. $\log s$, pak všechny body x_n jsou stejně pravděpodobné, takže volba výjimečné množiny pro určení charakteristické množiny je úplně indiferentní, stačí, aby její míra byla menší než pevně dané η.

Korolár 9.1. *V případě stacionárního a ergodického zdroje s konečnou abecedou a s rychlostí entropie H (tato rychlost entropie se shoduje s rychlostí informace zdroje vzhledem ke kanálu přímého pozorování) mohutnost $N_n(\eta)$ „množiny o velké pravděpodobnosti" (která rozšířena případně o bod „množiny o malé pravděpodobnosti" je asymptoticky charakteristickou množinou) splňuje nerovnosti*

$$\mathrm{e}^{n(H-\varepsilon)} < N_n(\eta) < \mathrm{e}^{n(H+\varepsilon)} \tag{9.2}$$

pro každé $\varepsilon > 0$ a $0 < \eta < 1$ (při čemž η shora omezuje pravděpodobnost výjimečné množiny, která je zde „množinou o malé pravděpodobnosti") pro dostatečně velká n. Z (9.2) *plyne*

$$\lim_{n\to\infty} \frac{1}{n} \log N_n(\eta) = H \tag{9.3}$$

pro každé $0 < \eta < 1$ (viz větu 3 v [22]).

Jak uvidíme v následujícím paragrafu, pro otázky přenosu (v širším smyslu) zdroje informace mají věty 7.1 a 8.1 spolu s větou 9.1 a korolárem 9.1 (pro případ konečné abecedy) základní důležitost pro uvažovaný stacionární případ. Speciálně věta 9.1 nám umožňuje *uspořádat množinu uvažovaných zdrojů informace podle rychlosti informace s hlediska možnosti přenosu regulárním sdělovacím kanálem.*

10. Možnost přenosu v širším smyslu zdroje informace sdělovacím kanálem. Stacionární případ. Kapacity C_1 a C_2

Definici možnosti přenosu zavedenou v § 8 lze zobecnit a přizpůsobit aplikacím, kde se často spokojujeme s aproximací toho, co nelze dosáhnout přesně nebo to aspoň není pohodlné. Tak na př. podle věty 7.1 čím větší je n tím větší je příslušná mohutnost ε-rozlišitelnosti splňující (7.2). Z tohoto hlediska je ideální pracovat s nekonečně dlouhými „texty“ ($n = \infty$), avšak prakticky je možné pracovat s „texty“ nebo „řetězy“ toliko s konečným počtem „písmen“ (n konečné). S ohledem na tuto skutečnost vyjádřenou obecněji pomocí rostoucích posloupností σ-algeber a když si ještě uvědomíme, že cílem teorie přenosu zpráv je vytvořit prostředky, které máme k disposici, co nejvýhodnější podmínky pro příjemce zpráv tak, aby tento mohl co nejlépe odhadovat vysílané zprávy na základě zpráv přijatých, pak dospějeme k následující definici možnosti přenosu.

Nechť $(X, \mathfrak{X}, \mu)$ je zdroj informace, $(\mathfrak{X}, \nu_x, \mathfrak{Y})$ kanál příjemce zpráv a $(\mathfrak{Z}, \nu_z, \mathfrak{Y})$ sdělovací kanál.

Nechť $\{\mathfrak{X}_n\}$ je stoupající posloupnost pod-σ-algeber σ-algebry $\mathfrak{X}$ a $\{\mathfrak{Y}_n\}$ a $\{\mathfrak{Z}_n\}$ analogické posloupnosti vzhledem k $\mathfrak{Y}$ a $\mathfrak{Z}$. Podle definice, *zdroj informace* $(X, \mathfrak{X}, \mu)$ *lze přenést v širším smyslu kanálem* $(\mathfrak{Z}, \nu_z, \mathfrak{Y})$ vzhledem k posloupnostem σ-algeber $\{\mathfrak{X}_n\}$, $\{\mathfrak{Y}_n\}$ a $\{\mathfrak{Z}_n\}$ a za limitních podmínek pozorování $\varepsilon_0 \geqq 0$, $\eta_0 \geqq 0$ a $(\mathfrak{X}, \nu_x, \mathfrak{Y})$ (kanál příjemce zpráv), když

a) kanál $(\mathfrak{Z}, \nu_z, \mathfrak{Y})$ je *regulární* vzhledem k $\{\mathfrak{Z}_n\}$ a $\{\mathfrak{Y}_n\}$,

b) ke každému $\varepsilon > \varepsilon_0$ a $\eta > \eta_0$ existuje $n_0(\varepsilon, \eta)$ tak, že pro $n > n_0$ zdroj $(X, \mathfrak{X}_n, \mu)$ lze přenést kanálem $(\mathfrak{Z}, \nu_z, \mathfrak{Y})$ v užším smyslu podle definice v § 8, při čemž podmínky pozorování jsou ε, η a $(\mathfrak{X}_n, p_{nx}, \mathfrak{Y}_n)$ (p_{nx} je podmíněná pravděpodobnost která pro každé $x \in X$ je podle předpokladu pravděpodobnostní míra na $\mathfrak{Y}_n$ a jako funkce x je měřitelná vzhledem k $\mathfrak{X}_n$ a odpovídá dvojitému zdroji indukovanému zdrojem informace a kanálem příjemce zpráv).

