Sobre o Teorema dos Códigos para um Canal sem Ruído (*)

por J. Marques Henriques
Universidade Técnica de Lisboa

0. Sumário.

É examinado o teorema dos códigos para um canal sem ruído dum ponto de vista probabilístico, num caso especial, tomando a expansão relativa à base θ dum ponto do intervalo unitário, considerando-a como a realização do processo estocástico a ser codificado e relacionando-a com propriedades conhecidas do intervalo unitário, em particular no que se refere à dimensão de HAUSDORFF BESICOVITCH e ao teorema de SHANNON-McMILLAN-BREIMAN.

1. Generalidades e propriedades básicas.

Seja $\Omega = (0,1]$ e $\mathfrak{R}$ o conjunto dos borelianos de Ω. Para cada $\omega \in \Omega$ associemos-lhe a sua expansão infinita relativamente a uma base $\theta \geqq 2$ (θ inteiro), ou seja,

$$\omega = \sum_{i=1}^{\infty} a_i(\omega)/\theta^i,$$

onde $a_i(\omega) = 0, 1, \cdots, \theta - 1$. Então cada $a_i(\omega)$ é uma função mensurável relativamente a $(\Omega, \mathfrak{R})$ (para detalhes relativamente à notação ver [4]).

Deste modo, se P for uma medida de probabilidade sobre $\mathfrak{R}$, $\{a_i(\omega) : i = 1, 2, \cdots\}$ é um processo estocástico relativamente a $(\Omega, \mathfrak{R}, P)$, com espaço-amostra finito, $\Xi = \{0, 1, \cdots, \theta - 1\}$, a que aqui chamamos *alfabeto*, sendo os elementos $0, 1, \cdots, \theta - 1$ as suas *letras*.

Inversamente, pode provar-se que se a medida de probabilidade P for não-atómica (isto é, se com $P(A) > 0$ existir sempre um $B \subset A$ com $0 < P(B) < P(A)$), então qualquer processo estocástico com alfabeto Ξ pode representar-se desta forma.

Neste trabalho ocupar-nos-emos em particular daquelas medidas não-atómicas P, relativamente às quais o processo estocástico $\{a_i(\omega) : i = 1, 2, \cdots\}$ é estacionário e ergódico.

Por definição o processo é *estacionário* se $P(\omega : a_1(\omega), \cdots, a_n(\omega)) = P(\omega : a_{k+1}(\omega), \cdots, a_{k+n}(\omega))$, com qualquer inteiro k. Quanto à *ergodicidade*, diremos que para a transformação T definida sobre Ω por $T\omega = \theta\omega \pmod 1$ A é invariante se para $A \in \mathfrak{R}$, $T^{-1}A = A$, e que P é ergódica relativamente a T se para todo o $A \in \mathfrak{R}$ invariante $P(A) = 0$ ou $P(A) = 1$.

DEFINIÇÃO 1. Chamaremos *código* a qualquer função χ contínua e não decrescente em $[0,1]$, tal que $\chi(0) = 0$ e $\chi(1) = 1$.

Em correspondência com cada ω obtemos então a expansão de $\chi(\omega)$ relativamente à

(*) Trabalho apresentado no «Simpósio sobre as Teorias da Informação e dos Sistemas», Lisboa, Setembro de 1970.

base θ, de tal modo que virá

$$\chi(\omega) = \sum_{i=1}^{\infty} b_i(\omega)/\theta^i.$$

Agora podemos interpretar o código χ como um esquema tal que a cada sucessão ou *mensagem original* $a = (a_1, a_2, \cdots)$ de letras de Ξ lhe faz corresponder uma outra sucessão ou *mensagem codificada* $b = (b_1, b_2, \cdots)$. Por uma questão de simplicidade só consideraremos códigos com o mesmo alfabeto Ξ para domínio e contradomínio, aos quais também chamaremos *input* e *output*.

Vamos assumir daqui em diante que todas as medidas de probabilidade P são não-atómicas; então o código χ é tal que com probabilidade 1 podemos determinar as primeiras n letras de b conhecido que seja um número finito de letras de a.

