

PROBABILITY EXPONENTS OF THE CLASSES OF CERTAIN PARTITIONS RELATIVE TO DISCRETE-PARAMETER STOCHASTIC PROCESSES (*)

by ANDREA SGARRO (in Trieste) (**)

SOMMARIO. - *Si considerano certe partizioni delle successive potenze cartesiane del codominio di un processo stocastico a parametro discreto. Si esamina il caso in cui asintoticamente l'esponente di probabilità di una qualsiasi unione di classi della partizione coincida con l'esponente di probabilità di un'unica classe dell'unione. L'interesse in tale proprietà proviene da certe questioni di Teoria dell'Informazione e di Statistica. Si investigano classi di processi stocastici più ampie di quelle finora considerate.*

SUMMARY. - *We consider certain partitions of the successive Cartesian powers of the range of a discrete-parameter stochastic process. We examine the case when asymptotically the probability exponent of any union of partition-classes is equal to the probability exponent of a single class in the union. Our interest is motivated by problems of Information Theory and Statistics. We investigate larger classes of stochastic processes than those which have been considered so far.*

1. Introduction.

In this paper we investigate certain probability-theoretic questions, which now we describe in loose terms. Assume that a discrete-parameter stochastic process (s. p.) $\{X_n\}_{n \in \mathbb{N}}$ is given. The common range of the random variables (r. v. 's) X_n is denoted by K and is supposed to be

(*) Pervenuto in Redazione il 18 gennaio 1978.

Lavoro eseguito nell'ambito del Gruppo Nazionale per l'Informatica Matematica del C.N.R.

(**) Indirizzo dell'Autore: Istituto di Matematica dell'Università - Piazzale Europa 1 - 34100 Trieste.

at most countable; the probability that the random vector (X_1, \dots, X_n) belongs to A_n , that is $Pr\{X_1, \dots, X_n \in A_n\}$, is denoted simply by $Pr(A_n)$ whenever this does not cause ambiguities (here A_n is any subset of K^n , the n -th Cartesian power of K); if A_n is a singleton, $A_n = \{x_n\}$, we shall write $Pr(x_n)$. The non-positive number $n^{-1} \log Pr(A_n)$ is called the *probability exponent* (p. e.) of A_n ; logs are taken to the base e . A *stratification* of $\{K^n\}_{n \in N}$ is a sequence of finite partitions of K^n :
$$K^n = \bigcup_{i \in I(n)} C_{n,i}, |I(n)| < \infty, n \in N.$$

We are interested in the case when the stratification meets certain demands. The first is the following: asymptotically the p. e. of *any union* of $C_{n,i}$'s is equal to the p. e. of *one* of the $C_{n,i}$'s in the union. Then the stratification is called an α -stratification for the s. p. $\{X_n\}_{n \in N}$. The second demand is the following: all the elements of any $C_{n,i}$ have the same probability. When this demand is met the stratification is called a β -stratification for the s. p. $\{X_n\}_{n \in N}$. A stratification which both α and β for a s. p. $\{X_n\}_{n \in N}$ will be called a γ -stratification for that process.

As it will be seen, α -stratifications are easier to construct when the cardinality $|I(n)|$ is « small », while β -stratifications are easier to construct when $|I(n)|$ is « large ». A very important case is known, however, when a trade-off is possible and there exists a γ -stratification. This is the case when the s. p. is finite range, stationary and memoryless. Then a γ -stratification is obtained by partitioning K^n in *composition classes* (two element of K^n belong to the same composition class iff they are a permutation of each other). The knowledge of this γ -stratification allows the use of powerful techniques in basic problems of both Information Theory and Statistics (for a heuristic description of this techniques see [1] and [2]).

Therefore we think that it is desirable to investigate the properties of stratifications for larger classes of s. p.'s, as we have done in this paper. In particular the results that we have obtained in the finite-range stationary Markov case appear to be promising and worth of further research-work in a more specific context, whether information-theoretic or statistical (we have in mind, e. g., Longo's conjecture, [3]).