Tato definice možnosti přenosu obsahuje jako speciální případ definici možnosti přenosu v užším smyslu jakož i následující definici uváděnou v diskretní teorii informace (v dalším *definice D*): zdroj informace $(X_0^I, \mathfrak{X}_0^I, \mu)$ s konečnou abecedou je schopný přenosu regulárním kanálem $(\mathfrak{Z}_0^I, \nu_z, \mathfrak{Y}_0^I)$ vzhledem k posloupnostem vstupních a výstupních σ-algeber $\{\mathfrak{Z}_0^n\}$ a $\{\mathfrak{Y}_0^n\}$, když ke každému $\varepsilon > 0$ existuje $n_0(\varepsilon)$ tak, že pro $n > n_0$ a po vhodném kodování před vstupem zdroje $(X_0^n, \mathfrak{X}_0^n, \mu)$ do kanálu $(\mathfrak{Z}_0^n, \nu_z, \mathfrak{Y}_0^n)$ je *pravděpodobnost chyby* e_n menší než ε.

Nyní budeme definovat pravděpodobnost chyby. Nechť $(X, \mathfrak{X}, \mu)$ je zdroj informace připojený přímo na vstup kanálu $(\mathfrak{X}, \nu_x, \mathfrak{Y})$ a předpokládejme, že prostor X je konečný, $X = \{x_1, x_2, \ldots, x_n\}$ a σ-algebra $\mathfrak{X}$ obsahuje všechny jednobodové množiny prostoru X.

Nechť $\boldsymbol{P} = \{G_1, G_2, \ldots, G_n\}$ je *rozklad* měřitelného prostoru $(Y, \mathfrak{Y})$, t. j. systém množin z $\mathfrak{Y}$ po dvou disjunktních jejichž sjednocení je Y. *Pravděpodobnost chyby* e, kterou uděláme když rozhodneme, že nastal jev x_i za předpokladu, že $y \in G_i$ je dána výrazem

$$e = 1 - \sum_{i=1}^{n} \mu(\{x_i\})\, \nu_{x_i}(G_i)\,. \tag{10.1}$$

Jak víme (viz na př. [25]), existuje vždy rozklad $\boldsymbol{P}$, který minimalisuje pravděpodobnost chyby (10.1). Tento rozklad a tedy i minimum pravděpodobnosti chyby závisí na „apriorní míře“ μ.

Věta 10.1. *Když konečný prostor X, jehož body odhadujeme, tvoří množinu ε-rozlišitelnou kanálem $(\mathfrak{X}, \nu_x, \mathfrak{Y})$ pak nezávisle na „apriorní míře“ minimum pravděpodobnosti chyby bude menší než ε.*

Kromě této věty jsou ještě jiné důležité vztahy mezi pojmem ε-rozlišitelnosti a minimem pravděpodobnosti chyby (viz [12]). Omezíme se tu na poznámku, že na základě těchto vztahů lze dokázat následující větu:

Věta 10.2. *Definice D možnosti přenosu (používaná v diskretní teorii informace) je speciálním případem definice možnosti přenosu v širším smyslu, kde kanál příjemce zpráv je kanálem přímého pozorování, $\varepsilon_0 = \eta_0 = 0$ a abeceda je konečná.*

Vidíme, že zavedení přirozeného pojmu kanálu příjemce zpráv nám umožnilo uvažovat *současně* případ konečné i libovolné abecedy. Účelnost tohoto hlediska ukážeme na následující větě, která jako speciální případ obsahuje t. zv. *„první větu Shannonovu“*, jak ji nazval Chinčin v [1].

Věta 10.3. *Nechť $(X_0^I, \mathfrak{X}_0^I, \mu)$ je stacionární a ergodický zdroj informace s libovolnou abecedou. Nechť $(\mathfrak{X}_0^I, \nu_x, \mathfrak{Y}_0^I)$ je kanál příjemce zpráv o němž předpokládáme, že je stacionární a který spolu se zdrojem informace indukuje dvojitý (ergodický) zdroj se zobecněnou vlastností E. Nechť R je příslušná rychlost informace* (6.14). *Dále nechť $(\mathfrak{Z}_0^I, \nu_z, \mathfrak{Y}_0^I)$ je stacionární sdělovací kanál regulární vzhledem k posloupnosti σ-algeber $\{\mathfrak{Z}_0^n\}$ (resp. $\{\mathfrak{Z}_{(n+m)}\}$,* viz větu 7.1) *a $\{\mathfrak{Y}_0^n\}$ a s kapacitou C_1 definovanou pomocí* (7.1). *Pak zdroj informace lze přenést sdělovacím kanálem vzhledem k posloupnostem σ-algeber $\{\mathfrak{X}_0^n\}$, $\{\mathfrak{Y}_0^n\}$ a $\{\mathfrak{Z}_0^n\}$ (resp. $\{\mathfrak{Z}_{(n+m)}\}$), když*

$$R < C_1 , \tag{10.2}$$

při čemž limitní pozorovací podmínky jsou určeny pomocí $\varepsilon_0 = \eta_0 = 0$ a kanálem příjemce zpráv.

Tato věta obsahuje jako speciální případ t. zv. první větu Shannonovu, která říká, že za předpokladu konečné abecedy a když kanál příjemce zpráv je kanálem přímého pozorování, pak zdroj informace lze přenést v předešlém smyslu (nebo, což je podle věty 10.2 ekvivalentní, ve smyslu definice D), když

$$H < C_1 , \tag{10.3}$$

kde H je rychlost entropie zdroje.

