

THE CAPACITY OF A GENERAL NOISELESS CHANNEL AND ITS CONNECTION WITH HAUSDORFF DIMENSION

MEIR SMORODINSKY¹

1. Introduction. Consider a discrete noiseless channel with elementary symbols S_1, \dots, S_n , where the transmission time of S_i is t_i , and assume that t_1, \dots, t_n are positive integers having no common divisor.

Let T be a positive integer and denote by $N(T)$ the number of different "words" $S_{\alpha_1}, \dots, S_{\alpha_k}$ whose transmission time $t_{\alpha_1} + \dots + t_{\alpha_k} = T$. Shannon [1] defines the capacity of the channel by

$$C = \lim_{T \rightarrow \infty} \frac{\log N(T)}{T}.$$

$N(T)$ satisfies an obvious difference equation, and most studies of the capacity of such a channel start from this equation.

It is of some interest to get rid of the assumption of commensurability of the times of transmission as well as of the finiteness of the set of elementary symbols.

In the following we shall show how the classical results can be generalized to this case.

Let S_1, \dots, S_n, \dots be a finite or countable set of elementary symbols and let $t_n > 0$ be the time of transmission of S_n ($n = 1, 2, \dots$). It is no longer assumed that the t_n are commensurable, but we make the assumption that in the countable case, $\lim_{n \rightarrow \infty} t_n = \infty$. Without loss of generality we may assume $0 < t_1 \leq t_2 \leq \dots$.

Let $\tau(W)$ be the time of transmission of the word $W = S_{\alpha_1}, \dots, S_{\alpha_n}$, that is,

$$\tau(W) = \sum_{i=1}^n t_{\alpha_i}$$

and denote by $G(T)$ the total number of different words W which can be transmitted within time T , i.e., which satisfy $\tau(W) \leq T$.

Received by the editors August 8, 1967.

¹ This paper is a part of the author's Ph.D. thesis written at the Hebrew University of Jerusalem under the supervision of Professor A. Dvoretzky.

The author is now on leave at the University of California, Berkeley.

Put

$$N(T) = G(T) - G(T - t_1);$$

$N(T)$ is the number of different words W which satisfy $T - t_1 < \tau(W) \leq T$. We shall show that²

$$(1) \quad C = \lim_{T \rightarrow \infty} \frac{\log N(T)}{T}$$

exists and relate this limit, "the capacity of the channel," to the Dirichlet series

$$(2) \quad f(x) = \sum_n \exp(-t_n x).$$

For the series (2) there exists an abscissa of convergence d such that (2) converges for $x > d$ and diverges for $x < d$.

(2) is continuous and is strictly decreasing on the real axis in its domain of convergence. Therefore either (i) there exists a unique $x_0 > d$ for which the series converges and $f(x_0) = 1$, or (ii) $f(x) < 1$ for all $x > d$.

We define

$$\begin{aligned} D &= x_0 && \text{in case (i),} \\ &= d && \text{in case (ii),} \end{aligned}$$

as the *abscissa of distribution* (a.o.d.) of the series (2) (or of the sequence $\{t_n\}$).

Our first result is the following

THEOREM 1. *The limit (1) always exists and we have*

$$(3) \quad C = D.$$

Then, applying Theorem 1 to generalize a result of J. R. Kinney [2, Theorem 2], the capacity is related to the Hausdorff dimension of a set of points associated with the message set of a code (§3, Theorem 2).

2. Proof of Theorem 1. We start with a lemma about the abscissa of distribution.

² It is immaterial what the basis of the logarithm is provided that \exp in (2) etc. is understood as the basis raised to the indicated power.

LEMMA 1. Let D be the a.o.d. of the sequence $\{t_n\}$, and let D_m be the a.o.d. of the finite sequence t_1, \dots, t_m ; then

$$(4) \quad \lim D_m = D,$$

and

$$(5) \quad f(D) = \sum_n \exp(-t_n D) \leq 1.$$

PROOF. D_m is an increasing sequence and $D_m \leq D$. Therefore, $\lim D_m = D' \leq D$. Our aim is to show that $D' = D$. We may thus assume $D' < \infty$. Then

$$1 = \sum_{n=1}^m \exp(-t_n \cdot D_m) \geq \sum_{n=1}^m \exp(-t_n D').$$

Since this holds for all m , we have

$$(6) \quad \sum_{n=1}^{\infty} \exp(-t_n D') \leq 1.$$

Therefore $D' \geq d$ (the a.o.d. of the sequence $\{t_n\}$), and in case (i), $D' \geq x_0$ because of the monotonicity of (2). Thus $D' \geq D$. This proves (4), and (5) follows from (6).

