Astérisque

JAN KWIATKOWSKI
FELIKS MANIAKOWSKI

On the transportation problem used in the definition of Ornstein's distance $\bar{d}$

Astérisque, tome 50 (1977), p. 223-238
<http://www.numdam.org/item?id=AST_1977__50__223_0>

Société Mathématique de France
Astérisque 50 (1977) p.223-238

ON THE TRANSPORTATION PROBLEM USED IN THE DEFINITION OF ORNSTEIN'S DISTANCE $\bar{d}$

Jan Kwiatkowski
Feliks Maniakowski

Summary.

The transportation problem used in the definition of Ornstein's $\bar{d}$ distance is considered. Some properties of optimal solutions are given. The $\bar{d}$ distance for Bernoulli measures and for some binary Markov stationary measures is calculated.

1. Introduction

In the paper [1] Ornstein defined a distance $\bar{d}$ between two stationary processes. A stationary process is a pair (T,P), where T is an ergodic automorphism of a Lebesgue's space Ω with a nonatomic measure m and where $P = (P_o, \dots, P_{s-1})$ is a finite partition of Ω. To any stationary process corresponds a measure μ defined on the σ-algebra of Borel subsets of the set X of doubly infinite sequences $x = (\dots, x_{-1}, x_o, x_1, \dots)$, where $0 \leq x_i \leq s-1$. The measure μ is defined as follows : if C is a cylinder i.e. $C = \{x \in X : x_o = i_o, \dots, x_{n-1} = i_{n-1}\}$, then $\mu(C) = m(\bigcap\{T^j P_{i_j} : j = 0,\dots,n-1\})$. It is well known that such a function can be extended to a measure on Ω and we denote this measure with the same letter μ.

Two stationary processes (T,P) , (S,Q) , where $Q = (Q_o,\ldots,Q_{n-1})$ are equivalent, if they define the same measure in the space X . Thus the notion of distance between two stationary processes may be reduced to the distance between two stationary measures on X , corresponding to these processes. Now, we shall give two equivalent definitions coming from the Ornstein's papers / compair : [1] and a foot-note of Vershik in the Russian translation [4] of [1]/. Let X_n be a set of n-tuples of symbols $0,1,\ldots,s-1$. If $b = (b_o,\ldots,b_{n-1})$, $c = (c_o,\ldots,c_{n-1}) \in X_n$, then

$$\bar{\rho}(b,c) = \frac{1}{n} |\{r: b_r \neq c_r \ , \ r=0,1,\ldots,n-1\}|$$

Let μ_n , for a measure μ on X , denote $\mu|X_n$ or, equivalently,

$$\mu_n(b_o,\ldots,b_{n-1}) = \mu(\{x \in X : x_o = b_o,\ldots,x_{n-1} = b_{n-1}\}) .$$

Let $Y_n = X_n \times X_n$ and let p_1,p_2 be projections of Y_n .

<u>Definition 1</u>

$$\bar{d}(\mu,\nu) = \sup d_n(\mu_n,\nu_n) \quad , \quad \text{where}$$

$$d_n(\mu_n,\nu_n) = \min_{\sigma_n} \sum_{b,c\in X_n} \bar{\rho}(b,c)\ \sigma_n(b,c)$$

σ_n being any measure on the / finite / set Y_n such that $p_1 \sigma_n = \mu_n$ and $p_2\sigma_n = \nu_n$.

It may be shown [2], that $\bar{d}(\mu,\nu) = \lim d_n(\mu_n,\nu_n)$.

It is clear that in order to find d_n it is sufficient to solve a transportation problem with the costs matrix $\bar{\rho}(b,c)$.

The equivalent definition of $\bar{d}$ is following. Consider the set Y of doubly infinite sequences $(\ldots,y_{-1},y_o,y_1,\ldots)$ elements of which are pairs y_i of symbols $0,1,\ldots,s-1$. The set $Y_n = X_n \times X_n$ can be treated as $(Y)_n$, it means as the set of sequences

$z = (z_o, \ldots, z_{n-1})$ with $z_1 \in \{0,1,\ldots,s-1\}^2$. Consider a measure σ on Y which is invariant under the shift or the space Y , with the property $p_1\sigma_n = \mu_n$, $p_2\sigma_n = \nu_n$ $n = 0,1,\ldots$. Define :

$$k_n(\sigma_n) = \sum_{b,c\in X_n} \bar{\rho}(b,c)\ \sigma_n(b,c)$$

It may be shown that $k_n(\sigma_n) = \sum_{i\neq j} \sigma_1(i,j)$, where $i,j = 0,\ldots,s-1$, and are treated as 1-tuples .

