DIMENSION AND ENTROPY FOR A CLASS OF STOCHASTIC PROCESSES

by

MATS RUDEMO[1]

Introduction

In the following we shall give a definition of dimension and entropy for a class of stochastic processes with the property that the sample functions are step functions with probability one.

Dimension and entropy together give a measure of the uncertainty associated with a random variable, or in our case a stochastic process, see [1] and [9].

In § 1 we give some examples of stochastic processes, whose sample functions a. s.[2] are step functions — such a process is called a purely discontinuous process, a PDP.

Some known properties of the dimension and the entropy for random variables and vectors, needed in the following, are stated in § 2.

In § 3 we define dimension and entropy for a class of PDP : s regarded on a finite interval $(0, T)$. As an example the dimension and entropy of a Poisson process are calculated. The asymptotic T-dependence of the dimension is studied in § 4 for ergodic Markov chains with a finite state space and for renewal processes.

For vector processes the corresponding definitions are made in § 5. An example with Poisson processes is discussed.

Finally in § 6 we give a method of approximating by PDP : s stochastic processes whose sample functions are a. s. continuous. Especially the Brownian motion is discussed and the dimension of the approximating PDP is studied.

§ 1. Purely discontinuous processes

By a stochastic process we shall mean a one-parameter family of random variables (measurable functions on a probability space), see [6]. The measure of the probability space will always be assumed to be complete. The parameter will be called time. We shall only consider processes on the interval $[0, \infty)$. The stochastic process is then written $\{X(t): t \in [0, \infty)\}$ or $\{X(t, \omega): t \in [0, \infty), \omega \in \Omega\}$. Here Ω is the space of elementary events. The sample function corresponding to ω is $X(\cdot, \omega)$.

A stochastic process is called a purely discontinuous process, a PDP, if the sample functions a. s. are step functions, with a finite number of jumps on every finite time interval. The time points corresponding to jumps will be called points of jump.

---

[1] Göteborg.

[2] a. s. is used as short for "almost surely", i.e. „with probability one".

If $X(t) = (X_1(t), \ldots, X_r(t))$, $t \in [0, \infty)$, where $\{X_i(t) \colon t \in [0, \infty)\}$, $i = 1, \ldots, r$, are PDP : s, then $\{X(t) \colon t \in [0, \infty)\}$ is called a purely discontinuous vector process.

General conditions for a process to be a PDP have been given by M. Fisz, see [7].

Important types of PDP : s are found among Markov processes and renewal processes.

A sufficient condition for a separable Markov process $\{X(t) \colon t \in [0, \infty)\}$ with stationary transition function $p(t, \xi, A) = \mathbf{P}\{X(s+t) \in A \mid X(s) = \xi\}$ to be a PDP is that

$$\lim_{t \to 0} p(t, \xi, \{\xi\}) = 1 \tag{1}$$

uniformly in ξ, see [6], theorem VI. 2.4.

By a renewal process, see [10], we shall mean a process $\{X(t) : t \in [0, \infty)\}$,

$$X(t) = \max_{\tau_n \leqq t} n\,,$$

where $\tau_1 \leqq \tau_2 \leqq \tau_3 \leqq \ldots$ are random variables taking values on the interval $[0, \infty)$ and such that $\tau_1, \tau_2 - \tau_1, \tau_3 - \tau_2, \ldots$ are independent and equidistributed. Let $\mathbf{P}\{\tau_1 = 0\} < 1$. Then $\{X(t) : t \in [0, \infty)\}$ is a PDP. For let $\alpha > 0$ be chosen so that $\mathbf{P}\{\tau_1 > \alpha\} > 0$ and let A_n be the event that $\tau_n - \tau_{n-1} > \alpha$. In order to show that $\{X(t) : t \in [0, \infty)\}$ is a PDP we have only to prove that $\mathbf{P}\{\lim_{n \to \infty} \tau_n = \infty\} = 1$. A sufficient condition for this relation is that $\mathbf{P}\{\limsup_{n \to \infty} A_n\} = 1$. But this equality follows from the Borel—Cantelli lemma since

$$\sum_{n=2}^{\infty} \mathbf{P}\{A_n\} = \mathbf{P}\{\tau_1 > \alpha\} \sum_{n=2}^{\infty} 1 = \infty\,.$$

§ 2. Information theoretic background

If ξ is a discrete random variable taking the value x_k, $k = 1, 2, \ldots$, with the probability p_k, $k = 1, 2, \ldots$, the entropy $H_0(\xi)$ of ξ is defined by

$$H_0(\xi) = -\sum_{k=1}^{\infty} p_k \log p_k\,,$$

if the series is convergent. The base of the logarithm is here arbitrary.

Let ξ be a real random variable, and put $\xi^{(n)} = \frac{1}{n}[n\xi]$, where $[x]$ is the integral part of x. If $H_0(\xi^{(1)}) < \infty$ we put, according to A. Rényi, see [9],

$$d = d(\xi) = \lim_{n \to \infty} \frac{H_0(\xi^{(n)})}{\log n}\,,$$

provided that the limit exists, and call d the dimension of ξ. If further the limit

$$H_d(\xi) = \lim_{n \to \infty} [H_0(\xi^{(n)}) - d \log n]$$

exists, it is called the d-dimensional entropy of ξ.