We now proceed to the proof of the theorem. First we show that the limit (1) exists. If W_1 is a word such that $T_1 - t_1 < \tau(W_1) \leq T_1$ and W_2 is such that $T_2 - t_1 < \tau(W_2) \leq T_2$, then either $T_1 + T_2 - t_1 < \tau(W_1 W_2) \leq T_1 + T_2$, or $T_1 + T_2 - 2t_1 < \tau(W_1 W_2) \leq T_1 + T_2 - t_1$, in which case $T_1 + T_2 - t_1 < \tau(W_1 W_2 S_1) \leq T_1 + T_2$. Therefore $N(T_1 + T_2) \geq N(T_1) \cdot N(T_2)$.

Hence

$$N(n\tau) \geq [N(T)]^n \quad \text{or} \quad \frac{\log N(nT)}{nT} \geq \frac{\log N(T)}{T}$$

and the existence of the limit easily follows.

Next we introduce the stochastic process $\{x_i\}$, ($i=1, 2, \dots$), where x_i are independent, identically distributed random variables assuming the values $R, S_1; S_2; \dots$ with probabilities

$$P[x_1 = S_n] = \exp(-t_n D), \quad (n = 1, 2, \dots),$$

$$P[x_1 = R] = 1 - f(D).$$

(The fictitious symbol R was introduced to assure that the probabilities add up to 1, even when D is defined according to case (ii).)

We shall associate the word $W = S_{\alpha_1}, \dots, S_{\alpha_n}$ with the event $[x_1 = S_{\alpha_1}, \dots, x_n = S_{\alpha_n}]$; its probability is $\exp(-\tau(W)D)$. Let $W_1, \dots, W_{N(T)}$ be the words satisfying

$$(7) \quad T - t_1 < \tau(W_i) \leq T.$$

Since for any two words W_i, W_j satisfying (7) we have $|\tau(W_i) - \tau(W_j)| < t_1$, no word in this set is a prefix of another word in the set. Hence they correspond to disjoint events and we have

$$\sum_{i=1}^{N(T)} \exp(-\tau(W_i) \cdot D) \leq 1.$$

Put $p_i = \exp(-\tau(W_i) \cdot D)$ and $q_i = 1/N(T)$ ($i = 1, \dots, N(T)$). The inequality³ [3, p. 13]

$$-\sum q_i \log q_i \leq -\sum q_i \log p_i$$

entails

$$\log N(T) \leq \frac{D}{N(T)} \sum_{i=1}^{N(T)} \tau(W_i) \leq D \cdot T.$$

This shows that $C \leq D$.

To establish the reverse inequality we first assume that case (i) holds, i.e., $d < D$. In this case, $f(x)$ is analytic at D and the series can be differentiated term by term. Then $\mu = \sum t_i \exp(-Dt_i)$, the mathematical expectation of the random variable $Y_i = \tau(x_i)$, is a positive finite number.

We put $Z_n = \sum_{i=1}^n Y_i$. Z_n is a sum of independent identically distributed random variables with finite variances (the second derivative of $f(x)$ is finite at $x = D$).

Z_n is either degenerate or asymptotically normal. Let A_n be the event $[Z_n \leq n\mu]$. In either case $P(A_n) > \frac{1}{3}$ for n large enough. Let

$$B_n = [Z_n/n > \mu - \epsilon];$$

then, by the law of large numbers, $P(B_n) \rightarrow 1$ for every $\epsilon > 0$. Therefore, if n is large enough,

³ The assumption $\sum p_i = 1$ in [3], can of course be replaced by $\sum p_i \leq 1$.

$$(8) \quad P(A_n \cap B_n) \geq P(A_n) + P(B_n) - 1 > \frac{1}{4}.$$

By adding an adequate number (≥ 0) of symbols S_1 every word (we thus abbreviate "word corresponding to an event") in A_n can be continued to one satisfying (7) with $T = n\mu$. Since the probability of every word in B_n is $\leq \exp(-nD(\mu - \epsilon))$, it follows, by (8), that the number of words in $A_n \cap B_n$ is $> \frac{1}{4} \exp(n \cdot D(\mu - \epsilon))$. A fortiori

$$\log N(n\mu)/n\mu > D - \epsilon/\mu - \log 4/n\mu,$$

which shows that $C \geq D$ in case (i).