<u>Definition 2</u>

$\bar{d}(\mu,\nu) = \inf \sum_{i\neq j} \sigma_1(i,j)$, where σ runs the set of invariant measure on Y such that $p_1\sigma_n = \mu_n$, $p_2\sigma_n = \nu_n$.

By occasion of this definition let us make the following remarks. The condition for a measure σ on Y to be invariant are following:

$$\sum_{i,j=0}^{s-1} \sigma_n(bi,cj) = \sum_{i,j=0}^{s-1} \sigma_n(ib,jc)$$

for any $(b,c) \in Y_n$ and for $n = 1,2,\ldots$ Consider any measure σ_n on Y_n satisfying these conditions. Let

$$\tilde{d}_n(\mu_n,\nu_n) = \inf \{k_n(\sigma_n) : p_1\sigma_n = \mu_n , p_2\sigma_n = \nu_n\} .$$

Define σ^n_{n-1} on Y_n by equality :

$$\sigma^n_{n-1}(b,c) = \sum_{i,j=0}^{s-1} \sigma_n(bi,cj)$$

for $(b,c) \in Y_{n-1}$. It is easy to see that σ^n_{n-1} is an invariant measure. Similarly define σ^n_{n-2} by means of σ^n_{n-1} , and so on. As a result one obtains a sequence of measures σ^n_{n-3} , $\sigma^n_{n-4},\ldots,\sigma^n_1$ such that

$$k_1(\sigma^n_1) = k_2(\sigma^n_2) = \ldots = k_{n-1}(\sigma^n_{n-1}) = k_n(\sigma_n).$$

Hence $\tilde{d}_1 \leq \tilde{d}_2 \leq \tilde{d}_3 \leq \ldots$

Of course $d_n \leq \tilde{d}_n$ for any n and from Definition 2 $\tilde{d}_n \leq \bar{d}$.

Therefore $\lim \tilde{d}_n = \bar{d}$. In what follows, we show that the case $\tilde{d}_n > d_n$ may occur.

In the present paper we consider the transformation problem appearing in the Definition 1 of Ornstein's $\bar{d}$ distance, describe some properties of optimal solutions of such problems and use them for a calculation of the distance $\bar{d}$ between Bernoulli measures and between some binary Markov measures.

2. The matrix of the transportation problem considered here is a square matrix with s^n columns and rows, s and n being positive integrs. Each number $k = 0,1,\dots,s^n-1$ is represented by a n-tuple $(k_o,k_1,\dots,k_{n-1})$ such that $k = \sum_{i=n}^{i=n-1} k_i d^{n-i-1}$. The elements of the matrix are denoted by ρ_{ij} and defined by the equality.

$$\rho_{ij} = |\{r:\ i_r \neq j_r\ ,\ r=0,1,\dots,n-1\}\ |\quad (i,j=0,1,\dots,s^n-1)\ .$$

where $|A|$ is the number of elements of a set.A. Thus ρ_{ij} is the number of places on which $(i_o,\dots,i_{n-1})$ and $(j_o,\dots,j_{n-1})$ have different symbols.

We use an unoriented graph G defined as follows. The vertices of G are numbers $k = 0,1,\dots,s^n-1$ or, equivalenty, n-tuples representing them. Two n-tuples $(i_o,\dots,i_{n-1})$, $(j_o,\dots,j_{n-1})$ are joined by an edge iff $\rho_{ij} = 1$. A pair of such vertices is refered as neighbour.

In order to solve a transportation problem one can use two functions (u_i) , (v_i) , $i = 0,1,\dots,s^{n-1}$ such that the numbers

(1) $d_{ij} = \rho_{ij} + u_i - v_j \quad (i,j=0,1,\dots,s^n-1)$

satisfy following conditions

(2) $d_{ij} \geq 0$,

(3) $d_{ij} = 0$ for (i,j) constituing a basis.