A random variable ξ is said to be the mixture with weights $\{q_k\}$, $q_k \geqq 0$, $\sum_k q_k = 1$, of the variables $\{\xi_k\}$ if the distribution function of ξ is $\sum_k q_k F_k$, where F_k is the distribution function of ξ_k, If N is a random variable, taking the values $\{k\}$ with probabilities $\{q_k\}$ and if the conditional distribution of ξ, given that $N = k$ is equal to the distribution of ξ_k, than ξ is the mixture of $\{\xi_k\}$ with weights $\{q_k\}$.

If ξ is the mixture of ξ_k, $k = 1, \ldots m$, with weights q_k, $k = 1, \ldots m$, then

$$\sum_k q_k H_0(\xi_k^{(n)}) \leq H_0(\xi^{(n)}) \leq \sum_k q_k H_0(\xi_k^{(n)}) - \sum_k q_k \log q_k .$$

The first inequality follows from Jensens inequality applied to the convex function $x \log x$, and the second inequality follows from the fact that for discrete variables the entropy of a mixture is always less than or equal to the sum of the weighted average of the entropies of the components in the mixture and the entropy of the mixing distribution, in this case $\{q_k\}$, see [1]. Dividing with $\log n$ and letting n pass to infinity we obtain

$$d(\xi) = \sum_{k=1}^{m} q_k d(\xi_k) . \tag{2}$$

If the components in the mixture have pairwise orthogonal[3] and elementary[4] distributions in the sense of [1] we further have, see [1],

$$H_d(\xi) = \sum_{k=1}^{m} q_k H_{d_k}(\xi_k) - \sum_{k=1}^{m} q_k \log q_k , \tag{3}$$

where $d_k = d(\xi_k)$.

Let ζ be a random vector, $\zeta = (\xi_1, \ldots \xi_r)$. We put

$$\zeta^{(n)} = \left(\frac{[n\,\xi_1]}{n}, \ldots, \frac{[n\,\xi_r]}{n}\right).$$

If $H_0(\zeta^{(1)}) < \infty$ the dimension of ζ is defined by

$$d = d(\zeta) = \lim_{n \to \infty} \frac{H_0(\zeta^{(n)})}{\log n} ,$$

if the limit exists. If further the limit

$$H_d(\zeta) = \lim_{n \to \infty} [H_0(\zeta^{(n)}) - d \log n]$$

exists, it is called the d-dimensional entropy of ζ. If ζ has an absolutely continuous distribution and if $H_0(\zeta^{(1)}) < \infty$ then according to [9]

$$d(\zeta) = r , \tag{4}$$

and

$$H_r(\zeta) = - \int f(x) \log f(x) \, dx , \tag{5}$$

see [9] and [4]. Here $f(x)$ is the density function of ζ and the integral is taken over the entire r-space.

---

[3] ξ and η have orthogonal distributions if there exist two disjoint Borel measurable subsets E and F of the real line, such that $P\{\xi \in E\} = P\{\eta \in F\} = 1$.

[4] The distribution of ξ is called elementary if it is the mixture of a finite discrete distribution and an absolutely continuous distribution whose density function is piecewise continuous, only has discontinuities of the first kind and is zero outside a finite interval.

§ 3. Dimension and entropy for purely discontinuous processes

Let $\{X(t, \omega) : t \in [0, \infty), \omega \in \Omega\}$ be a PDP. We shall regard the process on the interval $0 < t < T$. Let $N(T) = N(T, \omega)$ be the number of jumps on this interval. We will first show that $N(T, \cdot)$ is a random variable.

Lemma 1. *$N(T, \cdot)$ is a measurable function.*

Proof. Let $\{r_i\}$ be a countable set of points dense in the interval $(0, T)$. As shown below there then exists a sequence $\{f_n\}$, where f_n for each n is a Borel measurable function from the n-space to the set of nonnegative integers $< n$, such that a. s.

$$\lim_{n\to\infty} f_n(X(r_1, \omega), \ldots, X(r_n, \omega)) = N(T, \omega) .$$

The lemma then follows from the fact that the limit of an a.e. convergent sequence of measurable functions is measurable.

The functions $\{f_n\}$ can be chosen in the following way. For fixed n, let i' and i'' be indices such that $r_{i'}$ and $r_{i''}$ are the left and right neighbours of r_i among $r_1, \ldots, r_n$. For the least and the largest of $r_1, \ldots, r_n$ only one neighbour exists. If $x_1, \ldots, x_n$ are given real numbers, we let $A(x_1, \ldots, x_n)$ be the set of those $x_i : s$ such that $x_i = x_{i'}$ or $x_i = x_{i''}$ and let $m(x_1, \ldots, x_n)$ be the number of different real numbers in $A(x_1, \ldots, x_n)$. Then we can put

$$f_n(x_1, \ldots, x_n) = m(x_1, \ldots, x_n) - 1 .$$

If the number of jumps of $X(\cdot, \omega)$ is finite on the interval $(0, T)$, an event which has probability one, then for fixed ω

$$f_n(X(r_1, \omega), \ldots, X(r_n, \omega)) = N(T, \omega)$$

for sufficiently large n.

QED.

Note. The slightly complicated structure of the $f_n : s$ in the proof is caused by the possibility that some of the $r_i : s$ may coincide with points of jump with probability greater than zero. If this were not the case, or if the sample functions a.s. were right or left continuous, we could have chosen $f_n(x_1, \ldots, x_n)$ to be equal to the number of different numbers among $x_1, \ldots, x_n$ minus one.