Em todo este artigo só nos ocuparemos de um *canal sem ruído*, isto é, um código ao qual estão associados um input e um output, cada um dos quais possui o alfabeto Ξ. Do ponto de vista matemático o canal sem ruído é simplesmente o alfabeto Ξ, ao qual se associam todas as possíveis mensagens. Se uma mensagem a é enviada através do canal, ela é recebida com perfeito rigor depois de codificada, b.

O teorema dos códigos para um canal sem ruído, de que é uma versão especial no nosso caso o Teorema 5 abaixo, pretende estabelecer condições segundo as quais uma mensagem, depois de codificada, $b = \chi(a)$, pode ser retransmitida do output para o input e aí reconstituída a mensagem inicial ([6] e [1]).

DEFINIÇÃO 2. Intervalo fundamental de ordem n de Ω é um subconjunto $\Delta^{(n)}_{k_1,\ldots,k_n}$, tal que

$$\Delta^{(n)}_{k_1,\ldots,k_n} := \{\omega \in \Omega : a_i(\omega) = k_i, \; i = 1, \cdots, n \text{ e } k_i \in \Xi\}.$$

É evidente que para qualquer n, $\Delta^{(n)}$ é um conjunto de $\mathfrak{R}$; inversamente, os conjuntos $\Delta^{(n)}$ geram a σ-álgebra $\mathfrak{R}$.

No que se segue omitimos os índices $k_1, \cdots, k_n$, visto não haver qualquer possibilidade de confusão, pois trata-se de elementos prèviamente escolhidos de Ξ. Também no caso de n ser bem definido escreveremos apênas $\Delta(\omega)$ ou Δ.

Da Definição 2 imediatamente se deduz que para todo o $n \geqq 1$ e $\omega \in \Omega$ existe um e um só intervalo fundamental de ordem n que contém ω. Indicá-lo-emos por $\Delta^{(n)}(\omega)$ ou $\Delta(\omega)$ sempre que seja necessário individualizá-lo. Este é então o intervalo θ-ádico do tipo $(l/\theta^n, (l+1)/\theta^n]$, com $l = 0, 1, \cdots, \theta^n - 1$ que contém ω, pois os θ^n intervalos de ordem n possuem todos o mesmo comprimento, pelo que se se designar por λ a clássica medida de LEBESGUE sobre Ω será $\lambda(\Delta) = \theta^{-n}$.

Mas agora, se os n primeiros símbolos da expansão de ω são conhecidos, também o é Δ, e é então evidente que $\chi(\omega) \in \chi(\Delta)$. Denotemos por $\{\chi(\Delta)\}$ o intervalo θ-ádico mais pequeno que contém $\chi(\Delta)$; deste modo o número de símbolos do alfabeto Ξ na expansão de $\chi(\omega)$ que pode ser determinado de maneira unívoca é exactamente a ordem do intervalo θ-ádico $\{\chi(\Delta)\}$ mas esta é $-\log_\theta \lambda \{\chi(\Delta)\}$, onde, tal como acima, λ é a medida de LEBESGUE. Isso é, aliás, óbvio, no caso de ser $\chi(\omega) \equiv \omega$, por vir $-\log_\theta \lambda(\Delta) = n$.

Deste modo, os primeiros n símbolos da expansão de ω determinam exactamente $-\log_\theta \lambda \{\chi(\Delta^{(n)}(\omega))\}$ símbolos na expansão de $\chi(\omega)$.

Assim, podemos medir a eficiência do código χ, considerando *a relação entre o número de símbolos determinados na expansão de $\chi(\omega)$ e o número de símbolos da expansão de ω necessários para a determinação daqueles*, a que chamamos a *compressão* originada

por χ, e que denotaremos por

$$C_n(\omega) := -n^{-1}\log_\theta \lambda\{\chi(\Delta^{(n)}(\omega))\}.$$

Para simplificar as deduções que vamos fazer substituiremos C_n por

$$D_n(\omega) := -n^{-1}\log_\theta \lambda(\chi(\Delta^{(n)}(\omega))).$$

Um código χ será *eficiente* se $C_n(\omega)$ for pequeno, no limite para n infinito. Assim, definimos

$$C_\chi(\omega) := \lim_{n\to\infty} C_n(\omega)$$

e

$$C^*_\chi(\omega) := \liminf_{n\to\infty} C_n(\omega)$$

no caso de existirem ambos os limites, e anàlogamente para $D_\chi(\omega)$ e $D^*_\chi(\omega)$.