By a basis, we mean any set of $2s^n-1$ cells of the matrix of a transportation problem not containing a closed loop. A feasible solution of transportation problem is optimal if for any cell (i,j) occupied by the solution, $d_{ij} = 0$.

It follows from results of Vershik [3] , that any transportation problem with the costs matrix ρ_{ij} described above has a solution which occupies all the cells $(i,i), i=0,\dots,s^n-1$ and we can suppose $d_{ii} = 0$ or $u_i = v_i (i=0,\dots,s^n-1)$. Thus we can use one fonction (u_i) instead of two (u_i) and (v_i) .

We call a potential any function $(u_i), i=0,\dots,s^n-1$ such that the matrix d_{ij} with elements

(4) $d_{ij} = \rho_{ij} + u_i - u_j$

satisfies conditions (2) and (3) .

It is well known that a solution of the transportation problem with matrix ρ_{ij} is optimal iff for all cells (i,j) occupied by this solution $d_{ij} = 0$ for some potential.

Our aim is to describe all potentials for one matrix ρ_{ij} and give an effective method of producing all potentials in the case $s = 2$ for any given n .

3. For the proof of the main theorem we need two easy lemmas.

Lemma 1.

If (u_i) is a potential and $u_o = 0$, then u_i is an integer for all $i = 0,1,\dots,s^n-1$.

Proof.

$d_{ij} = 0$ for all (i,j) from some basis. Taking $u_o = 0$, we can calculate step by step all others u_i from the equation

$\rho_{ij} + u_i - u_j = 0$.

Because ρ_{ij} are integers, all u_i are integers.

Remark.

Note, we can suppose $u_o = 0$ without any loss of generality, because if we substract a constant from the function (u_i), we don't change the array d_{ij}.

Lemma 2.

For every two vertices $i = (i_o,\dots,i_{n-1})$, $j = (j_o,\dots,j_{n-1})$ of the graph G the inequality

(5) $|u_i - u_j| \leq \rho_{ij}$

holds for any potential (u_i).

Proof.

If the vertices i,j are neighbour, then the condition $d_{ij}=0$ is equivalent to the following :

(6) $u_i - u_j = \pm 1$.

Inequality (5) is an another form of $d_{ij} \geq 0$. The condition (6) is implied by the equality $\rho_{ij} = 1$ for all (i,j) such that i,j are vertices of the same edge.

<u>Theorem</u>

A function $(u_i : i=0,1,\dots,s^n-1)$ is a potential iff two following conditions are satisfied :

a) $|u_i - u_j| \leq \rho_{ij}$ for all $i,j=0,1,\dots,s^n-1$,

b) there exists a partial graph D of G which is a tree and such that for every edge (i,j) of D $|u_i - u_j| = 1$.

Proof.

First we are going to show that the condition a) and b) imply that (u_i) is a potential. Given a tree D which is a partial graph of G , the set of the cells corresponding to the edges of D together with the cells $(i,i),(i=0,1,\dots,s^n-1)$ form a basis because the number of these cells is $2s^n-1$ and they are independent i.e. they do not form a closed loop. In fact, suppose the cells

$$(i_o,i_1),(i_1,i_2),\dots,(i_{k-1},i_k) = (i_o,i_k)$$

form a cycle and that $(i_j,i_{j+1}) \in D$ whenever $i_j \neq i_{j+1}$. It is clear that vertices $i_o,i_1,\dots,i_k$ from a cycle in D . This is impossible, because D is a tree. Now, conditions (2),(3) are obviously fullfiled.

For necessity of conditions a), b) suppose B is a basis of cells with the property $d_{ij} = 0$ for all $(i,j) \in B$. We need to build a tree D for which a) and b) hold. Let H be a graph with vertices $0,1,\dots,s^n-1$, and edges joining exactly those vertices i,j for which the cell (i,j) belongs to B . H is a connected graph, because every cell in the matrix of our transportation problem may be included in a cycle formed by the edges of the graph H . Define a graph H_1 as follows. An edge $(i,j) \in H_1$ iff it lies on a shortest path in G from k to l , where $(k,l) \in H$. Of course, H_1 is a connected graph. We'll show, that $d_{ij} = 0$ for $(i,j) \in H_1$. Let