Let $t_1(\omega), \ldots t_{N(T)}(\omega)$ be the points of jump on the interval $(0, T)$ ordered such that $0 < t_1 < \ldots < t_{N(T)} < T$. For simplicity we define $t_k(\omega) = T$ if $k > N(T, \omega)$. Then $\{t_k\}$ become random variables according to

Lemma 2. *t_k, $k = 1, 2, \ldots$, are measurable functions.*

(Sketch of) **Proof.** As the proof is rather similar to the proof of lemma 1, we will only indicate how it can be carried through. Let $\{r_i\}$ be given as in the proof of lemma 1. Given $X(r_i, \omega)$, $i = 1, \ldots, n$, it is possible to choose $t_k^{(n)}(\omega)$, $k = 1, 2, \ldots$, among the numbers $r_1, \ldots, r_n$ such that a.s.

$$\lim_{n\to\infty} t_k^{(n)}(\omega) = t_k(\omega) .$$

The number $t_k^{(n)}(\omega)$ can further be chosen in such a way that the choice only depends on the mutual size relations among $X(r_i, \omega)$, $i = 1, \ldots, n$. Then $\{t_k^{(n)}(\cdot)\}$ and therefore also $\{t_k(\cdot)\}$ become measurable functions. QED.

A similar proof, which is completely left out, is also valid for the following

Lemma 3. *$X(t_k+0, \cdot)$, $k = 1, 2, \ldots$ are measurable functions.*

From now on we will drop ω in the notations for random variables and stochastic processes.

If we disregard $X(t_1)$, $X(t_2)$, ..., a sample function of our process can, on the interval $(0, T)$, a.s. be described by the $(2N(T)+1)$-tuple

$$\xi(T) = \big(t_1, \ldots, t_{N(T)}, X(+0), X(t_1+0), \ldots, X(t_{N(T)}+0)\big).$$

In case $X(t_1), X(t_2), \ldots$, are essential for the description of the stochastic process the procedure has to be modified. However, if we know that the sample functions a.s. are continuous from the left, or if we know that they a.s. are continuous from the right, then $\xi(T)$ a.s. determines the sample functions on the interval $(0, T)$.

Now $\xi(T)$ can be regarded as the mixture with weights

(6) $$q_n(T) = P\{N(T) = n\},\ n = 0, 1, \ldots,$$

of the variables

(7) $$\xi_n(T) = \big(t_1, \ldots, t_n, X(+0), X(t_1+0), \ldots, X(t_n+0)\big),\ n = 0, 1, \ldots,$$

where the $(2n+1)$-tuple $\xi_n(T)$ has the conditional distribution of the points of jump and the corresponding X-values, given that $N(T) = n$. If $q_n(T) > 0$ this conditional distribution exists according to lemmata 1, 2, and 3. From the definition of a PDP it follows that

$$\sum_{n=0}^{\infty} q_n(T) = 1.$$

If $\xi_n(T)$ has a dimension and an entropy we denote them

(8) $$d_n(T) = d\big(\xi_n(T)\big)$$

and

(9) $$H_{d_n}(T) = H_{d_n(T)}\big(\xi_n(T)\big).$$

As $\xi_n(T)$ and $\xi_m(T)$ for $n \neq m$ take values in spaces of different dimensions it is natural to regard the corresponding distributions as orthogonal. The relations (2) and (3) then motivate the following definitions.

Definition 1. If $\xi_n(T)$ for each n has the dimension $d_n(T)$, the dimension of the process $\{X(t) : t \in [0, \infty)\}$ regarded on the interval $(0, T)$ is defined by

(10) $$d^T = d^T(X) = \sum_{n=0}^{\infty} q_n(T)\, d_n(T),$$

if the series is convergent.

Definition 2. If $\xi_n(T)$ for each n has the dimension $d_n(T)$ and the entropy $H_{d_n}(T)$ and if the series (10) converges, the d^T-dimensional entropy of the process $\{X(t) : t \in [0, \infty)\}$ regarded on the interval $(0, T)$ is defined by

(11) $$H^T = H^T(X) = \sum_{n=0}^{\infty} q_n(T)\, H_{d_n}(T) - \sum_{n=0}^{\infty} q_n(T) \log q_n(T),$$

if the series are convergent.

Note. In these definitions we treat the points of jump and the corresponding process values in the same way (see the definition of $\xi_n(T)$). It may in some instances be more natural to have different "scales" of uncertainty for time values and process values.

For later application we note that if

$$d_n(T) = an + b, n = 0, 1, \ldots,$$

then

(12) $$d^T = am(T) + b \; ,$$

where $m(T)$ is the average number of jumps on the interval $(0, T)$,

$$m(T) = \sum_{n=0}^{\infty} n \, q_n(T) \, .$$

Example. Let $\{X(t) : t \in [0, \infty)\}$ be a Poisson process with parameter λ, i.e. $X(0) = 0$ and

$$\mathbf{P}\{X(t) = n \mid X(\tau) = m\} = \frac{[\lambda(t-\tau)]^{n-m}}{(n-m)\,!} e^{-\lambda(t-\tau)}, n \geqq m, t > \tau \, .$$

Let further the process be separable. As (1) is satisfied uniformly for $\xi = 0, 1, \ldots$, the process is a PDP. We will now calculate its dimension and entropy.

