

SINGULAR FUNCTIONS ASSOCIATED WITH MARKOV CHAINS¹

JOHN R. KINNEY²

We suppose $\{x_i\}$, $i=1, 2, \dots$ to be a chain with a finite number of states, $0, 1, \dots, M-1$, and consider the random variable $X = \sum_{i=1}^{\infty} x_i M^{-i}$ and its associated distribution function $F(x) = \text{Prob} \{X < x\}$. We write $F(A) = \text{Prob} \{X \in A\} = \int_A dF(x)$. $F(A)$ is a completely additive probability measure on the Borel field of sets in $[0, 1]$ generated by sets of the form $\{F(x) < \alpha\}$. Harris [1] has shown that under very general conditions on the stationarity of the chain that $F(x)$ is a purely singular function and that $\phi(t)_{t \rightarrow 0} \rightarrow 0$ where $\phi(t)$ is the Fourier-Stieltjes transform $\phi(t) = \int e^{itz} dF(x)$. Wiener and Wintner [2] used the connection between the Lipschitz condition satisfied by $F(x)$ and the behavior of $\phi(t)$ to show that there are purely singular functions $F(x)$ for which $\phi(t)_{t \rightarrow \infty} = O(t^{-\alpha})$ for all $\alpha < 1/2$.

Salem [3] showed the connection between the Hausdorff measure of the set E on which $F(A)$ is concentrated and the behavior of $\phi(t)$ for large t . Although in our case $\phi(t)_{t \rightarrow \infty} \rightarrow 0$, the Lipschitz condition and the Hausdorff dimension of E still play a role. Namely, when the x_i form a stationary Markov chain, with a single ergodic class, they are the entropy, in the sense of Shannon [4], of the sequence $\{x_i\}$ considered as the sequence of states of a symbol-generating source.

The dimensional number $\beta(E)$ of a set $E \subset [0, 1]$ is defined as follows: If $\mu \geq \max_i |I_i|$, where $\{I_i\}$ is a set of intervals, and $E \subset \bigcup I_i$, we say $C_\mu = \bigcup I_i$ is a covering of E of norm μ . We let

$$\Gamma(\gamma, C_\mu, E) = \sum |I_i|^\gamma.$$

The γ -dimensional Hausdorff measure of E is

$$\Gamma(\gamma, E) = \lim_{\mu \rightarrow 0} \text{g.l.b.} \Gamma(\gamma, C_\mu, E)$$

where the greatest lower bound is taken over all coverings of norm μ . $\beta(E)$ is the number such that, for all $\epsilon > 0$,

$$\Gamma[\beta(E) - \epsilon, E] = \infty, \quad \Gamma[\beta(E) + \epsilon, E] = 0.$$

We suppose then the $\{x_i\}$ $i=1, 2, \dots$, to be a Markov chain,

Received by the editors August 12, 1957, and in revised form March 19, 1958.

¹ The research in this document was supported jointly by the Army, Navy and Air Force under contract with the Massachusetts Institute of Technology.

² Staff Member, Lincoln Laboratory, Massachusetts Institute of Technology.

with stationary transition probabilities with matrix $\|a_{ij}\|$, a single ergodic class, initial probabilities M^{-1} , and stationary probabilities b_i .

We let $\alpha = -\sum_{i,j} b_i a_{ij} \log_M a_{ij}$, a number proportional to the entropy in the sense of Shannon [4].

THEOREM 1. *There is a set $E \subset [0, 1]$ such that*

- (1) $F(E) = 1$,
- (2) $\beta(E) = \alpha$,
- (3) $x \in E$, $\epsilon > 0$ imply
 - (a) $\lim_{h \rightarrow 0} F(x - h, x + h) h^{-\alpha + \epsilon} \rightarrow 0$,
 - (b) $\lim_{h \rightarrow 0} F(x - h, x + h) h^{-\alpha - \epsilon} \rightarrow \infty$.

We let $I(n, k) = [kM^{-n}, (k+1)M^{-n}]$, $I(n, x)$ be that $I(n, k)$ which contains x . If $k = \sum_{i=0}^{n-1} k_{n-i} M^i$ is the expansion of k in the base M , $F[I(n, x)] = M^{-1} \prod_{i=2}^n a_{k_{i-1}, k_i}$. By the ergodic properties of the chain, the number of times $(k_{i-1}, k_i) = (j, k)$ is $b_j a_{jk} [n + O(1)]$ [4]. Hence

$$(1) \quad F[I(n, x)] = \prod_{ij} a_{ij}^{[n + O(1)] b_i a_{ij}} = M^{-[n + O(1)] \alpha}$$

except for a set of zero measure, which we delete to form E^* , $F(E^*) = 1$. On E^* , (3.b) is satisfied.