$$(k,i_1) = (i_o,i_1) \ , \ (i_1,i_2),\dots,(i_{p-1},i_p) = (i_{p-1},l)$$

be a shortest path in G from k to l and let $(k,l) \in H$. Since $d_{kl} = 0$ and $\rho_{kl} = p$, $|u_k-u_l| = p$. On the other hand,

$$|u_k-u_l| \le \sum_{r=0}^{r=p-1} |u_{i_r} - u_{i_{r+1}}| \le \sum_{r=0}^{r=p-1} \rho(i_r,i_{r+1})$$

and as $\rho(i_r,i_{r+1}) = 1$, $|u(i_r) - u(i_{r+1})| = 1$ for any edge (i_r,i_{r+1}) $(r=0,1,\dots,p-1)$. Thus we have shown that $\rho_{ij} = 1$ and $|u_i - u_j| = 1$ for any edge (i,j) of the connected graph H_1 . Now for D we can take any partial graph of H_1, which is a tree.

Corollary :

In the case $s = 2$ a potential is any fonction $(u_i : i=0,1,\dots,s^n-1)$ such that $|u_i - u_j| = 1$ for every pair of neighbour vertices of the graph G.

Proof.

If a partial graph D is a tree, then every two vertices i,j of G may be connected by means of a path in D. The length of the path is an odd number, if i,j are neighbour.

As $u_i - u_j = \sum(u(k_r) - u(k_{r+1}))$, where (k_r,k_{r+1}) are edges of the path, $|u(k_r) - u(k_{r+1})| = 1$ and $|u_i - u_j| \leq \rho_{ij} = 1$, we have $|u_i - u_j| = 1$.

4. A process of generating of all potentials in the case $s = 2$.

We consider a potential as a function, which domain is the set of all vertices of the graph G. This set is a union of sets $A_0,A_1,\dots,A_n$ corresponding to the number of 1's. Precisely

$A_i = \{(\alpha_1,\dots,\alpha_n) : \sum_{j=1}^{j=n} \alpha_j = i\}$ $(i=0,1,\dots,n)$.

It is clear that any two vertices k,l joined by an edge in the graph G belong to A_i and A_{i+1} /or conversly/ for a suitable $i=0,1,\dots,n-1$

The following procedure generates all possible potentials for a given n.

(1) Take $u_0 = 0$.

(2) For $i \in A_1$ define $u_i = 1$ or $u_i = -1$

(3) For $r = 2,3,\dots,n$ and for $i \in A_r$ examine the set $U_i = \{u_j : j \in A_{r-1}, \rho_{ij}=1\}$.

If $U_i = \{k\}$, then take $u_1 = k-1$ or $u_i = k+1$.

If $U_i = \{k-1, k+1\}$, than take $u_i = k$.

Proof of correctness of the procedure.

It is obvious that any function defined by this procedure has even values on the sets A_i with even i and odd values on the A_i with odd i.

We will be done if we show that in the step (3) of the procedure only two possibilities can occur i.e. that always

(7) $U_i = \{k\}$ or $U_i = \{k-1 , k+1\}$

for a suitable k.

We shall prove this by induction on r. Suppose, (7) holds for $r \leq p-1$ and there exist $j_1 , j_2 , j_3 \in A_{p-1}$ such that $\rho_{ij} = 1$ for $j = j_1 , j_2 , j_3$ and $u(j_1) \leq k-2$, $u(j_2) = k$, $u(j_3) \geq k+2$ for some k. Only one position, say m_l, distinguishes j_l from $i : j_l(m_l) \neq i(m_l)$ $(l=1,2,3)$. Without loss of generality one can assume that $m_l = l$ $(l=1,2,3)$. Then $i = (111\ldots)$, $j_1 = (011\ldots)$, $j_2 = (101\ldots)$, $j_3 = (110\ldots)$, where the symbols replaced by dots, the same in all j's, are inessential. Now consider $k_1 = (001\ldots)$, $k_2 = (010\ldots)$, $k_3 = (100\ldots) \in A_{p-2}$. One of k_l, namely k_2, is joined with j_1 and j_3 and on the $(p-1)$-th step U_{k_2} contains two numbers one of wich is $\leq k-2$ and the second on is $\geq k+2$. This is a contradiction to the induction hypotesis.