Regard the process on the interval $(0, T)$. The number of jumps on this interval is a.s. $X(T)$. We then have to determine the dimension and entropy of the random vector $(t_1, \ldots, t_n)$ given that $X(T) = n$. As before, $t_1, \ldots, t_n$ are the points of jump ordered in such a way that $0 < t_1 < \ldots t_n < T$. Since the relation $X(t_k + 0) = k$ holds a.s. we need not take into account the X-values at the points of jump when calculating the dimension and the entropy. The distribution of $(t_1, \ldots, t_n)$ is uniform in that region V of the n-space where $0 < x_1 < \ldots < x_n < T$, $\{x_i\}$ being the coordinates, see [11] p 86. The region V has the volume $\frac{1}{n!} T^n$. According to (4) and (5) we have

$$d(t_1, \ldots, t_n) = n$$

and

$$H_n(t_1, \ldots, t_n) = -\int_V \frac{n!}{T^n} \log \frac{n!}{T^n} \, dx_1 \ldots dx_n = \log \frac{T^n}{n!} \, .$$

The dimension of the process is

$$d^T = \sum_{n=0}^{\infty} n \, \mathbf{P}\{X(T) = n\} = \lambda \, T.$$

The corresponding entropy is

$$H^T = \sum_{n=0}^{\infty} \frac{(\lambda \, T)^n}{n\,!} e^{-\lambda T} \log \frac{T^n}{n!} - \sum_{n=0}^{\infty} \frac{(\lambda \, T)^n}{n\,!} e^{-\lambda T} \log \left[\frac{(\lambda \, T)^n}{n\,!} e^{-\lambda T} \right] = \left(\lambda \log \frac{e}{\lambda} \right) T.$$

Let $\{X_i(t) : t \in [0, \infty)\}$, $i = 1, \ldots, m$, be stochastic processes. We say that $\{X(t) : t \in [0, \infty)\}$ is the mixture with weights p_i, $i = 1, \ldots, m$, $p_i \geqq$

$\geq 0, \Sigma p_i = 1$, of the processes $\{X_i(t) : t \in [0, \infty)\}$, $i = 1, \ldots, m$, if there is a random variable N, taking the value i with the probability p_i, and such that if $N = i$, then $X(t) = X_i(t)$ for all $t \in [0, \infty)$.

Theorem 1. *Let $\{X(t) : t \in [0, \infty)\}$ be the mixture with weights $p_i, i = 1, \ldots m$, of the PDP: s $\{X_i(t) : t \in [0, \infty)\}$, $i = 1, \ldots m$. Then $\{X(t) : t \in [0, \infty)\}$ is a PDP and*

$$d^T(X) = \sum_{i=1}^{m} p_i \, d^T(X_i) ,$$

provided that the right member exists.

Proof. It is obvious that $\{X(t) : t \in [0, \infty)\}$ is a PDP. Let $q_n(T)$, $\xi_n(T)$ and $d_n(T)$ be defined according to (6), (7) and (8) and let $q_n^{(i)}(T)$, $\xi_n^{(i)}(T)$ and $d_n^{(i)}(T)$ be the corresponding quantities for $\{X_i(t) : t \in [0, \infty)\}$, $i = 1, \ldots, m$. Then $\xi_n(T)$ is the mixture with weights

$$r_i = \frac{p_i \, q_n^{(i)}(T)}{q_n(T)} , \quad i = 1, \ldots m,$$

of $\xi_n^{(i)}(T)$, $i = 1, \ldots m$. Since

$$q_n(T) = \sum_{i=1}^{m} p_i \, q_n^{(i)}(T) ,$$

we have $\Sigma r_i = 1$. From the relation (2), or rather from the corresponding relation for random vectors, we have

$$d_n(T) = \sum_{i=1}^{m} r_i \, d_n^{(i)}(T) .$$

Therefore

$$d^T(X) = \sum_n q_n(T) \, d_n(T) = \sum_i p_i \sum_n q_n^{(i)}(T) \, d_n^{(i)}(T) = \sum_i p_i \, d^T(X_i) .$$

QED

§ 4. Dimension for Markov chains and renewal processes.

Let $\{X(t) : t \in [0, \infty)\}$ be a Markov chain with stationary transition function and finite state space, which, we assume, consists of the integers $1, \ldots N$. Put

$$p_{ij}(t) = \mathbf{P}\{X(t + s) = j \,|\, X(s) = i\} , \quad i, j = 1, \ldots N.$$

As $t \to \infty$, $p_{ij}(t)$ tends to a limit, see [6]. If this limit is independent of i,

$$\lim_{t \to \infty} p_{ij}(t) = P_j ,$$

the chain is said to be ergodic. Then irrespective of the initial distribution

$$\lim_{t \to \infty} \mathbf{P}\{X(t) = j\} = P_j ,$$

and if the initial distribution is equal to $\{P_j\}$, the stationary distribution, then for all t

$$\mathbf{P}\{X(t) = j\} = P_j . \tag{13}$$

Suppose that $p_{ij}(t) \to \delta_{ij}$ as $t \to 0$, i.e. the condition (1) is satisfied. If $\{X(t) : t \in \in [0, \infty)\}$ is separable then it is a PDP. Further, see [6], the transition intensities

$$q_i = \lim_{t \to 0} \frac{1 - p_{ii}(t)}{t}$$

exist.