Although (1) immediately implies (3b), its lack of symmetry does not allow us to conclude that (3a) holds for all points of E^* . We eliminate the points affected by this lack of symmetry to form a set $E \subset E^*$ for which (3a) holds and $F(E) = 1$. We proceed as follows: We let $C(\epsilon)$ be those x for which

$$(2) \quad F[I(n, x)] \geq |I(n, x)|^{\alpha - \epsilon}$$

for an infinite number of n . Since $C(\epsilon) \subset E^*$, $F[C(\epsilon)] = 0$. We choose the covering $C_n(\epsilon)$ of $C(\epsilon)$ by assigning to each $x \in C(\epsilon)$ that interval $I(m, x)$ for which (2) holds for the smallest $m > n$. Thus, the interval assigned to x in the n th covering includes the interval assigned to x in any higher covering. Hence, $C_n(\epsilon) \supset C_{n+s}(\epsilon)$ in the sense of set inclusion. Also, $C_n(\epsilon) \downarrow C(\epsilon)$. From the complete additivity of $F(A)$, $\lim_{n \rightarrow \infty} F[C_n(\epsilon)] = 0$. For each $C_n(\epsilon)$ we construct a $C_n(r, \epsilon)$ and $C_n(l, \epsilon)$ as follows: If $I(s, k) \in C_n(\epsilon)$ we assign $I(s, k+1)$ to $C_n(r, \epsilon)$, $I(s, k-1)$ to $C_n(l, \epsilon)$. We let $D_n(\epsilon) = C_n(\epsilon) \cup C_n(r, \epsilon) \cup C_n(l, \epsilon)$. If $I(s, k+1) \in C_n(r, \epsilon)$ and $I(s, k+1) \subset C_n(\epsilon)$, (2) does not hold and hence $F[I(s, k+1)] < |I(n, k+1)|^{\alpha - \epsilon} = |I(n, k)|^{\alpha - \epsilon} \leq F[I(s, k)]$. Hence,

$$F[C_n(r, \epsilon)] = F[C_n(r, \epsilon) \cap C_n(\epsilon)] + F[C_n(r, \epsilon) \cap {}^c C_n(\epsilon)] \leq 2F[C_n(\epsilon)].$$

Similarly, $F[C_n(l, \epsilon)] \leq 2F[C_n(\epsilon)]$. So $F[D_n(\epsilon)] \leq F[C_n(\epsilon)] + F[C_n(r, \epsilon)] + F[C_n(l, \epsilon)] \leq 5F[C_n(\epsilon)]$. We note that $D_n(\epsilon) > D_{n+s}(\epsilon)$ and let $D(\epsilon) = \lim_{n \rightarrow \infty} D_n(\epsilon)$. $F[D(\epsilon)] \leq 5 \lim F[C_n(\epsilon)] = 0$. We obtain E by deleting from E^* all points belonging to $D(\epsilon_n)$ for a sequence $\epsilon_n \downarrow 0$. Having eliminated only a countable number of null sets, we still have $F[E] = 1$. However, for $x \in E$, for any ϵ , we can choose m and k so large that if $x \in I(m, j)$, $F[I(m, j+i)] < M^{-m(\alpha-\epsilon)}$ for $i = -1, 0, 1$. Hence, for $h < M^{-n}$, $F(x-h, x+h) < 3M^{-n(\alpha-\epsilon)}$ from which we may deduce (3a).

We cover E by a set $C_n = \{I_j\}$ of norm M^{-n} , by assigning for each x the $I(m, x)$ for which $F[I(m, x)] > |I(m, x)|^{\alpha+\epsilon}$ for the smallest $m \geq n$. The I_j are disjoint, since none is included in another and they cannot overlap. Since $E \subset \bigcup_j I_j$, $1 \geq \sum F(I_j) > \sum |I_j|^{\alpha+\epsilon}$. Hence $\Gamma(\alpha+\epsilon, E) \leq 1$ for all $\epsilon > 0$, so $\beta(E) \leq \alpha$.