5. Example showing that $\bar{d}_n \neq d_n$

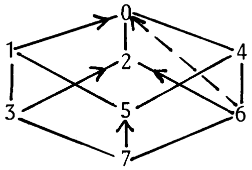
Let $s=2$, $n=3$. We have

$\mu_1 = (0.07, 0.11, 0.02, 0.20, 0.11, 0.11, 0.20, 0.18)$,

$\mu_2 = (0.12, 0.08, 0.10, 0.15, 0.08, 0.17, 0.15, 0.15)$.

We get the transportation problem given by Table 1 with a solution given by Table 2. We shall show this solution is optimal. It is sufficient to find a potential for which $d_{ij} = 0$ for $(i,j) = (1,0)$, $(3,2)$, $(4,5)$, $(6,0)$, $(6,2)$, $(7,5)$.

We represent these cells by arrows in the diagram



It is easy to see, that if we take $u_o = 0$, then we have $u_1=u_2=u_4=-1$, $u_5=u_6=-2$, $u_5=0$, $u_7=1$. d_{ij} are given in the Table 3. All d_{ij} _ 0 and for all cells of Table 2 occupied by numbers 0 we have $d_{ij} = 0$. Thus the solution is optimal. Moreover, every optimal solution has the form given in Table 4. /A basis is formed of cells (i,i)(i=0,1,...,7), (i,0)(i=0,1,2,3,4,6), (4,5), (6,7)/.

Now we shall show that no optimal solution can be an invariant measure. Conditions for a measure σ_{ij} to be invariant in this case are following :

Table 1.

15	3	2	2	1	2	1	1	0
15	2	3	1	2	1	2	0	1
17	2	1	3	2	1	0	2	1
8	1	2	2	3	0	1	1	2
15	2	1	1	0	3	2	2	1
10	1	2	0	1	2	3	1	2
8	1	0	2	1	2	1	3	2
12	0	1	1	2	1	2	2	3
	7	11	2	20	11	11	20	18

Transportation problem

Table 2.

15								15
15							15	3
17					3	11		
8					8			
15				15				
			2	5			3	
8		8						
12	7	3					2	
	7	11	2	20	11	11	20	18

A optimal solution

Table 3.

1	4	2	2	0	2	2	0	0
2	4	4	2	2	2	4	0	4
0	2	0	2	0	0	0	0	2
1	2	2	2	2	0	2	0	4
2	4	2	2	0	4	4	2	4
1	2	2	0	0	2	4	0	4
1	2	0	2	0	2	2	2	4
0	0	0	0	0	0	2	0	4
	0	-1	-1	-2	-1	0	-2	-1

Valuations d_{ij}

Table 4.

15				$\sigma 37$			$\sigma 67$	$\sigma 77$
15							$\sigma 66$	
17		$\sigma 15$		$\sigma 35$	$\sigma 45$	$\sigma 55$	$\sigma 65$	
8					$\sigma 44$		$\sigma 64$	
15				$\sigma 33$				
10		$\sigma 12$	$\sigma 22$				$\sigma 62$	
8		$\sigma 11$	$\sigma 31$					
12	$\sigma 00$	$\sigma 10$	$\sigma 20$	$\sigma 30$	$\sigma 40$		$\sigma 60$	

General optimal solution

(8) $\sigma(x,y) + \sigma(x+4,y) + \sigma(x,y+4) + \sigma(x+4,y+4) =$

$\sigma(2x,2y) + \sigma(2x+1,2y) + \sigma(2x,2y+1) + \sigma(2x+1,2y+1)$

for $x,y = 0,1,2,3$.

From these conditions for $(x,y) = (1,3),(2,3)$ we get $\sigma_{37} = 0$ and $\sigma_{67} = 0$ respectively and hence $\sigma_{77} = 15$, $\sigma_{75} = 0$. From the same condition for $(x,y) = (0,1)$, $(0,2)$, $(1,2)$, $(2,1)$ one obtains $\sigma_{45} = \sigma_{15} = \sigma_{35} = \sigma_{65} = 0$ and from that $\sigma_{75} = 6$. The contradiction shows that no invariant measure σ is an optimal solution of our transportation problem. Therefore, in this example $\tilde{d}_3 > \bar{d}_3$.