Theorem 2. *If $\{X(t) : t \in [0, \infty)\}$ is a separable, ergodic Markov chain with finite state space and stationary transition function, and if the transition probabilities $p_{ij}(t)$ tend to δ_{ij} as $t \to 0$, then the dimension $d^T(X)$ exists for every $T > 0$ and is a differentiable function of T such that*

$$\lim_{T \to \infty} \frac{d}{dT} d^T(X) = \sum_i P_i q_i .$$

Here $\{P_i\}$ is the stationary distribution and $\{q_i\}$ are the transition intensities. If the initial distribution is equal to $\{P_i\}$, then

$$d^T(X) = (\sum_i P_i q_i)\, T$$

for all $T > 0$.

Note. Chintschin (1953) showed for a finite stationary and ergodic Markov chain with discrete parameter that the r-step entropy is r times the onestep entropy. The later part of theorem 2 gives a corresponding property for Markov chains with continuous parameter.

Proof. If ξ and η are random variables or vectors, whose dimensions exist,

$$\max\big(d(\xi), d(\eta)\big) \leq d(\xi, \eta) \leq d(\xi) + d(\eta) .$$

This follows easily from the corresponding property of the entropy. Therefore

$$d(t_1, \ldots t_n) \leq d\big(\xi_n(T)\big) \leq d(t_1, \ldots t_n) + \sum_{k=0}^{n} d\big(X(t_k + 0)\big) ,$$

where $t_0 = 0$. As $X(t_k + 0)$, $k = 0, \ldots, n$, are random variables taking a finite number of values their dimensions are zero. Consequently

$$d\big(\xi_n(T)\big) = d(t_1, \ldots t_n) .$$

Now $(t_1, \ldots, t_n)$ has an absolutely continuous distribution. In order to prove that this is the case it is sufficient to show that if $0 \leq a_1 \leq b_1 \leq a_2 \leq \ldots \ldots \leq b_n \leq T$ then

$$\mathbf{P}\{a_i \leq t_i \leq b_i,\, i = 1, \ldots n \,|\, N(T) = n\} \leq C \prod_{i=1}^{n} (b_i - a_i) ,$$

where C is a constant independent of $\{a_i\}$ and $\{b_i\}$. As before $N(T)$ denotes the number of jumps in the interval $(0, T)$. Now, if $0 \leq a < b$

$$\mathbf{P}\{N(b) - N(\mathrm{a}) = 0 \,|\, X(a) = i\} = \mathbf{P}\{X(t) = i,\, a \leq t < b \,|\, X(a) = i\} = e^{-q_i(b-a)} ,$$

see [6]. Put

$$q = \max_i q_i .$$

Let Λ be a set in the Borel field generated by $\{X(t) : 0 \leqq t \leqq a\}$. Then

$$\mathbf{P}\{N(b) - N(a) \geqq 1 \mid \Lambda\} = \sum_i \mathbf{P}\{N(b) - N(a) \geqq 1,\ X(a) = i \mid \Lambda\} =$$

$$= \sum_i \mathbf{P}\{N(b) - N(a) \geqq 1 \mid X(a) = i\} \, \mathbf{P}\{X(a) = i \mid \Lambda\} =$$

$$= \sum_i (1 - e^{-q_i(b-a)}) \, \mathbf{P}\{X(a) = i \mid \Lambda\} \leq (1 - e^{-q(b-a)}) \sum_i \mathbf{P}\{X(a) = i \mid \Lambda\},$$

which gives

$$\mathbf{P}\{N(b) - N(a) \geqq 1 \mid \Lambda\} \leqq q(b-a) .$$

Consequently

$$\mathbf{P}\{a_i \leqq t_i \leqq b_i\,,\, i = 1, \ldots n \mid N(T) = n\} \leqq \frac{1}{\mathbf{P}\{N(T) = n\}} \mathbf{P}\{N(b_i) -$$

$$- N(a_i) \geqq 1\,,\, i = 1, \ldots n\} = \frac{1}{\mathbf{P}\{N(T) = n\}} \mathbf{P}\{N(b_1) -$$

$$- N(a_1) \geqq 1\} \prod_{i=2}^{n} \mathbf{P}\{N(b_i) - N(a_i) \geqq 1 \mid N(b_j) - N(a_j) \geqq 1\,, j =$$

$$= 1, \ldots i-1\} \leqq \frac{q^n}{\mathbf{P}\{N(T) = n\}} \prod_{i=1}^{n} (b_i - a_i) .$$

As the distribution of $(t_1, \ldots, t_n)$ is concentrated to a bounded set in the n-space

$$H_0([t_1], \ldots [t_n]) < \infty ,$$

and according to (4)

$$d(\zeta_n(T)) = d(t_1\,, \ldots t_n) = n .$$

Then (12) shows that $d^T(X) = m(T)$, the average number of jumps on the interval $(0, T)$. Let $V_{ij}(T)$ be the average number of jumps to j given that $X(0) = i$. Then according to [3], theorem II. 16. 2, $V_{ij}(\cdot)$ is differentiable and

$$\lim_{T\to\infty} V'_{ij}(T) = P_j q_j .$$

As

$$m(T) = \sum_{i,j} \mathbf{P}\{X(0) = i\} V_{ij}(T) ,$$

$m(\cdot)$ is differentiable and

$$\lim_{T\to\infty} m'(T) = \sum_{i,j} \mathbf{P}\{X(0) = i\} P_j q_j = \sum_j P_j q_j .$$

This proves the first part of the theorem. The second part now follows from the fact that if the initial distribution is equal to the stationary distribution, then according to stationarity (13)

$$m(T_1 + T_2) = m(T_1) + m(T_2) ,$$

all T_1, $T_2 > 0$. QED.