Section 2 is devoted to introducing some more notation. In Section 3, α , β and γ -stratifications are defined rigorously and some basic properties are stated. Section 4 is devoted to the finite-range stationary Markov case. In Section 5 we examine the case when $|I(n)|$ is « large », and show that non-trivial α -stratifications exist even in this case.

2. Notation.

Although we refer to texts like [4] or [5] for standard nomenclature and notation, we wish to make clear some points. By symbols like $X^\infty = \{X_n\}_{n \in \mathbb{N}}$ we denote a discrete-time stochastic process (s. p.); the random vector (X_1, \dots, X_n) is denoted simply by X^n . If it is not otherwise stated, all s. p.'s are assumed to have a common range K . Stratifications and probability exponents (p. e.'s) have been defined in the introduction. Stratifications will be denoted by symbols like $\{\mathcal{P}_n\}_{n \in \mathbb{N}}$; here $\mathcal{P}_n = \{C_{n,i}\}_{i \in I(n)}$ is a partition of K^n , $n \in \mathbb{N}$. A stratification $\{\mathcal{P}_n\}_{n \in \mathbb{N}}$ is said to be finer (grosser) than the stratification $\{\mathcal{Q}_n\}_{n \in \mathbb{N}}$ when, for all n , the partition \mathcal{P}_n is finer (grosser) than the partition \mathcal{Q}_n . P. e.'s can be infinite (when the probability is zero). Expressions like $\max_{i \in J(n)} Pr(C_{n,i})$ or $n^{-1} \log Pr(A_n) - n^{-1} \log Pr(B_n)$ are to be interpreted as zero when $J(n) = \emptyset$ or $Pr(A_n) = Pr(B_n) = 0$; these conventions have the unique purpose of avoiding trivial specifications.

3. Various Types of Stratifications.

Let $\{\mathcal{P}_n\}_{n \in \mathbb{N}}$ be a stratification of $\{K^n\}_{n \in \mathbb{N}}$, $\mathcal{P}_n = \{C_{n,i}\}_{i \in I(n)}$, $n \in \mathbb{N}$; let $\{X_d\}_{d \in D}$ be a family of s. p.'s with range K .

DEFINITION 1: $\{\mathcal{P}_n\}_{n \in \mathbb{N}}$ is said to be an α -stratification for $\{X_d\}_{d \in D}$ when:

$$(1) \quad \lim_n \left[\sup_{d \in D} \left[\max_{J(n) \subseteq I(n)} \left[n^{-1} \log Pr \left(\bigcup_{i \in J(n)} C_{n,i} \right) - n^{-1} \log \max_{i \in J(n)} Pr(C_{n,i}) \right] \right] \right] = 0.$$

Roughly stated, equality (1) means that if one is interested in p. e.'s rather than in the probabilities themselves, then asymptotically any union of classes of the partition \mathcal{P}_n can be replaced by a single class in the union. The sup and the max which appear after the lim sign in (1) are simply a uniform convergence proviso.

A very simple condition can be given which assures that a stratification is an α -stratification for any family of s. p.'s. First we give a definition. Let $\{\mathcal{P}_n\}_{n \in \mathbb{N}}$, $\mathcal{P}_n = \{C_{n,i}\}_{i \in I(n)}$ be a stratification; if $\lim_n n^{-1} \log |I(n)| = 0$, the stratification is said to be sparse; otherwise
 " $(\limsup_n n^{-1} \log |I(n)| > 0)$ the stratification is said to be *thick*.

PROPOSITION 1: *A sparse stratification is an α -stratification for any family of stochastic processes.*

PROOF: let $J(n)$ be a subset of indices. One has:

$$\begin{aligned} \Pr \left(\bigcup_{i \in J(n)} C_{n,i} \right) &= \sum_{i \in J(n)} \Pr (C_{n,i}) \leq |J(n)| \max \Pr (C_{n,i}) \leq \\ &\leq |I(n)| \max \Pr (C_{n,i}). \end{aligned}$$

To complete the proof take n^{-1} log of the extreme sides and use the sparseness of the stratification.