6. Applications

Let μ_p and μ_q be two Bernoulli measures on X given by two probability vectors $p = (p_1,\dots,p_s)$, $q = (q_1,\dots,q_s)$ Suppose that $p_i \le q_i$ $(i=1,2,\dots,k)$ and $p_i \ge q_i$ $((i=k+1,\dots,s)$. Define a Bernoulli measure r on $X = X \times X$ with a vector $\overline{r} = (r_{ij})(i,j=1,2,\dots,s)$ such that

$$r_{ii} = p_i \ (i=1,\dots,k)\ ,\ r_{ii} = q_i \ (i=k+1,\dots,s)$$

and the remaining r_{ij} arbitrarily preserving only conditions $\sum_i r_{ij} = p_j$, $\sum_j r_{ij} = q_i$ $(j,i=1,\dots,s)$. It is clear that $r_{ij} = 0$ if $i \neq j$ and $i = 1,\dots,k$ or $j = k+1,\dots,s$. Futhermore, it is easy to verify that $p_1 r_n = (\mu_p)_n$, $p_2 r_n = (\mu_q)_n$ for $n=1,2,\dots$ and

$$k_n(r_n) = k_1(r_1) = \sum_{i \neq j} r_1(i,j) = \sum_{i=k+1}^{i=s} |p_i - q_i| = \frac{1}{2} \sum_{i=1}^{i=s} |p_i - q_i|$$

as r is an invariant measure on Y .

It follows from definition 2 that $\overline{d}(\mu_p,\mu_q) \le \frac{1}{2} \sum_{i=1}^{i=s} |p_i - q_i|$. Since $d_1((\mu_p)_1)$, $((\mu_q)_1) = \frac{1}{2} \sum_{i=1}^{i=s} |p_i - q_i|$ and $\overline{d}(\mu_p,\mu_q) \ge d_1((\mu_p)_1,(\mu_q)_1)$ we have $\overline{d}(\mu_p,\mu_q) = \frac{1}{2} \sum_{i=1}^{i=s} |p_i - q_i|$. Taking into account potentials it is possible to find that $d_n((\mu_p)_n$, $((\mu_q)_n)$ is exactly equal to $\frac{1}{2} \sum_{i=1}^{i=s} |p_i - q_i|$.

Now we'll show that $k_n(r_n) = d_n(\mu_p,\mu_q)(n=1,2,\dots)$

Let u be a potential defined as follows :

$$u(i_o,\dots,i_{u-1}) = |\{j : k+1 \le i_j \le s\}|.$$

/If $s=2$ u is the number of 1's in the sequence $i_o,\dots,i_{n-1}$/

and let $S_o = \{(i,j) : i=j \text{ or } k+1 \le i \le s \text{ and } 1 \le j \le k\}$. Then the set Y_n of zeros of u may be characterized by the equivalence

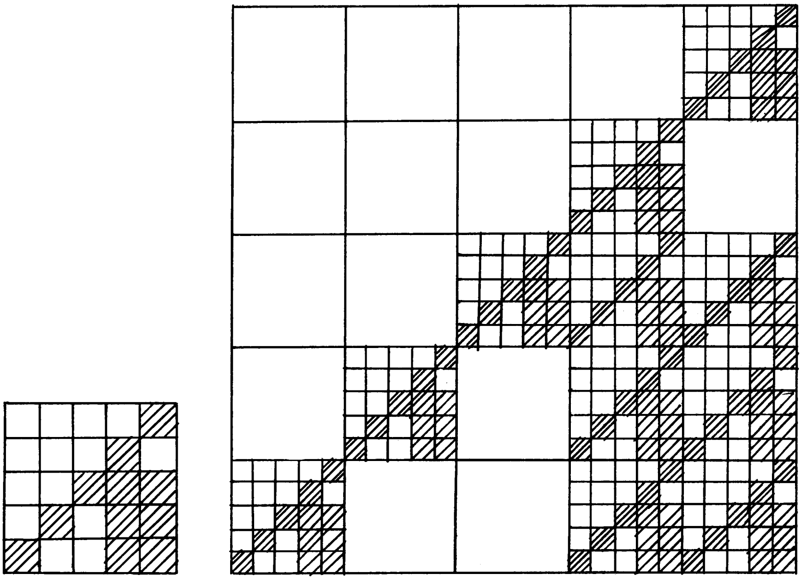
$$(b,c) \in Y_n \quad \text{iff} \quad (b_i,c_i) \in S_o \ ,$$

where $b = (b_o,\dots,b_{n-1})$, $c = (c_o,\dots,c_{n-1}) \in X_n$. For $n = 1$ and $n = 2$, $s = 5$ Y_n is the set cancelled on the diagram 2. It is easy to see that the measure r_n is concentrated on Y_n . Thus