Suponhamos agora que é dada uma medida de probabilidade P sobre $(\Omega, \mathfrak{R})$, tal que P é estacionária, não-atómica e ergódica. Se for $F(x) := P(0, x]$ a função de distribuição de P, então F é um código. Mostraremos abaixo que para este código é

$$P(\omega : D_F(\omega) = h) = 1,$$

onde h é a entropia relativa de $\{a_i\}$ em relação a P, ou seja, $h = \lim_{n\to\infty} \int_\Omega D_n(\omega)\, dP$ (a entropia relativa será simplesmente a entropia dividida por $\log\theta$).

Mostraremos ainda que para qualquer outro código χ é:

$$P(\omega : D^*_\chi(\omega) < h) = 0,$$

de modo que F possui a eficiência máxima, como aliás seria de esperar. Quer isto dizer que para um dado processo um código será optimal se considerado como função sobre Ω, for precisamente a função de distribuição deste processo.

2. O código F.

Se, tal como acima, designarmos por F a função de distribuição de P, então F é um código que goza da propriedade de que com $\omega \neq \omega'$,

$$P(\omega : F(\omega) = F(\omega')) = 0$$

e isto por P ser não-atómica; por outras palavras, a mensagem original pode, com probabilidade 1, deduzir-se da mensagem codificada.

TEOREMA 1. *Se P for não-atómica, estacionária e ergódica, então*

$$P(\omega : D_F(\omega) = h) = 1,$$

onde h é a entropia relativa de $\{a_i\}$ relativamente a P.

DEM.: Por ser P não-atómica, vem $\lambda(F(\Delta)) = P(\Delta)$ para qualquer intervalo Δ (isto é o mesmo que afirmar que se X for uma variável aleatória com função de distribuição $G(x)$, então $G(X)$ é uma variável aleatória uniformemente distribuída no intervalo unitário). Por conseguinte

$$D_n(\omega) = -n^{-1}\log_\theta P(\Delta^{(n)}(\omega)),$$

e a afirmação é agora uma consequência do teorema de SHANNON-MCMILLAN-BREIMAN [1, pág. 197]. Q. E. D.

3. Alguns tópicos sobre a dimensão de Hausdorff-Besicovitch de subconjuntos de Ω.

Nesta Secção apresentaremos apenas a definição e propriedades mais importantes da dimensão de HAUSDORFF-BESICOVITCH de subconjuntos de Ω, que nos será de utilidade posterior.

Consideremos de novo o espaço de probabilidade $(\Omega . \mathfrak{K}, P)$ não-atómico e seja A um subconjunto qualquer de Ω (não necessàriamente em $\mathfrak{K}$). Considere-se uma *cobertura* de A por conjuntos Z_i de $\mathfrak{K}$, em número finito ou infinito denumerável:

$$A \subset \bigcup_{i=1}^{\infty} Z_i, \quad Z_i \in \mathfrak{K}.$$

Agora, uma vez que o espaço de probabilidade é não atómico, será sempre possível, dado $\rho > 0$ encontrar conjuntos $\{Z_i\}$ em $\mathfrak{K}$ que cobrem A, com $P(Z_i) < \rho$; chamar-lhe-emos uma *cobertura* ρ *de* A. Então, fixado $\alpha > 0$ definimos para qualquer $A \subset \Omega$:

$$\Gamma_\alpha(A;\rho) := \inf\left\{\sum_{i=1}^{\infty} P(Z_i)^\alpha\right\};$$

tomado o ínfimo para todas as coberturas ρ de A (com $P(Z_i) < \rho$; nada impede aqui que $\Gamma_\alpha(A;\rho)$ tome o valor $+\infty$),

Quando $\rho \downarrow 0$, $\Gamma_\alpha(A;\rho)$ é não decrescente, pois com $\rho' < \rho$ toda a cobertura ρ' de A é também uma cobertura ρ de A e deste modo o ínfimo é tomado para classes cada vez mais reduzidas de coberturas de A, e portanto o limite de $\Gamma_\alpha(A;\rho)$ para $\rho \downarrow 0$ existe. Então definimos

$$\Gamma_\alpha(A) := \lim_{\rho \downarrow 0} \Gamma_\alpha(A;\rho).$$

A $\Gamma_\alpha(A)$ chama-se por vezes a medida exterior α-dimensional de A. É evidente que $\Gamma_\alpha(\cdot)$ é monótona não decrescente: para A, $B \subset \Omega$, se $A \subset B$, então $\Gamma_\alpha(A) \leqq \Gamma_\alpha(B)$.