Q. E. D.

Examples of thick α -stratifications will be given in Section 5.

DEFINITION 2: A stratification $\{\mathcal{P}_n\}_{n \in \mathbb{N}}$, $\mathcal{P}_n = \{C_{n,i}\}_{i \in I(n)}$, is a β -stratification for the family of s. p.'s $\{X_d^\infty\}_{d \in D}$ when the following equality holds for all values of the index d , for all $C_{n,i}$'s and for all couple of elements of $C_{n,i}$, x_n and y_n ($n \in \mathbb{N}$, $i \in I(n)$):

$$\Pr \{X_d^n = x_n\} = \Pr \{X_d^n = y_n\}.$$

PROPOSITION 2: *If the stratification $\{\mathcal{P}_n\}_{n \in \mathbb{N}}$ is a sparse α -stratification for a family of s. p.'s, then any grosser stratification $\{\mathcal{Q}_n\}_{n \in \mathbb{N}}$ is also a sparse α -stratification for that family of s. p.'s; if the stratification $\{\mathcal{P}_n\}_{n \in \mathbb{N}}$ is a β -stratification for a family of s. p.'s, then any finer stratification $\{\mathcal{Q}_n\}_{n \in \mathbb{N}}$ is again a β -stratification for that family of s. p.'s.*

PROOF: The first part of Proposition 2 follows from Proposition 1 since also $\{\mathcal{Q}_n\}_{n \in \mathbb{N}}$ is sparse; the second part is obvious.

Q. E. D.

In the next section we shall give examples of (sparse) stratifications which are both α and β for suitable families of s. p.'s.

DEFINITION 3: A stratification which is both an α -stratification and a β -stratification for a family of s. p.'s is said to be a γ -stratification for that family of s. p.'s.

PROPOSITION 3: *A sparse β -stratification for the family of s. p.'s $\{X_d^\infty\}_{d \in D}$ is also a γ -stratification for $\{X_d^\infty\}_{d \in D}$.*

PROOF: Use Proposition 1.

Q. E. D.

4. The Finite-Range Stationary Markov Case.

Let m be any non-negative integer.

THEOREM. *If K is finite there exists a sparse stratification which is a γ -stratification for any family of stationary Markov processes with range K and memory at most m .*

PROOF: Let $K = \{c_1, \dots, c_{|K|}\}$, $|K| < \infty$.

We shall subdivide the proof into three parts.

a) First we exhibit a partition of K^n . Let $r = |K^m|$; denote by $\gamma_1, \dots, \gamma_r$ the elements of K^m in an arbitrary but henceforth fixed order. We shall define an equivalence relation over K^n . If $n \leq m$, two elements of K^n are equivalent iff they are equal. Let n be larger than m , and let x_n be any element of K^n . Denote by x_m and \tilde{x}_{n-m} the two substrings of x_n such that $x_n = x_m \tilde{x}_{n-m}$ and $l(x_m) = m$ ($l(\cdot)$ is the length of its argument). With x_n we shall associate r strings in the following way: the j -th string, $1 \leq j \leq r$, is made up by all the components of \tilde{x}_{n-m} which are preceded by γ_j , in the same order as they appear in \tilde{x}_{n-m} . Let $\alpha_1, \dots, \alpha_r$ be the r strings which have been associated with x_m ; one has $0 \leq l(\alpha_j) \leq n - m$, ($1 \leq j \leq r$), $\sum_{1 \leq j \leq r} l(\alpha_j) = n - m$. Let $y_n = y_m \tilde{y}_{n-m}$ be another element of K^n ($l(y_m) = m$) and let β_1, \dots, β_r be the strings which have been associated with it. We say that x_n and y_n are equivalent iff:

- i) $x_m = y_m$;
- ii) β_j is a permutation of α_j , $1 \leq j \leq r$.