Věta 10.3 je dokázána v [12] na základě vět 7.1, 8.1 a 9.1.

Bohužel kapacita C_1, definovaná pomocí (7.1), nám neumožňuje tvrdit opak, t. j. když $R > C_1$ (v případě konečné abecedy $H > C_1$), pak zdroj nelze přenést.

Avšak když se nezabýváme tím, jak obtížně lze kapacitu kanálu vypočítat, pak nám nic nebrání v tom, abychom definovali jinou *kapacitu* C_2 jako supremum rychlostí informace R všech stacionárních a ergodických zdrojů (patřících do tak velké množiny jakou právě potřebujeme), které lze přenést daným sdělovacím kanálem jako ve větě 10.3, pro které platí (9.1).

Uvažujeme-li případ konečné abecedy a předpokládáme-li, že kanál příjemce zpráv je kanálem přímého pozorování, pak rychlost informace R je nahražena rychlostí entropie H a relace (9.1) relací (9.2). Pak dostaneme přesně podmínky, za kterých J. Nedoma [2] s použitím jisté myšlenky Fortetovy definoval kapacitu kanálu C_2' jako supremum rychlostí entropie H všech stacionárních a ergodických zdrojů, které lze přenésti (ve smyslu ekvivalentním definici D) uvažovaným sdělovacím kanálem.

Nahradíme-li kapacitu C_1 kapacitou C_2, pak celá obtíž je naráz odstraněna neboť platí

Věta 10.4. *Když stacionární a ergodický zdroj informace s rychlostí informace R lze přenést ve smyslu věty* 10.3, *pak každý stacionární a ergodický zdroj s rychlostí informace $R' < R$ lze přenést týmž kanálem za předpokladu, že podmínky věty* 9.1, *které se týkají odpovídajících kanálů příjemce zpráv, jsou splněny.*

Poznamenejme, že právě regularita sdělovacího kanálu nám umožňuje dokázat větu 10.4 a dává nám tak možnost uspořádat zdroje informace podle jejich rychlostí informace (vzhledem k příslušnému kanálu příjemce zpráv) s hlediska možnosti přenosu, jak už bylo uvedeno na konci § 9.

Na základě věty 10.4 můžeme formulovat další větu, ve které se používá kapacity C_2 (resp. C_2').

Věta 10.5. *Když ve větě* 10.3 *kapacita C_1 se nahradí kapacitou C_2, pak zdroj lze nebo nelze přenést (ve smyslu věty 10.3) podle toho, je-li $R < C_2$ nebo $R > C_2$.*

Když zdroj informace s konečnou abecedou má rychlost entropie H a kanál příjemce zpráv je kanálem přímého pozorování, pak zdroj lze nebo nelze přenést (ve smyslu definice D) podle toho, je-li $H < C_2'$ nebo $H > C_2'$.

11. Závěr

Nakonec několik poznámek.

Článek podává přehled současného stavu teorie informace s libovolnou abecedou, avšak s „*diskretním časem*", takže naše texty jsou posloupnosti „písmen" abecedy. Avšak v praxi se vyskytují také „texty", které jsou funkcemi času, při čemž čas probíhá všechna reálná čísla (radiofonické signály a pod.).

Teorie informace se „*spojitým časem*" je na samém počátku svého rozvoje (viz [26] a [27]). Toliko v technickém smyslu byl tento problém uvažován v [15] a to ještě ve speciálních případech. Všeobecně lze říci, že kdykoliv v technické literatuře šlo o otázky teorie informace se „spojitým časem", vždy se autoři snažili redukovat problém na případ s „diskretním časem" jako na př. pomocí t. zv. *Kotělnikovy věty* [28] (v anglické literatuře *sampling theorem*, viz [15], [29], [30] a [31]), o jejímž přibližném charakteru v nesingulárních případech se zmiňuje Kolmogorov v [26] a [27].

Uvedená věta zhruba říká, že když frekvenční pásmo W uvažovaných signálů je omezené, pak signály jsou již určeny svými hodnotami v diskretních bodech, které lze volit ekvidistantně se vzdáleností $\frac{1}{2W}$. Avšak signály s omezeným frekvenčním pásmem jsou *singulární* v tom smyslu, že znalost jejich průběhu v minulosti úplně určuje jejich průběh v budoucnosti a nemohou tedy nést žádnou informaci.

Z toho, co bylo řečeno, je vidět, že je třeba ještě mnoho udělat v teorii informace pro „spojitý čas". Tím ovšem zdaleka není řečeno, že pro případ „diskretního času" je problematika uzavřena. Naopak, jak už jsme konstatovali, zbývá ještě mnoho vykonat mimo jiné také pokud jde o aplikace.

Zdůrazněme ještě jednou, že abstraktní pravděpodobnostní charakter teorie informace jako vědy, která se snaží proniknout do podstaty pojmu zprávy a co se s ní stane, je zcela podstatný a přirozený.

Impuls pro vznik a budování teorie informace pochází od techniky přenosu zpráv. Aby se umožnil resp. usnadnil přenos zpráv v přítomnosti různých rušivých vlivů, je třeba využít výhod kodování, které při správné volbě v jistém smyslu chrání zprávu před jejím znehodnocením a bere současně ohled jak na konkretní podmínky přenosu a prostředky k jeho realisaci, tak také na př. na vysílací energii, šířku frekvenčního pásma, čas a pod.