E, dada uma sucessão de conjuntos A_n de Ω, podem escolher-se para cada um deles coberturas $\{Z_{ni}\}$, com $P(Z_{ni}) < \rho$, tais que

$$\sum_{i=1}^{\infty} P(Z_{ni})^\alpha < \Gamma_\alpha(A_n;\rho) + \varepsilon 2^{-n} \leqq \Gamma_\alpha(A) + \varepsilon 2^{-n}.$$

Deste modo a totalidade dos Z_{ni} constitui uma cobertura ρ da união $\bigcup_n A_n$ tal que

$$\sum_n \sum_i P(Z_{ni})^\alpha < \sum_n \Gamma_\alpha(A_n) + \varepsilon$$

e assim vemos que $\Gamma_\alpha(\cdot)$ é uma função sub-aditiva, ou seja:

$$\Gamma_\alpha\left(\bigcup_n A_n\right) \leqq \sum_n \Gamma_\alpha(A_n).$$

Quer dizer, tomada como função de A, $\Gamma_\alpha(A)$ é, de facto, uma medida exterior no sentido de CARATHÉODORY.

TEOREMA 2. *Com* $\alpha > 0$, $\Gamma_\alpha(A) < \infty$ *implica* $\Gamma_{\alpha+\varepsilon}(A) = 0$ *para todos os* $\varepsilon > 0$.

DEM. Se for $\{Z_i\}$ uma cobertura ρ de A para a qual

$$\sum_i P(Z_i)^\alpha \leqq \Gamma_\alpha(A;\rho) + \varepsilon \leqq \Gamma_\alpha(A) + 1 = K < \infty$$

(da Definição de $\Gamma_\alpha(A;\rho)$, uma tal cobertura existe sempre), então também

$$\Gamma_{\alpha+\varepsilon}(A;\rho) \leqq \sum_i P(Z_i)^{\alpha+\varepsilon} \leqq$$

$$\leqq \rho^\varepsilon \sum_i P(Z_i)^\alpha < \rho^\varepsilon K.$$

E, por ser $\varepsilon > 0$, fazendo $\rho \downarrow 0$ logo se vê que $\Gamma_{\alpha+\varepsilon}(A) = 0$. Q. E. D.

Anàlogamente se mostra que se $\Gamma_\alpha(A) > 0$ então para todos os ε tais que $0 < \varepsilon < \alpha$ vem $\Gamma_{\alpha-\varepsilon}(A) = \infty$.

Quer dizer: há um ponto α_0, tal que $\Gamma_\alpha(A) = \infty$ para $\alpha < \alpha_0$ e $\Gamma_\alpha(A) = 0$ para $\alpha > \alpha_0$. O valor de $\Gamma_\alpha(A)$ em α_0 pode ser nulo, finito e positivo ou $+\infty$. A este número α_0, bem determinado, chamamos

então a dimensão de HAUSDORFF-BESICOVITCH de A (relativamente à medida de probabilidade P):

DEF.

$$\dim_P A := \sup\{\alpha : \Gamma_\alpha(A) = \infty\} = \\ = \inf\{\alpha : \Gamma_\alpha(A) = 0\}.$$

No caso de ser $\Gamma_\alpha(A) < \infty$ para todos os $\alpha > 0$, definiremos $\dim A$ como sendo 0 (omitiremos a indicação da medida de probabilidade P sempre que não haja qualquer motivo para dúvidas, o que aliás sucede sempre abaixo; no caso de $P \equiv \lambda$, $\dim A$ coincide com a vulgar dimensão de HAUSDORFF de A).