The classes of the equivalence relation which we have just defined will be called *composition classes of order n and memory m* .

b) Now consider the stratification of $\{K^n\}_{n \in \mathbb{N}}$ in composition classes (c. c.'s) of order n and memory m . This is a β -stratification for all stationary Markov processes with range K and memory at most m . In fact in such processes the probability of an element of K^n , $x_n = a_1, \dots, a_n$ ($n > m$), is given by:

$$\begin{aligned} Pr \{X^n = a_1, \dots, a_n\} &= \prod_{m+1 \leq j \leq n} Pr \{X_j = a_j | X_{j-m} X_{j-m+1}, \dots, X_{j-1} = \\ &= a_{j-m} a_{j-m+1}, \dots, a_{j-1}\}; \end{aligned}$$

but this number depends only on the c. c. to which x_n belongs, and not on x_n itself.

c) Now we prove that the stratification of $\{K^n\}_{n \in \mathbb{N}}$ in c. c.'s of order n and memory m is sparse.

Assume $n > m$. Let $C_{n,i}$ be a c. c. of order n and memory m and let x_n be any of its element. $C_{n,i}$ is identified by the common « head » of its elements, x_m , and by the r strings $\alpha_1, \dots, \alpha_r$ associated with x_n ($x_n = \tilde{x}_m x_{n-m}$, $l(x_m) = m$); on the other hand $C_{n,i}$ identifies the γ_j 's only up to a permutation of the latter. Therefore it is convenient for us to replace each γ_j with a K -length string v_j whose h -th component is the number of occurrences of c_h in γ_j ; $1 \leq j \leq r$, $1 \leq h \leq |K|$.

Now different c. c.'s identify and are identified by different $(r+1)$ -tuples (x_m, v_1, \dots, v_r) . The components of the v_j 's are subject to several constraints, of which we shall retain only the following; they must be integers ranging between 0 and $n-m$. Thus an upper bound to the number of c. c.'s of order n and memory m is given by:

$$r [(n-m+1)^{|K|}]^r = |K|^m (n-m+1)^{|K|^{m+1}}.$$

Since $\lim_{n \rightarrow \infty} n^{-1} \log |K|^m (n-m+1)^{|K|^{m+1}} = 0$, the stratification in c. c.'s of order n and memory m is sparse.

Use Proposition 3 to complete the proof of the theorem.

Q. E. D.

5. Thick Stratifications.

It is apparent by now that the simple and powerful properties of sparse stratifications make their use most desirable; basically this depends on Proposition 1, which asserts that sparse α -stratifications are « universal ». Instead, in this section we shall give « non-universal » examples of thick α -stratifications. Notice that trivial examples of thick α -stratifications for a family of s. p.'s $\{X_d\}_{d \in D}$ are easily found; we have in mind the case when $\lim_{n \rightarrow \infty} n^{-1} \log |I(n) - J(n)| = 0$, where $J(n)$ denotes the subset of indices whose corresponding partition-classes have zero-probability for all the s. p.'s X_d , $d \in D$. The example below are not trivial.

I) We begin by a non-stationary memoryless example. Let $K = \{0, 1\}$. To specify a non-stationary memoryless process $\{X_n\}_{n \in \mathbb{N}}$

over K it is enough to assign a sequence of positive real numbers, $\{r_n\}_{n \in N}$, and set $p_n = Pr\{X_n=1\} = r_n/(1+r_n)$, $q_n = Pr\{X_n=0\} = 1/(1+r_n)$; r_n is the ratio p_n/q_n , $n \in N$. The sequence $\{r_n\}_{n \in N}$ will be chosen according to the following recursive set of rules:

- 1) $r_1=5$;
- (2)
- n) $r_n=1+r_{n-1} \cdot \prod_{1 \leq i \leq n-1} r_i, \quad n \geq 2.$