By choosing h_0 sufficiently small, we can, for any fixed ϵ , find a subset $A(h_0)$ of E such that $F[A(h_0)] > 0.5$, and, for $x \in A(h_0)$, $h \leq h_0$ we have $F(x-h, x+h) < (2h)^{\alpha-\epsilon}$. We choose any covering $C_{h_0} = \bigcup I_j$. To each I_i we choose a point $x_i \in A(h_0)$ and take I'_i to be the smallest interval containing I_i symmetric about x_i . We note that $|I'_i| < 2|I_i|$. We then have

$$\sum (2|I_i|)^{\alpha-\epsilon} \geq \sum |I'_i|^{\alpha-\epsilon} > \sum F(I'_i) \geq F[A(h_0)] > 0.5.$$

So, for $\epsilon > 0$, $\Gamma[\alpha-\epsilon, C_{h_0}, A(h_0)] > 0.5$. Since C_{h_0} was an arbitrary covering of $A(h_0)$ of norm h_0 , $\Gamma[\alpha-\epsilon, A(h_0)] > 0.5$. Hence $\beta[A(h_0)] \geq \alpha$. Since $E \supset A(h_0)$, $\beta(E) \geq \alpha$. Hence (2) holds. This establishes our theorem.

An example: H. G. Eggleston [5] has shown that the set

$$S = \left\{ x \left| \lim_{n \rightarrow \infty} \sum_{i=1}^n x_i/n \leq a \leq (M-1)/2 \right. \right\} \text{ has } \beta(S) = \alpha,$$

where $M^\alpha = [kr^a]^{-1}$, $k = (1-r)/(1-r^M)$, for r the positive real root of $\sum_{i=0}^{M-1} (i-a)r^i = 0$. He proves this by showing $\alpha \geq \beta(S^*) \geq \beta(S_*) \geq \alpha$, where S^* , S_* are obtained by replacing the limit in the definition of S by limit inferior and limit superior and using rather sophisticated methods to obtain his coverings.

We give a proof using the point of view of information theory. We let $x_j = i$, Prob p_i , $i = 0, \dots, M-1$, independently of j , subject to the restrictions $E(x_j) = \sum i p_i = a$, $\sum p_i = 1$. We vary the probabilities p_i so as to maximize the entropy $H(X) = -\sum p_i \log_M p_i$, of the sequence $\{x_i\}$ considered as a source, subject to the given side conditions. We find the maximum to be given by kr^i . Our restrictions imply $k \sum_{i=0}^{M-1} r^i = 1$, $k \sum_{i=0}^{M-1} ir^i = a$. This yields $k = [\sum_{i=0}^{M-1} r^i]^{-1}$

$= 1 - r/1 - r^M$. $\sum_{i=0}^{M-1} (i-a)r^i = 0$. We have also

$$\begin{aligned} H(X) &= - \sum_0^{M-1} kr^i \log_M kr^i = - \sum kr^i \log_M k - \sum kir^i \log_M r \\ &= - \log_M k - a \log_M r = - \log_M kr^a, \end{aligned}$$

so $H(X) = \alpha$. However, our measure $F(A)$ imposed on the interval $[0, 1]$ by this assignment of probabilities, applied to intervals $I(n, x)$ corresponding to the partial sums of x , gives

$$F[I(n, x)] = k^n r^{\sum_{i=1}^n x_i}$$

so the set S_0 which corresponds to the set E of Theorem 1 is

$$S_0 = \left\{ x \left| \lim_{n \rightarrow \infty} \sum_{i=1}^n x_i/n = a \right. \right\}, \text{ and, by Theorem 1, } \beta(S_0) = \alpha.$$

We note that, in terms of $F(A)$, S^* is the set

$$\begin{aligned} S^* &= \{ x \mid F[I(n, x)] \\ &= k^n r^{\sum_{i=1}^n x_i} \geq M^{-n\alpha} = |I(n, x)|^\alpha \text{ for infinitely many } n \}. \end{aligned}$$

We construct the covering C_N of S^* of norm M^{-N} by assigning to each $x \in S^*$ that interval $I(m, x)$ for which, for the smallest $m > N$, $F[I(m, x)] \geq M^{-m\alpha} = |I(m, x)|^\alpha$. The $I(m, x)$ are disjoint, so

$$1 \geq F(C_N) = \sum F[I(m, x)] > \sum |I(m, x)|^\alpha = \Gamma(\alpha, C_N, S^*).$$

Hence, $\Gamma(\alpha, S^*) \leq 1$, so $\beta(S^*) \leq \alpha$. Since $S^* \supset S \supset S_0$, $\alpha \geq \beta(S^*) \geq \beta(S) \geq \beta(S_0) = \alpha$. So $\beta(S) = \alpha$.