Mostra-se com facilidade que $\dim \phi = 0$ (ϕ é o conjunto vazio e $\dim \Omega = 1$. Também é válido o

TEOREMA 3. *A dimensão de Hausdorff-Besicovitch goza das propriedades seguintes:*

a) A, $B \subset \Omega$ *então* $dim\, A \leqq dim\, B$;

b) $dim\, A = 1$ *para todos os* $A \in \mathfrak{R}$ *com probabilidade positiva;*

c) $0 \leqq dim\, A \leqq 1$;

d) $dim \bigcup_{i=1}^{\infty} A_i = sup_i\, dim\, A_i$.

DEM. *a*)-*c*) deduzem-se imediatamente da Definição;

d) é ligeiramente mais complexa: se $\alpha > \sup \dim A_i$, então $\Gamma_\alpha(A_i) = 0\ (i = 1, 2, \cdots)$ e por isso, pela subaditividade de $\Gamma_\alpha(\cdot)$ como medida exterior sobre Ω, também

$$\Gamma_\alpha\Big(\bigcup_i A_i\Big) = 0,$$

o que implica

$$\dim \bigcup_i A_i \leqq \alpha;$$

agora, para provar a desigualdade no outro sentido, se $\alpha < \sup_i \dim A_i$, para i_0 convenientemente fixado é evidente que $\alpha < \dim A_{i_0}$ e deste modo $\Gamma_\alpha(A_{i_0}) = \infty$. Isto acarreta $\Gamma_\alpha\Big(\bigcup_i A_i\Big) = \infty$ (por *a*) acima) e por consequência $\dim \bigcup_i A_i \geqq \alpha$. Q. E. D.

Um resultado importante, mas que pela extensão da demonstração não provaremos aqui (ver [2] e [4]) é o seguinte:

TEOREMA 4. *Se* P *for uma medida de probabilidade não atómica sobre* $\mathfrak{R}$, *e* Π *uma outra medida sobre* $\mathfrak{R}$ *com* $\Pi(\Omega) < \infty$, *então para o conjunto*

$$A := \left\{\omega : \liminf_{n\to\infty} \frac{log\, \Pi(\Delta^{(n)}(\omega))}{log\, P(\Delta^{(n)}(\omega))} \leqq \delta\right\}$$

vem $dim\, A \leqq \delta$.

4. O código geral.

Nesta secção vamos mostrar que o código F da Secção 2 é optimal no sentido de que para nenhum outro código χ se obtém uma compressão inferior a h.

TEOREMA 5. *Sendo* P *não-atómica, estacionária e ergódica e* χ *um código qualquer, então*

$$P(\omega : D^*_\chi(\omega) < h) = 0.$$

DEM. Considere-se a medida de probabilidade Π sobre $\mathfrak{R}$, cuja função de distribuição é χ. Como χ é contínua, Π é não-atómica e para qualquer intervalo Δ, $\Pi(\Delta) = \lambda(\chi(\Delta))$. Assim,

$$D_n(\omega) = -\frac{1}{n}\log_\theta \Pi(\Delta^{(n)}(\omega)) =$$

$$= \frac{\log_\theta \Pi(\Delta^{(n)}(\omega))}{\log_\theta P(\Delta^{(n)}(\omega))}\left[-\frac{1}{n}\log_\theta P(\Delta^{(n)}(\omega))\right].$$

Mas, de acordo com o Teorema 1, o segundo factor do lado direito converge para h, a menos de um conjunto de medida P nula. Por isso,

$$D^*(\omega) = h \liminf_{n\to\infty} \frac{\log_\theta \Pi(\Delta^{(n)}(\omega))}{\log_\theta P(\Delta^{(n)}(\omega))} [P].$$

Ora, aplicando o Teorema 4 com $\delta = \beta/h$ e $\beta < h$ vem

$$\dim\left(\omega : \liminf_{n\to\infty} \frac{\log_\theta \Pi(\Delta^{(n)}(\omega))}{\log_\theta P(\Delta^{(n)}(\omega))} \leqq \frac{\beta}{h}\right) < 1.$$