Fix n . We describe a partition of K^n . If $x_n = a_1, \dots, a_n$ is an element of K^n , its substrings $h(x_n) = a_1, \dots, a_{s(n)}$ and $t(x_n) = a_{s(n)+1}, \dots, a_n$ will be called the head and the tail of x_n , respectively: $s(n)$ is defined as the integer par of $n/2$. We say the two elements of K^n are equivalent iff they have the same tail. Denote by $\{\mathcal{P}_n\}_{n \in N}$, $\mathcal{P}_n = \{C_{n,i}\}_{i \in I(n)}$, $n \in N$, the corresponding stratification of $\{K^n\}_{n \in N}$. We shall prove that $\{\mathcal{P}_n\}_{n \in N}$ is a thick α -stratification for $\{X_n\}_{n \in N}$. We subdivide the proof into three parts. For brevity, in the sequel expressions like $Pr\{X_h X_{h+1}, \dots, X_{h+k} = \alpha\}$ will be written simply as $Pr(\alpha)$ whenever this does not cause ambiguity; $h \geq 1, k \geq 0, \alpha \in K^{k+1}$.

a) Let $x_n = a_1, \dots, a_n$ and $y_n = b_1, \dots, b_n$ belong to K^n ($x \neq y$), and let m be the index of the rightmost component where they differ. We prove that $Pr(x_n) > Pr(y_n)$ if $a_m = 1$.

In fact, if $a_m = 1$:

$$\frac{Pr(x_n)}{Pr(y_n)} = \frac{Pr(a_1) \dots Pr(a_{m-1}) Pr(a_m)}{Pr(b_1) \dots Pr(b_{m-1}) Pr(b_m)}$$

$$\geq \frac{q_1 \dots q_{m-1} p_m}{p_1 \dots p_{m-1} q_m} = (r_1 \dots r_{m-1})^{-1} r_m \stackrel{(i)}{>} r_{m-1} > 1.$$

In (i) we have used the fact that $r_m > r_1, \dots, r_{m-2} r_{m-1}^2$ (see (2)). In the same way one proves that $Pr(y_n) > Pr(x_n)$ if $b_m = 1$.

b) Take now an arbitrary union of (at least two) classes of \mathcal{P}_n . One has:

$$(3) \quad Pr\left(\bigcup_{i \in J(n)} C_{n,i}\right) < Pr(C_n^{(1)}) + 2^n Pr(C_n^{(2)}),$$

where $C_n^{(1)}$ denotes a class in the family whose probability is maximal, while $C_n^{(2)}$ denotes a class of maximal probability in the family obtained from $\{C_{n,i}\}_{i \in J(n)}$ by suppressing $C_n^{(1)}$.

Denote by $t_n^{(1)} = a_{s(n)+1}, \dots, a_n$ and $t_n^{(2)} = b_{s(n)+1}, \dots, b_n$ the tails of the elements of $C_n^{(1)}$ and $C_n^{(2)}$, respectively. One has:

$$(4) \quad \frac{Pr(C_n^{(1)})}{Pr(C_n^{(2)})} = \frac{\sum_{x_n \in O_n^{(1)}} Pr(x_n)}{\sum_{y_n \in O_n^{(2)}} Pr(y_n)} = \frac{Pr(t_n^{(1)}) \sum_{x_{s(n)} \in K^{s(n)}} Pr(x_{s(n)})}{Pr(t_n^{(2)}) \sum_{y_{s(n)} \in K^{s(n)}} Pr(y_{s(n)})}$$

$$= \frac{Pr(t_n^{(1)})}{Pr(t_n^{(2)})} = \frac{Pr(a_{s(n)+1}) \dots Pr(a_n)}{Pr(b_{s(n)+1}) \dots Pr(b_n)}.$$