$k_n(r_n) = d_n$ and therefore

$$d_n = \frac{1}{2}\sum_{i=1}^{i=s} |p_i - q_i| = \bar{d}(\mu_p,\mu_q) \quad .$$

Diagram 2



The set Y_i . Cases n=1 and n=2 . /s=5 and k=3/.

In case s=2 the same potential can be used to determine the Ornstein's distance between two Markov measures. Let μ,ν be two Markov measures given by

$$p = (p_0,p_1) \quad , \quad P = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix}$$

$$q = (q_0,q_1) \quad , \quad Q = \begin{bmatrix} q_{00} & q_{01} \\ q_{10} & q_{11} \end{bmatrix}$$

respectively. We'll define a Markov measure $\bar{\mu}$ on the set

$Y = \prod_{-\infty}^{\infty} \{0,1\} = \{0,1\}^{Z}$ such that

(9) $p_1 \overline{\mu}_n = \mu_n$, $p_2 \overline{\mu}_n = \nu_n$.

In order to define the measure $\overline{\mu}$ it is sufficient to give the transition matrix $\tilde{R}$, because the initial vector $\tilde{r}$ can be obtained from the equation $\tilde{r}\tilde{R} = \tilde{r}$. First we define an operation on probabilities vectors. When $p = (p_0,p_1)$, $q = (q_0,q_1)$ are such two vectors, let $v = p \vee q$ denotes a vector $v = (v_{00},v_{01},v_{11})$ with $v_{ii} = \min(p_i,q_i) (i=0,1)$, $v_{01} = p_0-v_{00}$, $v_{10} = p_1-v_{11}$. We have $v_{00}+v_{01} = p_0$, $v_{10}+v_{11} = p_1$, $v_{00}+v_{10} = q_0$, $v_{01}+v_{11} = q_1$. Let us denote

$$\overline{p}_i = (p_{i0},p_{i1})$$
$$\overline{q}_i = (q_{i0},q_{i1})$$
$$\overline{r}(i,j) = \overline{p}_i \vee \overline{q}_i$$

for $i,j = 0,1$. Now the matrix $\tilde{R} = (r((i,j),(k,l)))$ is defined in such a way that the (i,j) - th row of R is equal to $\overline{r}(i,j)$.

A proof of the equalities (9) requires some preparations. Let $S = \{0,1,\dots,s-1\}$, $X = \prod_{-\infty}^{\infty} S$ and μ be a stationary Markov measure on X given by the probability vector $r = (r_0,\dots,r_{s-1})$ and the transition matrix $R = (r_{ij})_{(0\le i,j\le s-1)}$. Further, let $\overline{S} = (\alpha_1\dots\alpha_t)$, $t \le s$ is a partition of S,$Y = \prod_{-\infty}^{\infty} \overline{S}$ and $P:X\to X$ is the mapping defined by $(P(x))_i = \alpha_i$ for $x_i \in \alpha_j$. Then we can define a stationary measure ν on Y as follows

$$\nu(B) = \mu(P^{-1}(B)) \quad \text{for } B \subset Y .$$

Lemma 3.

If $\sum_{j\in\beta} r_{ij} = \sum_{j\in\beta} r_{kj}$, whenever $i,k \in \alpha$, then ν is a Markov measure.

Proof.

Let μ_n, ν_n be the measures on X_n and Y_n determined by respectively, where $X_n = \prod_0^{n-1} S$, $Y = \prod_0^{n-1} \overline{S}$. Put $P(r) = (r'_{\alpha_1}, r'_{\alpha_2}, \ldots, r'_{\alpha_t})$, where $r' = \sum_{i\varepsilon\alpha} r_i$ and $P(R) = \tilde{r}_{\alpha\beta}$, $(\alpha,\beta\varepsilon\overline{S})$ with $\tilde{r}_{\alpha\beta} = \sum_{j\varepsilon\beta} r_{ij}$, $i\varepsilon\alpha$. We show by induction on n that $P\mu_n = \nu_n$. It is easy to check this for $n=1,2$.