Různé typy modulací při přenosu zpráv (na př. amplitudová, frekvenční, fázová, pulskodová a Δ-modulace) používané v současné době v technice přenosu ([29] a [30]) využívají ve větší nebo menší míře prostředky, které jsou k disposici, aby se přenos uskutečnil co nejhospodárněji, avšak je známo, že nepronikají do jádra problému. Naproti tomu obecná teorie informace tyto základní problémy zkoumá, avšak prozatím jen v tom smyslu, že určuje maximální možnosti, kterých lze dosáhnout a také to, čeho dosáhnout nelze, a její výsledky jsou tedy hlavně existenčního typu což neposkytuje metody pro stanovení konkretní realisace (kodování). Kromě toho uvádí problémy přenosu zpráv do vztahu se statistickou thermodynamikou a poskytuje prostředky pro rozvíjení statistické thermodynamiky samotné.

Nesmí se také podceňovat ta okolnost, že již ve svém současném stavu otevírá teorie informace nové obzory, což nepochybně bude mít příznivý vliv na rozvoj nejenom oboru přenosu zpráv, stejně tak jako statistická thermodynamika nám umožňuje lepší orientaci v problémech fysiky, chemie atd.

Teorie informace je totiž povolána k tomu, aby osvětlila celou řadu dalších problémů, kde pojem zprávy hraje roli. Jak známo tvoří teorie informace podstatný prvek *kybernetiky*, která se snaží proniknout do problémů „*řízení*", při čemž abstrahuje od všech jeho realisací jak v živých organismech tak také ve strojích, a poskytnout člověku možnost, jak využít znalost přírodních a společenských zákonů k dosažení stanovených cílů s maximálním vyloučením nežádoucích náhodných vlivů. Speciálně umožňují tyto metody konstruovat stroje, které mohou realisovat proces řízení na stále vyšším a vyšším stupni. Je samozřejmé, že tyto stroje mohou čím dál tím ve větší míře být schopny *samočinného přizpůsobení* statistickým vlastnostem zpráv nebo obecněji vlivu svého okolí.

Základní důležitost pro tento proces přizpůsobování má *zpětná vazba*, která umožňuje srovnávat žádaný výsledek s odhadem a podle chyby stanovené tímto srovnáním proces řízení vhodně přizpůsobovat.

Schema přenosu zpráv je speciálním případem schematu řízení. Procesu řízení podléhá právě vysílaný signál, který na straně příjemce se má reprodukovat co nejvěrněji, i když se v procesu vyskytují rušivé náhodné vlivy. Tento boj proti náhodným rušivým vlivům lze někdy uskutečnit použitím zpětné vazby, která realisuje srovnávání vysílaných a přijímavých zpráv. Avšak toto ve velké většině případů nelze realisovat a je proto třeba mimo kodování využít ještě jiných prostředků, které máme k disposici pro vyloučení rušivých vlivů (zvýšení vysílacího výkonu, rozšíření frekvenčního pásma, prodloužení doby vysílání a pod.).

Mezi otázkami teorie informace, které by bylo třeba v blízké budoucnosti řešit jsou důležité tyto:

1) Analysa zdrojů informace s hlediska „srozumitelnosti", což v naší terminologii vede na otázky *kanálu příjemce zpráv* a příslušné *charakteristické množiny*. To je důležité proto, aby pro příjemce byl splněn požadavek srozumitelnosti, ale nic víc (ekonomie). Je na př. známo, že pro telefonní přenos lidského hlasu by stačilo asi desetkrát užší frekvenční pásmo než je běžně používané, aniž se tím zhorší srozumitelnost (Vocoder a pod).

2) Řešit problémy *kodování* (na straně vysílací) a *dekodování* (na straně přijímače), aby se tím umožnil přenos charakteristických množin co nejhospodárnějším způsobem.

3) Prohloubit znalosti o vztahu teorie informace k termodynamice.

4) Prohloubit znalosti o vztahu teorie informace k teorii statistických rozhodovacích funkcí.

5) Prohloubit znalosti o vztahu teorie informace ke kybernetice.

6) Precisovat roli pojmu informace při vyjádření závislosti dvou náhodných elementů.

7) Rozšířit dosažené výsledky na případ „spojitého času".

8) Rozšířit výsledky na nestacionární a neergodické zdroje informace a nestacionární kanály (viz [4] a [5]).

9) Nalézt metody pro numerický výpočet entropie a informace (viz [32]) a vybudovat statistickou teorii těchto veličin (viz [13]).

LITERATURA

[1] *А. Я. Хинчин:* Об основных теоремах теории информации (УМН, XI, 1956, str. 17—75).

[2] *J. Nedoma:* Capacity of a discrete channel (Transactions of the First Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, 1957, str. 143—181).

[3] *И. М. Гельфанд, А. Н. Колмогоров* и *А. М. Яглом:* К общему определению количества информации (ДАН СССР, *111*, 1956, № 4, str. 745—748).

[4] *М. Розенблат-Рот:* Энтропия стохастических процессов (ДАН СССР, *112*, 1957, № 1, str. 16—19).

[5] *М. Розенблат-Рот:* Теория передачи информации через стохастические каналы (ДАН СССР, *112*, 1957, № 2, str. 202—205).