Let m be the index of the rightmost component where $t_n^{(1)}$ and $t_n^{(2)}$ differ. One has $a_m = 1$, $b_m = 0$ (use (a)). If $m = s(n) + 1$ the last side of (4) is $r_{s(n)+1}$. If $m > s(n) + 1$, one has:

$$\frac{Pr(C_n^{(1)})}{Pr(C_n^{(2)})} = \frac{Pr(a_{s(n)+1}) \dots Pr(a_m)}{Pr(b_{s(n)+1}) \dots Pr(b_m)} = \frac{Pr(b_{s(n)+1}) \dots Pr(a_{m-1}) r_m}{Pr(b_{s(n)+1}) \dots Pr(b_{m-1})}$$

$$\stackrel{(ii)}{\geq} (r_1 \dots r_{m-1})^{-1} r_m > r_{m-1} \geq r_{s(n)+1}.$$

In (ii) we have again used (2). Therefore:

$$\frac{Pr(C_n^{(2)})}{Pr(C_n^{(1)})} \leq (r_{s(n)+1})^{-1}$$

A substitution in (3) yields:

$$(5) \quad Pr\left(\bigcup_{i \in J(n)} C_{n,i}\right) < Pr(C_n^{(1)}) [1 + 2^n (r_{s(n)+1})^{-1}]$$

$$\stackrel{(iii)}{<} 2 Pr(C_n^{(1)}).$$

In (iii) we have used the fact that $r_{s(n)+1} \geq 2^{2(s(n)+1)} > 2$ (use (2)). It is enough to take $n^{-1} \log$ of the extreme sides of (5) to prove that $\{\mathcal{P}_n\}_{n \in \mathbb{N}}$ is an α -stratification for $\{X_n\}_{n \in \mathbb{N}}$.

c) The above α -stratification is thick. In fact the number of the classes of \mathcal{P}_n is the same as the number of the possible tails, that is $2^{n-s(n)} = 2^{n-\text{int}(n/2)}$. On the other hand

$$\lim_n n^{-1} \log 2^{n-\text{int}(n/2)} = \frac{1}{2} \log 2.$$

Thus we have exhibited a thick α -stratification for $\{X_n\}_{n \in \mathbb{N}}$.

II) The non-stationary example which we have given can be used to exhibit thick α -stratifications for different types of s. p.'s. Assume, for example, that $\{Y_n\}_{n \in N}$ is a stationary memoryless s. p.; the Y_n 's are, say, Poisson r. v.'s with parameter λ , $\lambda > 0$. Now the range is $N_0 = N \cup \{0\}$. Let f_n be the largest quantile of order p_n ; here p_n is the same number as in the example above. Let us define a mapping ψ_n of N_0^n onto $K^n = \{0, 1\}^n$: if $x_n = a_1, \dots, a_n$, $\psi_n(x_n)$ is the n -tuple which is obtained from x_n by replacing a_i with 1 if $a_i \leq f_i$, with 0 otherwise; $1 \leq i \leq n$. Consider the following equivalence relation over N_0^n : x_n and y_n are equivalent iff $\psi_n(x_n)$ and $\psi_n(y_n)$ are equivalent in K^n according to the equivalence relation which has been used in the example above. It is immediate to see that the corresponding stratification of $\{N_0^n\}_{n \in N}$ is a thick α -stratification for $\{y_n\}_{n \in N}$.

Acknowledgement. The author thanks G. Longo for his help and for pointing him out the equivalence relation used in Section 4.

REFERENCES

- [1] G. LONGO and A. SGARRO, *The Source Coding Theorem Revisited: a Combinatorial Approach*. (To appear on I. E. E. E. Trans. on Information Theory).
- [2] G. LONGO and A. SGARRO, *The Error Exponent for the Testing of Simple Statistical Hypotheses: a Combinatorial Approach*. Submitted for publication.
- [3] G. LONGO, *On the Error Exponent for Markov Sources*. 2nd Intern. Symp. on Information Theory, Tsahkadsor, Soviet Armenia, 1971.
- [4] H. R. PITT, *Integration, Measure and Probability*. Oliver & Boyd, Edinburgh and London, 1963.
- [5] J. L. DOOB, *Stochastic Process*. Wiley, New York, 1953.