Let $\tau_1, \tau_2, \ldots$ be nonnegative random variables such that $\tau_1, \tau_2 - \tau_1, \tau_3 - \tau_2, \ldots$ are independent and have the same distribution function F. Then F will be called the gap distribution function of the renewal process $\{X(t) : t \in [0, \infty)\}$ where,

$$X(t) = \max_{\tau_n \leqq t} n .$$

If

$$\mu = \int_0^\infty x dF(x) < \infty$$

we get the corresponding stationary renewal process by letting τ_1 have the distribution function

$$F^*(x) = \frac{1}{\mu} \int_0^x [1 - F(y)]\, dy ,$$

see [11].

Theorem. 3. *If $\{X(t) : t \in [0, \infty)\}$ is a renewal process with absolutely continuous gap distribution function F, then $d^T(X)$ exists for every $T > 0$ and*

$$\lim_{T\to\infty} \frac{d^T(X)}{T} = \frac{1}{\mu} ,$$

where

$$\mu = \int_0^\infty x dF(x) .$$

If $\mu = \infty$ then $\frac{1}{\mu}$ is to be interpreted as zero. If $\mu < \infty$ the corresponding statiouary renewal process has the dimension

$$d^T = \frac{T}{\mu} .$$

Proof. Knowledge of the points of jump $t_1 < t_2 < \ldots < t_n$ makes it possible to determine the corresponding X-values and therefore

$$d(\xi_n(T)) = d(t_1, \ldots, t_n) .$$

The condition that the gap distribution is absolutely continuous implies the absolute continuity of the distribution of $(t_1, \ldots, t_n)$ given that $N(t) = n$. Indeed if $0 \leqq x_1 \leqq x_2 \ldots \leqq x_n \leqq T$ we have

$$\mathbf{P}\{t_i \leqq x_i , i = 1, \ldots n \mid N(T) = n\} =$$

$$= \frac{1}{\mathbf{P}\{N(T)=n\}} \int_0^{x_1} \int_{y_1}^{x_2} \int_{y_2}^{x_3} \cdots \int_{y_{n-1}}^{x_n} f(y_1) f(y_2 - y_1) f(y_3 - y_2) \ldots f(y_n - y_{n-1}) [1 -$$

$$- F(T - y_n)]\, dy_1 \ldots dy_n ,$$

where f is the derivative of F, and differentiation with respect to $x_1, x_2, \ldots, x_n$ gives the desired result. In the stationary case the proof is similar.

According to (4) one has

$$d(t_1, \ldots, t_n) = n$$

(cf. the proof of theorem 2). Therefore $d^T(X) = m(T)$, the average number of jumps on the interval $(0, T)$. Now

$$\lim_{T \to \infty} \frac{m(T)}{T} = \frac{1}{\mu},$$

see [10], and in the stationary case

$$m(T) = \frac{T}{\mu}$$

for all T, see [11].

QED.

If the gap distribution function has moments of higher orders, one can give a more detailed description of $m(T)$: s asymptotic behaviour, see [10]. The same is true of the dimension d^T.

§ 5. Dimension and entropy for purely discontinuous vector processes

Let $\{X(t) : t \in [0, \infty)\}$ be a purely discontinuous vector process, i.e.

$$X(t) = \big(X_1(t), \ldots, X_r(t)\big), \, 0 \leq t < \infty,$$

where $\{X_k(t) : t \in [0, \infty)\}$, $k = 1, \ldots, r$, are PDP : s. By $N_k(T)$ we denote the number of jumps on the interval $(0, T)$ of the $k : th$ PDP, $k = 1, \ldots, r$. Put

$$N(T) = \big(N_1(T), \ldots, N_r(T)\big).$$

If $n = (n_1, \ldots, n_r)$, where $\{n_k\}$ are nonnegative integers we put

$$q_n(T) = P\{N(T) = n\}.$$

The points of jump of the $k : th$ process are denoted $t_{k1} < t_{k2} < \ldots$. Measurability questions are here treated in the same way as in § 3. For $n = (n_1, \ldots, n_r)$ we let

$$\xi_n(T) = \big(t_{11}, \ldots, t_{1n_1}, t_{21}, \ldots, t_{rn_r}, X_1(+0),$$
$$X_1(t_{11} + 0), \ldots, X_r(t_{rn_r} + 0)\big)$$

have the conditional distribution of the points of jump and the corresponding X_k-values, $k = 1, \ldots, r$, given that $N(T) = n$. If the dimension and entropy of $\xi_n(T)$ exist we denote them $d_n(T)$ and $H_{d_n}(T)$. If we change $\sum\limits_{n=0}^{\infty}$ into

$$\sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \cdots \sum_{n_r=0}^{\infty}$$

we can use the definitions 1 and 2 literally to define the dimension

$$d^T = d^T(X) = d^T(X_1, \ldots X_r)$$

and the d^T-dimensional entropy

$$H^T = H^T(X) = H^T(X_1, \ldots X_r)$$

of $\{X(t) : t \in [0, \infty)\}$ regarded on the interval $(0, T)$.