If case (ii) holds, we consider the channel with the finite number of symbols S_1, \dots, S_m and its capacity $C_m = D_m$. (For finite $m \geq 2$, case (i) holds.) Let $N_m(T)$ be the function $N(T)$ for this channel; then obviously $N_m(T) \leq N(T)$, and therefore

$$C = \lim \frac{\log N(T)}{T} \geq D_m;$$

hence, by (4), $C \geq D$ and the proof of Theorem 1 is completed in both cases.

ADDED IN PROOF. Theorem 1 can be sharpened to take the following version:

THEOREM 1'. *Let t_1, t_2, \dots be the transmission times of a noiseless channel.*

(a) *If the numbers $\{t_1, t_2, \dots\}$ are commensurable, let δ_0 be the maximum positive number such that every $t_i = m_i \delta_0$ where m_i are integers. Then*

$$\lim \frac{N(T)}{\exp(n \cdot \delta_0)} = \frac{\delta_0}{\mu} \sum_{k=0}^{m_1-1} \exp(-k\delta_0 \cdot C).$$

(b) *If $\{t_1, t_2, \dots\}$ are noncommensurable, then*

$$\lim \frac{N(T)}{\exp(CT)} = \frac{t_1}{\mu} \int_0^{t_1} \exp(-t \cdot C) dt$$

where

$$\mu = \sum_{i=1}^{\infty} t_i \exp(-t_i \cdot C).$$

Theorem 1' can be proved by using the renewal theorem.

3. Codes and their Hausdorff dimensions. Let $A = \{a_1, \dots, a_n, \dots\}$ be a finite or countable set of symbols, and $\{q_1, \dots, q_n, \dots\}$ a set of probabilities associated with them:

$$q_i > 0 \quad (i = 1, 2, \dots), \quad \sum q_i = 1.$$

Consider the space $A^I = \Omega$ of all infinite sequences of symbols from A , and the product probability induced on it by the distribution $\{q_n\}$.

If $\omega \in \Omega$ is the sequence $\omega = a_{\alpha_1}, a_{\alpha_2}, \dots, a_{\alpha_n}, \dots$, we associate with ω a real number in the unit interval in the following way:

$$(9) \quad \omega \rightarrow Q_{\alpha_1} + q_{\alpha_1}Q_{\alpha_2} + \dots + q_{\alpha_1} \dots q_{\alpha_k}Q_{\alpha_{k+1}} + \dots,$$

where $Q_k = \sum_{i < k} q_i$. This correspondence is one to one except for a countable set of sequences ω . There is thus no harm in identifying ω with the real number associated with it (which we shall again denote by ω).

The correspondence (9) transfers the probability measure on Ω to the Lebesgue measure on $[0, 1]$.

Let there be given a code, i.e. a set (of words) $V = \{S_1 \dots S_n \dots\}$ where each S_k is a finite sequence of symbols $\{a_n\}$

$$S_k = a_{\alpha_1}, \dots, a_{\alpha_i(k)}.$$

We assume that V is a uniquely decodable code, that is, each $\omega \in \Omega$ can be decomposed into a sequence of words of V in at most one way. A sequence ω which is decomposable into a sequence of words is called a *message*. Let M be the set of all messages.

With each S_k we associate the *time of transmission*: $t_k = -\sum_{j=1}^{i(k)} \log q_{\alpha_j}$. Let C denote the capacity of the channel S_1, \dots, S_n, \dots with the transmission times t_1, \dots, t_n, \dots .