Suppose $P\mu_n = \nu_n$ for some $n \geq 2$. Then for $C = (c_0,\ldots,c_n) \varepsilon Y_{n+1}$ we have

$$\nu_{n+1}(C) = \sum_{i_0\varepsilon c_0} \sum_{i_1\varepsilon c_1} \cdots \sum_{i_n\varepsilon c_n} \mu_{n+1}(i_0,\ldots,i_n) =$$

$$= \sum_{i_0} \sum_{i_1} \cdots \sum_{i_n} r_{i_0} r_{i_0 i_1} \cdots r_{i_{n-1} i_n} =$$

$$= \sum_{i_0} \sum_{i_1} \cdots \sum_{i_{n-1}} r_{i_0} r_{i_0 i_1} \cdots r_{i_{n-2} i_{n-1}} \sum_{i_n\varepsilon c_n} r_{i_{n-1} i_n} =$$

$$= \tilde{r}_{c_{n-1} c_n} \sum_{i_0} \sum_{i_1} \cdots \sum_{i_{n-1}} r_{i_{n-2} i_{n-1}} =$$

$$= \tilde{r}_{c_{n-1} c_n} \nu_n(c_0,\ldots,c_{n-1}) .$$

This means that ν is a Markov measure given by the probability vector $P(r)$ and the transition matrix $P(R)$.

Now we are able to prove the equalities (9).

Let $S = \{(00),(01),(11)\}$ and let $\overline{S}_1 = \{\overline{\alpha}_0,\overline{\alpha}_1\}$, $\overline{S}_2 = \{\overline{\beta}_0,\overline{\beta}_1\}$ be the partitions of S defined as follows $\overline{\alpha}_0 = \{(00),(01)\}$, $\overline{\alpha}_1 = \{(10),(11)\}$, $\overline{\beta}_0 = \{(00),(10)\}$, $\overline{\beta}_1 = \{(01),(11)\}$. One can verify that the partitions $\overline{S}_1,\overline{S}_2$ satisfy the conditions of lemma 3 and that $P_1(\tilde{r}) = \overline{p}$, $P_1(\tilde{R}) = P$, $P_2(\tilde{R}) = \overline{q}'$ $P_2(\tilde{R}) = Q$, and P_1,P_2 are the mappings determined by $\overline{S}_1,\overline{S}_2$ respectively. Thus we obtain $P_1(\overline{\mu}) =$, $P_2(\overline{\mu}) = \nu$.

Now analizing use of Lemma 3 and definition 2 we conclude that $\bar{d}(\mu,\nu) \leq r_{01}+r_{10}$. Moreover, if $p_{00},p_{10} \leq < q_{00},q_{10}$ then it is easy to see that measure $\bar{\mu}_n$ is concentrated on the set of zeros of the potential u given above. Consequently $\bar{d}(\mu,\nu) = r_{01}+r_{10}$ and then $\bar{d}(\mu,\nu) = q_0-p_0$, as $p_0 \leq q_0$ and $r = (p_0,0,q_0-p_0,q_1)$.

References :

[1] D.S. Ornstein, An application of ergodic theory to probability theory. Ann. Prob., vol.1 (1973), 43-65.

[2] R.M. Gray, D.L. Neuhoff, P.C. Shields, A generalisation of Ornstein's $\bar{d}$ distance with applications to information theory. Ann. Prof., vol. 3 (1975), 315-328.

[3] A.M. Vershik, Nieskolko zamieczanij o biezkoniechnych zadaczach liniejnogo programmirowanija, Usp. Mat. Nauk, 25 wyp. 5 (1970) 117-124.

[4] D.S. Ornstein, Russian translation of [1], Usp. Mat. Nauk, 30 wyp. 3 , (1975), 123-146.

J. Kwiatkowski
F. Maniakowski
Institute of Mathematics
ul. Chopina 12/18
87-100 Toruń
Poland