With these definitions we find for example:
If $\{X_k(t) : t \in [0, \infty)\}$, $k = 1, \ldots, r$, are independent processes, then

$$d^T(X_1, \ldots X_r) = \sum_{k=1}^{r} d^T(X_k)$$

and

$$H^T(X_1, \ldots X_r) = \sum_{k=1}^{r} H^T(X_k),$$

if the respective right members exist. The proofs of these relations are straightforward computations from the definitions.

With $r = 2$, i.e. with two PDP : s, one can now define counterparts to mutual information and conditional entropy for pairs of random variables. For example, the conditional dimension and entropy of the process $\{X(t) : t \in \in [0, \infty)\}$ regarded on the interval $(0, T)$ given the process $\{Y(t) : t \in [0, \infty)\}$ on the same interval are defined by

$$d^T(X/Y) = d^T(X, Y) - d^T(Y)$$

and

$$H^T(X/Y) = H^T(X, Y) - H^T(Y),$$

if the right members exist, cf. § 4 of [1].

An example with Poisson processes. Let $\{X(t) : t \in [0, \infty)\}$ and $\{Z(t) : t \in [0, \infty)\}$ be independent separable Poisson processes with parameters λ and μ. Let $Y(t) = X(t) + Z(t)$, for $0 \leq t < \infty$. Then $\{Y(t) : t \in [0, \infty)\}$ becomes a separable Poisson process with parameter $\lambda + \mu$. We shall now calculate $d^T(X, Y)$, $H^T(X, Y)$, $d^T(X/Y)$ and $H^T(X/Y)$.

Let $N_1(T)$ and $N(T)$ be the number of jumps of $X(t)$ and $Y(t)$ on the interval $0 < t < T$. For $N(T) = n$ and $N_1(T) = n_1$ we let

$$t = (t_1, \ldots, t_n), t_1 < t_2 < \ldots < t_n,$$

and

$$\tau = (\tau_1, \ldots, \tau_{n_1}), \tau_1 < \tau_2 < \ldots < \tau_{n_1},$$

denote the points of jump of $Y(\cdot)$ and $X(\cdot)$. Then a.s. for every k, $1 \leq k \leq n_1$, $\tau_k = t_l$ for some l, $1 \leq l \leq n$. The conditional distribution of τ given t is finite with $\binom{n}{n_1}$ possible outcomes, all of which have the same probability. Therefore, cf. the example of § 3 in this paper and also § 4 of [1],

$$d(t, \tau) = d(t) = n$$

and

$$H(t, \tau) = \log \frac{T^n}{n!} + \log \binom{n}{n_1}.$$

From the definitions of dimension and entropy we then have

$$d^T(X, Y) = \sum_{n=0}^{\infty} \sum_{n_1=0}^{n} P\{X(T) = n_1, Y(T) = n\} n,$$

$$H^T(X, Y) = \sum_{n=0}^{\infty} \sum_{n_1=0}^{n} P\{X(T) = n_1, Y(T) = n\} \left[\log \frac{T^n}{n!} + \log \binom{n}{n_1}\right] -$$

$$- \sum_{n=0}^{\infty} \sum_{n_1=0}^{n} P\{X(T) = n_1, Y(T) = n\} \log P\{X(T) = n_1, Y(T) = n\},$$

which after some calculations gives that

$$d^T(X, Y) = (\lambda + \mu) T,$$

$$H^T(X, Y) = (\lambda + \mu) T \log e - \lambda T \log \lambda - \mu T \log \mu.$$

As $d^T(Y) = (\lambda + \mu) T$ and $H^T(Y) = (\lambda + \mu) T[\log e - \log (\lambda + \mu)]$,

see the example in § 3, one has that

$$d^T(X/Y) = 0,$$

$$H^T(X/Y) = [(\lambda + \mu) \log (\lambda + \mu) - \lambda \log \lambda - \mu \log \mu] T.$$

As the conditional entropy in this case is zero-dimensional it ought to be possible to calculate it without the use of the dimension concept. This is also the case. Let namely $\Delta = \{t_1, \ldots, t_n\}$ be a partition of the interval $(0, T)$, $0 < t_1 < t_2 < \ldots < t_n < T$, and put

$$|\Delta| = \max_{1 \leq i \leq n+1} (t_i - t_{i-1}),$$

where $t_0 = 0$ and $t_{n+1} = T$. Then if $h(t_1, \ldots, t_n)$ is the conditional entropy of the random vector $(X(t_1), \ldots, X(t_n))$ given the random vector $(Y(t_1), \ldots$ $\ldots, Y(t_n))$ one can show that

$$h(t_1, \ldots, t_n) = [(\lambda + \mu) \log(\lambda + \mu) - \lambda \log \lambda - \mu \log \mu] T + O(|\Delta|),$$

i.e. $h(t_1, \ldots, t_n) \to H^T(X/Y)$ as $|\Delta| \to 0$.