If $W = S_{\alpha_1} \dots S_{\alpha_k}$, put $\tau(W) = \sum_{i=1}^k t_{\alpha_i}$. The sequences ω which start with S form an interval $I(W)$ whose length is given by $|I(W)| = \exp(-\tau(W))$. W can also be considered as an event in the process $\{x_i\}$ defined in §2. Its probability is

$$P(W) = \exp(-\tau(W)C) = |I(W)|^C.$$

The γ -dimensional measure, $\Gamma(\gamma, E)$, of a set $E \subset [0; 1]$ is defined as follows. Let $\mathcal{@}(\epsilon)$ be a set of intervals $\{I_i\}$ such that $E \subset \cup I_i$; and $|I_i| < \epsilon$

$$\Gamma(\gamma, E) = \liminf_{\epsilon \rightarrow 0} (\sum_{\mathcal{@}(\epsilon)} |I_i|^\gamma).$$

The Hausdorff dimension $\beta(E)$ of E is defined by

$$\beta(E) = \inf\{\gamma \mid \Gamma(\gamma, E) = 0\}.$$

Let V be a finite code; then there exists a set $W'_1 \cdots W'_k$ (see [4]) such that every event in the process $\{x_i\}$ which contains M on the unit interval is of probability $\geq \sum_{i=1}^k P(W'_i)$.

LEMMA 2. *If V is a finite code, then*

$$\Gamma(C, M) \geq \sum_i |I(W'_i)|^c > 0.$$

PROOF. If V is finite, then M is closed. Let $\{I_j\}$ be a covering of M . Each I_j contains a longest interval of the form $I(W_j)$. Therefore,

$$\sum_j |I_j|^c > \sum_j |I(W_j)|^c = \sum_j P(W_j) \geq P\left(\bigcup_j (W_j)\right).$$

But $\{I_j\}$ is a covering of M and so is $\bigcup_j W_j$; hence

$$P\left(\bigcup_j W_j\right) \geq \sum P(W'_i) = \sum_i |I(W'_i)|^c.$$

THEOREM 2. *Let V be a code with a capacity C and message set M ; then $C = \beta(M)$.*

(When $A = \{0, 1\}$ and $q(0) = q(1) = \frac{1}{2}$ and V is finite, Theorem 2 reduces to the above mentioned result of Kinney.)

PROOF. If V is finite, then, by Lemma 2, $\beta(M) \geq C$. Assume V to be infinite, $V = \{S_1 \cdots S_n \cdots\}$ and denote by V_m the finite code, $V_m = \{S_1, \cdots, S_m\}$ and by C_m and M_m its capacity and message set respectively. Then $M \supset M_m$. Therefore $\beta(M) \geq \beta(M_m) \geq C_m$ and, by Lemma 1, $\beta(M) \geq C$.

Now, there are $N(T + t_1)$ intervals $I(W)$ for which $T < \tau(W) \leq T + t_1$, $N(T + 2t_1)$ intervals $I(W)$ for which $T + t_1 < \tau(W) \leq T + 2t_1$, etc. Every message must be contained in some $I(W)$. We cover M by those intervals and note that if $\gamma > C$,

$$\begin{aligned} & \sum_{n=1}^{\infty} N(T + nt_1) \{ \exp(-(T + (n - 1)t_1)) \}^\gamma \\ & \leq \sum_{n=1}^{\infty} \exp((T + nt_1)C) \exp(-\gamma(T + (n - 1)t_1)) \\ & = \exp(T(C - \gamma) + t_1\gamma) \sum_{n=1}^{\infty} \exp(nt_1(C - \gamma)) = K \exp(T(C - \gamma)) \end{aligned}$$

with K finite and not depending on T . Therefore, letting $T \rightarrow \infty$, we see that the γ -dimensional measure of M is zero. This shows that $\beta(M) \leq C$.

REFERENCES

1. C. Shannon, *A mathematical theory of communication*, Bell System Tech. J. **27** (1948), 379–423.
2. J. R. Kinney, *Singular functions associated with Markov chains*, Proc. Amer. Math. Soc. **9** (1958), 603–608.
3. A. Feinstein, *Foundations of information theory*, McGraw-Hill, New York, 1958.
4. M. Smorodinsky, *On infinite decodable codes*, Information and Control **11** (1968), 607–612.

TEL-AVIV UNIVERSITY