Mas como qualquer conjunto de medida P positiva possui dimensão 1, então isso acarreta que com $\beta < h$

$$P\left(\omega : \liminf_{n\to\infty} \frac{\log_\theta \Pi(\Delta^{(n)}(\omega))}{\log_\theta P(\Delta^{(n)}(\omega))} \leqq \frac{\beta}{h}\right) = 0,$$

ou seja

$$P(\omega : D^*_\chi(\omega) \leqq \beta) = 0.$$

Q. E. D.

Antes, na Secção 1, tínhamos assumido que $\chi(0) = 0$ e $\chi(1) = 1$. Contudo podemos relaxar ligeiramente esta hipótese fazendo apenas $0 \leqq \chi(0) \leqq \chi(1) \leqq 1$ e definindo $\Pi(0, x] := \chi(x) - \chi(0)$. Neste caso será ainda $\Pi(\Omega) < \infty$, embora não uma medida de probabilidade, e a demonstração permanece inalterável.

Como $P(\omega : D^*_\chi(\omega) > h) = 1$, pelo Lema de FATOU vem

$$\liminf_{n\to\infty} \int_\Omega D_n(\omega)\, dP \geqq h,$$

o que quer dizer que h é não só a compressão média, mas também a compressão mínima, de acordo com o Teorema 5. Este resultado é devido a KHINCHIN [5, pág. 24], embora partindo de definições diferentes e vindo aí o limite superior em vez do limite inferior.

Como exemplo de aplicação consideremos o conjunto de CANTOR, isto é, com $\theta = 3$

$$K := \{\omega : a_n(\omega) = i; \quad i = 0, 2; \quad n = 1, 2, \cdots\}$$

ao qual corresponde a medida de probabilidade singular

$$P(\omega : a_n(\omega) = i) = p_i,$$

com $p_0 = p_2 = 1/2$ e $p_1 = 0$. Mostra-se com facilidade que $\lambda(K) = 0$. Neste caso o código optimal consiste em substituir na expansão de qualquer ω e K os dígitos 2 por dígitos 1, considerando as mensagens constituídas por 0's e 1's como expansões binárias tomadas relativamente à base 3. A compressão que assim se obtém é $\log 2/\log 3$, eventualmente o mesmo valor que para a dimensão de HAUSDORFF-BESICOVITCH deste mesmo conjunto, tomada relativamente à medida de LEBESGUE.

Um exemplo clássico é devido a SHANNON: com $\theta = 4$, façamos $P(\omega : a_n(\omega) = i) = p_i$ com $p_0 = 1/2$, $p_1 = 1/4$, e $p_2 = p_3 = 1/8$. Então substituindo cada letra numa mensagem a codificar por um conjunto de 0's e de 1's, de acordo com o código $0 \to 0$, $1 \to 10$, $2 \to 110$, $3 \to 111$, vem esta mensagem substituída por uma outra de dígitos binários. Estes símbolos são agora agrupados dois a dois e traduzidos para a base $\theta = 4$ de acordo com a tabela $00 \to 0$, $01 \to 1$, $10 \to 2$ e $11 \to 3$. Quer dizer, qualquer $\omega = (a_1(\omega), a_2(\omega), \cdots)$ vem codificado $\chi(\omega) = (b_1(\omega), b_2(\omega), \cdots)$, onde χ é a função de

de distribuição de P, definida acima. Então o código χ é optimal, obtendo-se a compressão média de 7/8 [6, pág. 63-4].

BIBLIOGRAFIA

[1] Ash, Robert B. (1965). *Information Theory*. John Wiley & Sons, New York — London.

[2] Billingsley, Patrick P. (1960). Hausdorff dimension in probability theory, *Ill. J. Math.* **4** 187-209.

[3] ——— (1965). *Ergodic Theory and Information*. John Wiley & Sons, New York—London.

[4] Henriques, J. Marques. (1967). *Hausdorff-Besicovitch Dimension for Probability Product Spaces*, Master's Paper, Department of Statistics, The University of Chicago, Chicago, Ill.

[5] Khinchin, A. Ya. (1957). *Mathematical Foundations of Information Theory*. Dover Publications, New York.

[6] Shannon, Claude E. (1949). *The Mathematical Theory of Communication*. The University of Illinois Press, Urbana, Ill. (with Warren Weaver).