[6] *A. Perez:* Nejistota, entropie, informace (sdělení na I. pracovní konferenci čsl. matematických statistiků, Praha, červen 1954).

[7] *A. Perez:* O konvergenci posloupností nejistot, entropií a informací odpovídajících rostoucím posloupnostem σ-algeber (sdělení na IV. sjezdu čsl. matematiků, Praha, září 1955).

[8] *A. Perez:* O teorii informace v případě abstraktní abecedy (sdělení na I. pražské konferenci o teorii informace, statistických rozhodovacích funkcích a náhodných procesech, Liblice, listopad 1956).

[9] *B. McMillan:* The Basic Theorems of Information Theory (AMS, *24*, 1953, str. 196 až 219).

[10] *A. Feinstein:* A New Basic Theorem of Information Theory (Trans. IRE, PGIT, 1954, str. 2—22).

[11] *A. Perez:* Notions généralisées d'incertitude, d'entropie et d'information du point de vue de la théorie de martingales (Transactions of the First Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, Praha 1957, str. 183—208).

[12] *A. Perez:* Sur la Theorie de l'information dans le cas d'un alphabet abstrait (tamtéž str. 209—243).

[13] *A. Perez:* Sur la convergence des incertitudes, entropies et informations échantillon (sample) vers leurs valeurs vraies (tamtéž str. 245—252).

[14] *N. Wiener:* Cybernetics, New York, 1948.

[15] *C. Shannon:* A Mathematical Theory of Communication (BSTJ, *27*, 1948, str. 379 až 423 a 623—654).

[16] *L. Szilard:* Über die Entropieverminderung in einem thermodynamischen System bei Eingreifen inteligenter Wesen (Zeits. für Phys., *53*, 1929, str. 840—856).

[17] *L. Brillouin:* Maxwell's Demon Cannot Operate (Jour. Appl. Phys., *22*, 1951, str. 334).

[18] *А. Н. Колмогоров:* О некоторых асимптотических характеристиках вполне ограниченных метрических пространств (ДАН СССР, *108*, 1956, № 3, str. 385—388).

[19] *P. Halmos:* Measure Theory, New York, 1950.

[20] *А. Я. Хинчин:* Метод произвольных функций и борьба против идеализма в теории вероятностей (Философские вопросы современной физики, Москва, 1952, str. 522—538).

[21] *Colin Cherry:* Information Theory, Third London Symposium, 1955, London, 1956.

[22] *А. Я. Хинчин:* Понятие энтропии в теории вероятностей (УМН, VIII, 1953, str. 3—20).

[23] *J. Doob:* Stochastic Processes, New York, 1953.

[24] *M. Loève:* Probability Theory, New York, 1955.

[25] *A. Perez:* Transformation ou σ-algèbre suffisante et minimum de la probabilité d'erreur (Czechosl. Mathem. Jour., Vol. 7, 1957, str. 115—123).

[26] *А. Н. Колмогоров:* Теория передачи информации, Москва, 1956.

[27] *A. N. Kolmogorov:* On the Shannon Theory of Information Transmission in the Case of Continuous Signals (IRE Transactions on Information Theory, Vol IT-2, No 4, December, 1956).

[28] *В. А. Котельников:* (Материалы к 1-му Всесоюзному сьезду по вопросам реконструкции связи, 1933.)

[29] *S. Goldman:* Information Theory, London 1953. (Ruský překlad, Moskva, 1957.)

[30] *D. A. Bell:* Information Theory and its engineering applications, Second edition, London, 1956.

[31] *А. А. Харкевич:* Очерки общей теории связи, Москва, 1955.

[32] *И. М. Гельфанд* и *А. М. Яглом:* О вычислении количества информации о случайной функции, содержащейся в другой такой функции (УМН, XII, 1957, № 1 (73), str. 3—52).

Резюме

МАТЕМАТИЧЕСКАЯ ТЕОРИЯ ИНФОРМАЦИИ

АЛЬБЕРТ ПЕРЕЗ (Albert Perez)

(Поступило в редакцию 24/VI 1957 г.)

В статье дается общий обзор математической теории информации в ее настоящем состоянии.

Что касается *дискретной* теории информации (*конечный* алфавит), можно получить удовлетворительную и точную картину при изучении статей [1] и [2]. Совсем другое положение встречаем в *общей* теории информации (*произвольный* алфавит). Только в последнее время начались появляться некоторые публикации, [3], [4], [5], представляющие собой серьезные попытки обобщить некоторые результаты дискретной теории информации.

Несмотря на их сжатую форму, уже в настоящее время мы можем констатировать, что эти публикации содержат ряд результатов, похожих на некоторые из тех, которые были сообщены на различных конференциях, состоявшихся в Чехословакии в 1954, 1955, 1956 годах, [6], [7], [8].

Настоящая статья основана в значительной степени на ссылках на работы [11], [12] и [13], где читатель можеть найти полные доказательства формулированых теорем. Она разделена на две части, первая из которых содержит следующие параграфы:

1. *Теория информации и различимость.*
2. *Вероятностная концепция понятия различимости.*
3. *Степень различимости и повторения. Понятие энтропии.*
4. *Свойства обобщенной энтропии и понятие информации.*

Вторая часть содержит следующие параграфы:

5. *Мощность различимости и тонкость наблюдений. Роль понятия информации. Понятие пропускной способности канала.*

6. *Предельная теорема Мак-Миллана и ее обобщение в случае произвольного алфавита. Скорость энтропии и информации. Избыточность (Редундация).*

7. *Обобщенный вид фундаментальной леммы Фейнштейна. Понятие регулярного канала и его пропускная способность.*

8. *Вероятностная концепция понятия различимости* (продолжение): *Проблема возможности передачи источника информации через канал связи. Кодирование.*

9. *Оценка характеристической мощности в стационарном случае.*

10. *Возможность передачи (в широком смысле) источника информации через канал связи. Стационарный случай. Пропускные способносты C_1 и C_2.*

11. *Заключение.*

В § 1 стараемся обяснить действительный смысл и предмет теории информации, которая является очень молодой и сильно развивающейся отраслью прикладных наук. *Вероятностная* точка зрения по праву преобладает в теории информации, так как само *сообщение* и его *судьба*, которые являются главным предметом этой теории, имеют вероятностный характер.

Эта концепция теории информации как науки, которая изучает понятие *сообщения* в абстракции от всех физических видов его реализаций (*сигналов*) и которая использует эти виды для борьбы с пертурбациями (шумом), обесценивающими сообщение, соответствует направлению, указанному Н. Винером [14] и К. Щенноном [15].

Далее мы выделяем существенные черты понятия сообщения и энтропии на основании некоторых термодинамических рассуждений.

В § 2 изложена определенная *вероятностная концепция понятия различимости*. Определены понятия *прямо различимой группы* и *ε-различимой* *группы* через *канал* ($\mathfrak{X}$, ν_x, $\mathfrak{Y}$), (см. (2,1)). Кроме того здесь определена *мощность ε-различимости канала* как точная верхняя грань мощностей всех ε-различимых групп через этот канал.

Из множества всех *каналов наблюдения источника информации* (X, $\mathfrak{X}$, μ) мы выделяем один как собственный для наблюдателя или адресата. Этот канал ($\mathfrak{X}$, ν_x, $\mathfrak{Y}$), который мы называем *канал наблюдения адресата*, только в исключительных случаях является *каналом прямого наблюдения*.

В § 3 введено естественным способом понятие *обобщенной энтропии* при исследовании того, как возрастает различимость при повторении наблюдений. Доказано, что энтропия, обычно определенная для случая конечного вероятностного поля с помощью формулы (3,10), является частным случаем обобщенной энтропии.

Теоремы 4,1 и 4,2 в § 4 выражают некоторые важные свойства обобщенной энтропии. Указывается связь с термодинамической энтропией. *Обобщенная информация* здесь определена как в обычном (конечном) случае, но с помощью обобщенных энтропий (см. (4,11)). Теоремы 4,3 и 4,4 выражают некоторые основные свойства обобщенной информации, и теорема 4,8 показывает, что она может быть аппроксимирована с помощью обычных информаций.

Понятия информации и ее *скорости* R и понятие *пропускной способности* C стационарного независимого канала (см. (5,4)) введены в § 5 естественным способом при изучении мощности различимости этого канала.

В § 6 *предельная теорема* Мак-Миллана [9] обобщена для случая *абстрактного алфавита*. Теорема 6,1 обеспечивает существование *обобщенной скорости энтропии* (6,6), теорема 6,2 является обобщенным видом предельной теоремы Мак-Миллана для энтропии и теорема 6,3 — для информации.

В § 7 на основании этой последней теоремы и именно на основании „*обобщенного свойства E*“, которое следует из нее, доказан обобщенный вид *фундаментальной леммы* Фейнштейна (см. [10] и [1]) для случая произвольных алфавитов (теорема 7,1).

В § 8 продолжаем изложение вероятностной концепции понятия различимости, которое было начато в § 2. Мы рассматриваем *проблему возможности передачи* (продукции) источника информации через *канал связи*. Этот канал надо в общем случае различать от канала наблюдения адресата не только потому, что возможно существование предварительного *кодирования* источника информации (кодирование — это преобразование, в общем измеримое, измеримого пространства источника информации в измеримое пространство на *входе* канала), но и потому, что факторы, независящие от природы адресата и его идеальных возможностей, часто способствуют тому, что этот канал является менее выгодным, чем канал наблюдения адресата.

Рассматривается *проблема характеризации источника информации с точки зрения его передачи*, принимая во внимание, с одной стороны, интересы адресата, а, с другой стороны, требования экономии. Таким образом дойдем к понятию *характеристической группы и ее мощности*, которые соотвествуют источнику, наблюдаемому через канал наблюдения адресата.

По определению источник информации *можно передать* (в узком смысле; его расширение рассматриваем в § 10) через канал связи, если можно закодировать его так, чтобы его характеристическая группа, различимая через канал наблюдения адресата с данной степенью различимости, осталась различимой через рассматриваемый канал с той же степенью различимости. Теорема 8,1 дает условия возможности передачи в выше приведенном смысле.

В § 9 доказана на основании теоремы 6,3 теорема 9,1, дающая оценку *характеристической мощности* стационарного источника информации, который при наблюдении через канал наблюдения адресата, который считаем стационарным, создает *двойной источник*, имеющий обобщенное свойство E. Теорема 3 в [22] является следствием теоремы 9,1 для случая конечного алфавита и канала прямого наблюдения адресата (следствие 9,1). Кроме того, теорема 9,1 позволяет *упорядочить* источники информации по их скоростям информации с точки зрения возможности их передачи через регулярный канал связи.

В § 10 расширяется понятие возможности передачи источника информации через канал связи, содержащее как специальный случай определение возможности передачи, обычно приводимое в дискретной теории информации (определение D) при помощи понятия *вероятности ошибки* (см. (10,1)). Эта вероятность связана с понятием ε-различимости посредством теоремы 10,1 (см. теорема 10,2).

Теорема 10,3 дает достаточное условие для возможности передачи источника информации в широком смысле (см. (10,2)), где используется пропускная способность C_1, обобщающая пропускную способность, обычно определяемую в дискретной теории информации [1]. Соответствующая теорема для случая источника информации с конечным алфавитом („Первая теорема Шеннона") является частным случаем теоремы 10,3 при условии, что канал наблюдения адресата эквивалентен каналу прямого наблюдения.

К сожалению, пропускная способность C_1, определенная выражением (7,1) не может обеспечить то, чтобы в случае, когда скорость информации (через канал наблюдения адресата) $R > C_1$, передача рассматриваемого источника была невозможной. Но ничто нам не мешает определить вторую *пропускную способность* C_2 как точную верхную грань скоростей информации R всех стационарных и эргодичных источников, которые, подобно тому как в теореме 10,3, можно передать через рассматриваемый канал и для которых имеет место соотношение (9,1). С помощью теоремы 10,4 (которая упорядочивает источники информации по их скоростям с точки зрения возможности их передачи) и этого определения пропускной способности можно доказать теорему 10,5, которая утверждает, что если $R < C_2$, то источник можно передать через рассматриваемый канал и если $R > C_2$, то это невозможно.

Résumé

THÉORIE MATHÉMATIQUE DE L'INFORMATION

ALBERT PEREZ

(Reçu le 24 juin 1957.)

Dans cet article on donne un aperçu général de la théorie mathématique de l'information, comme elle se présente dans son état actuel.

En ce qui concerne la théorie *discrète* de l'information (alphabet *fini*) on pourrait obtenir une image très satisfaisante et précise en étudiant les articles [1] et [2]. Toute autre est la situation avec la théorie *générale* de l'information (alphabet *quelconque*). C'est seulement dans le tout dernier temps que commencent à faire leur apparition certaines publications, [3], [4], [5], représentant des tentatives sérieuses de généralisation de certains résultats de la théorie discrète de l'information. Malgré leur forme succincte, nous pouvons dès à présent constater que ces publications contiennent une série de résultats semblables à ceux qui ont été communiqués en différentes conférences ayant eu lieu en Tchécoslovaquie pendant les années 1954, 1955 et 1956, [6], [7], [8].

Le présent article est dans une large mesure basé sur les références [11], [12] et [13], où le lecteur pourrait trouver les démonstrations complètes des théorèmes formulés. Il est divisé en deux parties, dont la première contient les paragraphes suivants:

1. *Théorie de l'information et discernabilité.*

2. *Sur une conception probabiliste de la notion de discernabilité.*

3. *Degré de discernabilité et répétition. Notion d'entropie.*

4. *Propiétés de l'entropie généralisée et notion d'information.*

5. *Puissance de discernabilité et finesse des observations. Le rôle de la notion d'information. Notion de capacité d'un canal.*

6. *Théorème limite de McMillan et sa généralisation au cas d'un alphabet quelconque. Débit d'entropie et d'information. Redundance.*

7. *Version généralisée du lemme fondamental de Feinstein. Notions de canal régulier et de capacité.*

8. *Sur une conception probabiliste de la notion de discernabilité* (suite): *problème de transmissibilité d'une source d'information à travers un canal de communication. Codage.*

9. *Estimation de la puissance caractéristique dans le cas stationnaire.*

10. *Transmissibilité (dans le sens large) d'une source d'information à travers un canal de communication. Cas stationnaire. Capacités* C_1 *et* C_2.

11. *Conclusion.*

Au § 1 on tâche de mettre en lumière le sens véritable et l'objet de la théorie de l'information proprement dite, branche très jeune des sciences appliquées se trouvant en pleine évolution. L'aspect *probabiliste* domine dans la théorie de l'information à bon droit justement à cause du fait que le *message* et son *destin*, qui constituent l'objet principal de cette théorie, sont de nature probabiliste.

Cette conception de la théorie de l'information comme science qui cherche à pénétrer la notion de *message* en faisant abstraction de toutes les formes physiques sous lesquelles il peut être réalisé (*signal*) et mettant à profit ces formes pour combattre les effets perturbateurs (*bruit*) tendant à désintégrer le message pendant sa transmission, est dans la ligne tracée par N. Wiener [14] et C. Shannon [15].

On fait ressortir les caractères essentiels des notions de message et d'entropie sur la base de certains raisonnements d'ordre thermodynamique.

Au § 2 est développée une certaine *conception probabiliste de la notion de discernabilité*. Sont définies les notions de *groupe directement discernable* et *ε-discernable* à travers un *canal* $(\mathfrak{X}, \nu_x, \mathfrak{Y})$. Voir (2,1). On définit la *puissance de ε-discernabilité d'un canal* comme le supremum des puissances de tous les groupes ε-discernables à travers ce canal.

Parmi les *canaux d'observation* d'une *source d'information* $(X, \mathfrak{X}, \mu)$ on distingue un comme propre à l'observateur ou destinataire, le *canal d'observation du destinataire* $(\mathfrak{X}, \nu_x, \mathfrak{Y})$, qui dans des cas exceptionnels seulement est un *canal d'observation directe.*

Au § 3 on introduit la notion d'*entropie généralisée* d'une façon naturelle en étudiant comment augmente la discernabilité par répétition des observations. On montre que l'entropie ordinairement définie pour le cas d'un champ de probabilité *fini* à l'aide de la formule (3.10) est un cas particulier de l'entropie généralisée.

Les théorèmes 4,1 et 4,2 en § 4 expriment certaines propriétés importantes de l'entropie généralisée. On indique son rapport à l'entropie thermodynamique. *L'information généralisée* est ici définie comme dans le cas ordinaire (fini) à l'aide des entropies généralisées (voir (4,11)). Les théorèmes 4,3 et 4,4 expriment certaines propriétés fondamentales de l'information généralisée et le théorème 4,8 montre qu'elle peut être approximée par des informations ordinaires.

Au § 5 les notions d'information et de son *débit R* ainsi que la notion de *capacité C* d'un canal stationnaire indépendant (voir (5,4)) sont introduites d'une façon naturelle en étudiant la puissance de discernabilité de ce dernier.

Au § 6 le *théorème limite de McMillan* [9] est généralisé au cas d'un *alphabet abstrait*. Le théorème 6,1 assure l'existence du *débit d'entropie généralisée* (6,6), le théorème 6,2 est la version généralisée du théorème limite de McMillan pour l'entropie et le théorème 6,3 est celle pour l'information.

Au § 7, sur la base de ce dernier théorème et notamment sur la base de la *„propriété E généralisée"* qui en résulte, on montre une version généralisée du *lemme fondamental de Feinstein* [10], [1], pour le cas des alphabets quelconques (théorème 7,1).

Au § 8 on poursuit le développement de la conception probabiliste de la discernabilité commencé au § 2. On considère le *problème de transmissibilité* d'une source d'information à travers un *canal de communication*, qui devra, en général, être distingué du canal d'observation du destinataire non seulement par le *codage* préalable éventuel de la source d'information (transformation, en général mesurable, de l'espace mesurable de la source d'information considérée dans l'espace mesurable d'*entrée* du canal), mais aussi par le fait que, le plus souvent, des facteurs indépendants de la nature même du destinataire et de ses possibilités idéales tendent à rendre ce canal moins préféré que le canal d'observation du destinataire.

On considère le *problème de caractérisation d'une source d'information en vue de sa transmission*, tenant compte du destinataire d'une part et de l'économie d'autre part. On aboutit ainsi à la notion de *groupe caractéristique* et de sa *puissance* correspondant à la source observée à travers le canal d'observation du destinataire.

Par définition, une source d'information sera *transmissible* (dans le sens étroit, élargi au § 10) à travers un canal de communication si, par un codage approprié, son groupe caractéristique, discernable à travers le canal d'observation du destinataire avec un degré de discernabilité fixé, reste discernable à travers le canal considéré avec le même degré de discernabilité. Le théorème 8,1 donne des conditions de transmissibilité dans le sens précédent.

Au § 9 on démontre sur la base du théorème 6,3 le théorème 9,1 donnant une estimation de la *puissance caractéristique* d'une source d'information stationnaire laquelle, observée à travers le canal d'observation du destinataire supposé stationnaire, engendre une *source double* ayant la propriété E généralisée. La théorème 3 de [22] en est un cas particulier correspondant à un alphabet fini et à un canal d'observation du destinataire équivalent à un canal d'observation directe (Corollaire 9,1). Le théorème 9,1 nous permet, en particulier, *d'ordonner* les sources d'information considérées en fonction de leurs débits d'information quant à leur transmissibilité à travers un canal de communication régulier.

Au § 10 on élargit la notion de transmissibilité d'une source d'information à travers un canal de communication, qui contient alors comme cas spécial la définition de transmissibilité donnée habituellement en théorie discrète de l'information (par la suite: définition D) à l'aide de la notion de *probabilité d'erreur* (voir (10,1)), liée à la notion de ε-discernabilité par le théorème 10,1 (voir théorème 10,2).

Le théorème 10,3 donne une condition suffisante pour la transmissibilité d'une source d'information dans le sens élargi (vois (10,2)) impliquant la capacité C_1, généralisant la capacité ordinairement définie dans la théorie discrète de l'information [1]. Le théorème correspondant au cas d'une source d'information à alphabet fini („Premier théorème de Shannon") est contenu comme un cas spécial dans le théorème 10,3, le canal d'observation du destinataire étant équivalent à un canal d'observation directe.

Malheureusement, la capacité C_1 définie par (7,1) n'est pas de nature à nous assurer que, si le débit d'information (à travers le canal d'observation du destinataire) $R > C_1$, la source considérée n'est pas transmissible. Rien ne nous empêche cependant de définir une seconde *capacité* C_2 comme le supremum des débits d'information R de toutes les sources stationnaires et ergodiques transmissibles à travers le canal considéré, comme dans le théorème 10,3, pour lesquelles la relation (9,1) est valable. A l'aide du théorème 10,4 (ordonnant les sources d'information en fonction de leurs débits quant à leur transmissibilité) et de cette définition de capacité il est alors possible de montrer le théorème 10,5, qui nous dit que si $R < C_2$ la source est transmissible et si $R > C_2$ la source n'est pas transmissible à travers le canal considéré.