§ 6. Approximation of continuous processes

Let $\{X(t) : t \in [0, \infty)\}$ be a stochastic process, whose sample functions a.s. are continuous. Further, suppose for simplicity, that $\mathbf{P}\{X(0) = x_0\} = 1$ for some x_0, $-\infty < x_0 < \infty$. We can then form an approximating PDP in the following way. Put $\tau_0 = 0$ and

$$\tau_{n+1} = \inf \{t : t > \tau_n, | X(t) - X(\tau_n) | = \varepsilon\}, n = 0, 1, \ldots.$$

If for some n we have $| X(t) - X(\tau_n) | < \varepsilon$ for all $t > \tau_n$ we put $\tau_{n+1} = \tau_{n+2} = \ldots = \infty$. As the sample functions a.s. are continuous, $\{\tau_n\}$ become random variables defined on a set of probability one and possibly taking the value ∞ with positive probability. A detailed proof of this fact can be made with the help of a sequence $\{r_i\}$ dense in $(0, \infty)$, cf. the proofs of lemmata 1 and 2.

We put

$$X_\varepsilon(t) = X(\tau_n), \tau_n \leqq t < \tau_{n+1}, n = 0, 1, \ldots .$$

As $\mathbf{P}\{\tau_n \to \infty\} = 1$, $X_\varepsilon(t)$ is defined a.s. for $0 \leqq t < \infty$ and $\{X_\varepsilon(t) : t \in [0, \infty)\}$ becomes a PDP approximating the original process in the sense that

$$\mathbf{P}\{\lim_{\varepsilon \to 0} (\sup_{0 \leqq t < \infty} | X_\varepsilon(t) - X(t) |) = 0\} = 1 .$$

Suppose that the dimension $d^T(X_\varepsilon)$ exists for every $T > 0$. We pu

$$d_\varepsilon(X) = \lim_{T \to \infty} \frac{d^T(X_\varepsilon)}{T},$$

if the limit exists. The asymptotic properties of $d_\varepsilon(X)$ as ε tends to zero can then be used to characterize the original process.

If $\{X(t) : t \in [0, \infty)\}$ has independent and stationary increments (for these concepts see [6]), the sequence $\{\tau_n\}$ constitutes a renewal process. If the gap distribution function F is absolutely continuous we have according to theorem 3

$$d_\varepsilon(X) = \frac{1}{\mu},$$

where

$$\mu = \int_0^\infty x dF(x) .$$

For by the calculation of the dimension of the process $\{X_\varepsilon(t) : t \in [0, \infty)\}$ we need not take account of the values of the process as $\{X(\tau_1), \ldots, X(\tau_n)\}$ for every n is a random vector taking a finite number of values.

Example. Let $\{X(t) : t \in [0, \infty)\}$ be a separable Brownian motion process. According to [5]

$$\int_0^\infty e^{-\lambda x} dF(x) = \frac{1}{\cosh \sqrt{2 \lambda} \varepsilon},$$

where F is the gap distribution function. Therefore

$$\mu = \lim_{\lambda \to 0} \left\{ - \frac{d}{d\lambda} \frac{1}{\cosh \sqrt{2 \lambda} \varepsilon} \right\} = \varepsilon^2 ,$$

and

$$d_\varepsilon(X) = \frac{1}{\varepsilon^2} .$$

This result can compared with JAGLOMS result

$$H_\varepsilon = \frac{4}{\pi}\frac{1}{\varepsilon^2} + o\left(\frac{1}{\varepsilon^2}\right),$$

see [8] page 107. H_ε is the KOLMOGOROV ε-entropy for the Brownian motion process on the interval $(0, 1)$, when the sample functions are considered as elements of $L^2(0, 1)$.

Acknowledgements

For stimulating discussions on the subject of this paper I thank Professor Harald Bergström. I also want to express my thanks to I. Csiszár and Jan Petersson for valuable remarks and to Statens Tekniska Forskningsgråd, from which I held a research fellowship during most of the time spent on the work on the subject of this paper.

(Received September 17, 1963)

REFERENCES

[1] BALATONI, J. and RÉNYI, A.: »Zum Begriff der Entropie«. (German translation from Hungarian in *Arbeiten zur Informationstheorie*, I, Berlin, 1957, pp. 117—134.)
[2] CHINTSCHIN, A. J.: »Der Begriff der Entropie in der Wahrscheinlichkeitsrechnung.« (German translation from Russian in *Arbeiten zur Informationstheorie*, I, Berlin, 1957, pp. 7—25.)
[3] CHUNG, K. L.: *Markov chains with stationary transition probabilities*. Berlin, Göttingen, Heidelberg, 1960.
[4] CSISZÁR, I.: „Some remarks on the dimension and entropy of random variables". *Acta Math. Acad. Sci. Hung.*, **12** (1961), 399—408.
[5] DARLING and SIEGERT: „The first passage problem for a continuous Markov process". *Ann. Math. Statist.*, **24** (1953), 624—639.
[6] DOOB, J. L.: *Stochastic processes*. New York, 1953.
[7] FISZ, M.: „Characterization of sample functions of stochastic processes by some absolute probabilities." *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol.* 2 (1961) 143—151.
[8] KOLMOGOROW, A. N.: „Theorie der Nachrichtenübermittlung". (German translation from Russian in *Arbeiten zur Informationstheorie*, I, Berlin, 1957, pp. 91—116).
[9] RÉNYI, A.: „On the dimension and entropy of probability distributions." *Acta. Math. Acad. Sci. Hung.*, **10** (1959), 193—215.
[10] SMITH, C.: „Renewal theory and its ramifications". *J. Roy. Statist. Soc., Ser. B*, **20** (1958), 243—282.
[11] TAKÁCS, L.: *Stochastic processes*. London, 1960.