We suppose ourselves restricted to sending two-state $[0, 1]$ pulses in the transmission of messages. We can then take symbols $\{y_i, i=1, \dots, \mu\}$ to be binary numbers of lengths respectively t_i . A message will be a sequence of the y_i 's. Z_j , the j th symbol, will be a function whose range are the y 's. An infinite message will be made to correspond to a point y in $[0, 1]$ by the dyadic expansion

$$Y = \sum Z_k 2^{-\sum_{i=1}^k t_i}.$$

The set of such points we will call the message set, Q . We assume restrictions on the choice of the y_i to have been made so that two different messages cannot correspond to the same point. This will imply that different finite messages of the same length will correspond to disjoint dyadic intervals of the line.

Each finite message of length n will correspond to an interval in $[0, 1]$ from the n th dyadic net. We suppose $t_j = M$ to be the length of the longest symbol. We let $N(t)$ be the number of finite messages of length t . It is clear that we can cover the message set Q by $N(t)$ intervals of length 2^{-t} , $N(t+1)$ intervals of length $2^{-(t+1)}$, \dots , $N(t+M)$ intervals of length $2^{-(t+M)}$. These may, of course, involve overlapping; however, we have an approximation

$$\Gamma[\gamma, C(2^{-t}), Q] \leq \sum_0^M N(t+i)2^{-(t+i)\gamma}$$

to the Hausdorff γ -dimensional measure of the set Q . We let $k(j)$ be the number of symbols of length j . As Shannon shows [1], $N(t)$ satisfies the difference equation

$$N(t) = \sum_1^M k(j)N(t-j)$$

so that $N(t)$ is asymptotically approximated by $k\lambda^t$ where λ is the largest root of

$$(3) \quad \lambda^t = \sum_1^M k(j)\lambda^{t-j}.$$

Hence,

$$\Gamma[\gamma, C(2^{-t}), Q] \leq K\lambda^t 2^{-\gamma t}$$

for large t . For $\gamma > \log_2 \lambda$,

$$\Gamma[\gamma, C(2^{-t}), Q] \xrightarrow{t \rightarrow \infty} 0;$$

so $\beta(Q) \leq \log_2 \lambda$.

We propose to send y_j independently of what has been sent before with probability $p_j = \lambda^{-t_j}$. By [3], $\sum p_j = 1$. On each of our intervals of length 2^{-t} corresponding to messages of length exactly t , the increase of $F(x)$, the distribution imposed on $[0, 1]$ by our mapping, and by our choice of message distribution will be λ^{-t} . Hence, on Q , our function $F(x)$ satisfies a Lipschitz condition $\text{Lip } \log_2 \lambda$, and no weaker Lipschitz condition. Since Q consists of the points of increase of $F(x)$, Q is closed. Hence we may cite the theorem of J. Gillis [6], to the effect that if $F(x)$ continuous and monotone takes its increase on a closed set E , satisfies a Lipschitz condition of no smaller order than $\text{Lip } \delta$, then the Hausdorff δ -dimensional measure of E is positive. Hence $\beta(Q) \geq \log_2 \lambda$.

Thus, if we let C denote the capacity of a channel in the sense of Shannon, we have, for the case described above.

THEOREM 2. $\beta(Q) = C$.

REFERENCES

1. T. E. Harris, *On chains of infinite order*, Pacific J. Math. vol. 5 (1955) pp. 707–724.
2. N. Wiener and A. Wintner, *Fourier-Stieltjes transform and singular finite convolutions*, Amer. J. Math. vol. 60 (1938) pp. 513–522.
3. R. Salem, *On singular monotone functions whose spectrum has a given Hausdorff dimension*, Ark. Mat. vol. 1 (1952) pp. 353–356.
4. C. Shannon, *A mathematical theory of communication*, Bell System Technical Journal vol. 27 (1948) pp. 379–423.
5. H. G. Eggleston, *Sets of fractional dimensions which occur in some problems of number theory*, Proc. London Math. Soc. vol. 54 (1953) pp. 42–93.
6. J. Gillis, *A note on a theorem of Myrberg*, Proc. Cambridge Philos. Soc. vol. 33 (1937) pp. 419–424.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY