The Annals of Probability
1985, Vol. 13, No. 2, 630–637

MUTUAL DEPENDENCE OF RANDOM VARIABLES AND MAXIMUM DISCRETIZED ENTROPY

By Carlo Bertoluzza[1] and Bruno Forte[2]

Università di Pavia and University of Waterloo

In connection with a random vector (X, Y) in the unit square Q and a couple (m, n) of positive integers, we consider all discretizations of the continuous probability distribution of (X, Y) that are obtained by an $m \times n$ cartesian decomposition of Q. We prove that Y is a (continuous and invertible) function of X if and only if for each m, n the maximum entropy of the finite distributions equals $\log(m + n - 1)$

In [2] a criterion for the discretization of a continuous random n-vector has been suggested. Based on the maximum entropy method, it presents indeed some advantages with respect to the methods previously introduced (see [1] and [2] for a full bibliography on the subject). In verifying the method, optimal bounds for the maximum discretized entropy have been found ([1] and [3]). The (best) upper bound in [3] is nothing but one of the "natural" properties of Shannon's entropy. Here we analyze the (best) lower bound derived in [1]. This bound is not trivial since it holds only under certain regularity assumptions on the probability distribution. These results are significant in pattern recognition. In digital image processing via a (black and white) video screen, for example, quantization is unavoidable. The image is a set of black or white squares covering the screen. The distribution of the black squares (pixels) on the screen defines a quantized probability distribution of a random 2-vector (X, Y). From this viewpoint the results of the present paper can be interpreted as follows:

a) If for some couple (m, n) of positive integers a discretization of the screen into $m \cdot n$ rectangles has entropy greater than $\log(m + n - 1)$ then the random variables X, Y are either functionally unrelated or if related, relation is *not* invertible.

b) Conversely, if for every discretization into $m \cdot n$ rectangles the entropy is not greater than $\log(m + n - 1)$ then the "picture" is a perfectly connected quantized line representing a monotone function [4].

1. Introduction.

Let (X, Y) be a real-valued random 2-vector in the unit square $Q = \{(x, y): 0 \le x \le 1, 0 \le y \le 1\}$; that is (X, Y) is an ordered couple of real-valued random variables such that

$$\text{Prob}\{X < 0\} = \text{Prob}\{X \ge 1\} = 0$$

$$\text{Prob}\{Y < 0\} = \text{Prob}\{Y \ge 1\} = 0.$$

Received March 1983; revised September 1984.

[1] Research supported in part by G.N.I.M. (C.N.R.—Italy).

[2] Research supported in part by N.S.E.R.C. Grant A 7677.

AMS 1980 *subject classifications.* Primary 60E99; secondary 62–07, 62B10.

Key words and phrases. Entropy, discretization, mutual dependence.

As usual we denote by $F(x, y) = \text{Prob}\{X < x, Y < y\}$ the probability distribution of (X, Y) and by $F_1(x) = F(x, 1)$, $F_2(y) = F(1, y)$ its marginal distributions.

For each m, n in the set of all positive integers $\mathbb{N}$ and each couple of sequences of real numbers

$$S_m := \{0 = s_0 < s_1 < \cdots < s_m = 1\}$$

$$T_n := \{0 = t_0 < t_1 < \cdots < t_n = 1\}$$

we consider the cartesian decomposition $S_m \times T_n$ of Q into mn rectangles

$$R_{i,j} := \{(x, y): s_i \leq x < s_{i+1},\ t_j \leq y < t_{j+1}\},$$

where $i = 0, 1, \cdots, m - 1$ and $j = 0, 1, \cdots, n - 1$. As in [1], the finite probability distribution

$$\pi(F, m, n) := \{\pi_{i,j}: i = 0, \cdots, m - 1, j = 0, \cdots, n - 1\}$$

with

$$\pi_{i,j} := F(s_{i+1}, t_{j+1}) + F(s_i, t_j) - F(s_i, t_{j+1}) - F(s_{i+1}, t_j)$$

is a "discretization" of the continuous probability distribution of the real-valued random vector (X, Y). Shannon's entropy of the discretized distribution, or simply the discretized entropy, is given by

$$H[\pi(F, m, n)] = -\textstyle\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \pi_{i,j} \log \pi_{i,j} \quad (0 \log 0 := 0)$$

where the log is taken in any fixed base.

It has been shown in [1] that if $F(x, y)$ is continuous and $x \to F(x, y)$, $y \to F(x, y)$ are strictly increasing then for each $m, n \in \mathbb{N}$, the maximum value of the discretized entropy satisfies the following inequalities

$$\log(m + n - 1) \leq \text{Max}_{\pi(F,m,n)} H[\pi(F, m, n)] \leq \log(mn) \tag{1}$$

where on the right-hand side, equality holds for all $m, n \in \mathbb{N}$, if and only if X and Y are stochastically independent $[F(x, y) = F_1(x) \cdot F_2(y)]$ [2].

Here, as in [1] and [2], we restrict ourselves to the case of a random 2-vector. In the case of a random p-vector ($p \geq 2$), inequalities (1) read as follows

$$\log(m_1 + m_2 + \cdots + m_p - p + 1) \leq \text{Max}_{\pi(F,m_1,\cdots,m_p)} H[\pi(F, m_1, \cdots, m_p)]$$
$$\leq \log(m_1 \cdot m_2 \cdot \cdots \cdot m_p).$$

A question arises, naturally. When is the left-hand side of (1) satisfied by equality for all $m, n \in \mathbb{N}$?

The answer to this question in a certain class of random 2-vectors (X, Y) will be given in the form of Theorem 2 in Section 3 of the present paper. But first in the form of Theorem 1 we shall give another proof of the l.h.s. of (1). The proof will be simpler than the one that can be found in [1] and it will provide a way for a better understanding of the proof of Theorem 2.

2. The lower bound for the discretized entropy. We assume $F(x, y)$ to be such that both $F_1(x)$ and $F_2(y)$ are strictly increasing and continuous. For all such random 2-vectors the following theorem holds,

THEOREM 1. *For each* $m, n \in \mathbb{N}$,

$$\text{(2)} \qquad \text{Max}_{\pi(F,m,n)} H[\pi(F, m, n) \geq \log(m + n - 1).$$

PROOF. Construct a partition α of the square Q into $m + n - 1$ rectangles by drawing horizontal and vertical lines, according to the following rules (see Figure 1).

(a) Draw $n - 1$ horizontal lines through the points $(0, t_j)$, $j = 1, 2, \cdots, n - 1$, such that $F_2(t_j) - F_2(t_{j-1}) = (m + n - 1)^{-1}$ $(t_0 := 0)$, this being possible (in an unique manner) since $F_2(y)$ is continuous (and strictly increasing).

(b) Draw $m - 1$ vertical lines down to the horizontal $y = t_{n-1}$, through the points $(s_i, 1)$, $i = 1, 2, \cdots, m - 1$, so that

$$F_1(s_i) + F(s_{i-1}, t_{n-1}) - [F_1(s_{i-1}) + F(s_i, t_{n-1})] = (m + n - 1)^{-1}$$

$(s_0 := 0)$. This is always possible since the function $x \to F_1(x) - F(x, t_{n-1})$ is continuous, hence it takes on all the values between 0 and $1 - F_2(t_{n-1}) = m/(m + n - 1)$.

Each element (atom) of the partition α has probability $1/(m + n - 1)$. The entropy of this uniform probability distribution equals $\log(m + n - 1)$.

On the other hand, by extending the vertical lines down to meet the horizontal

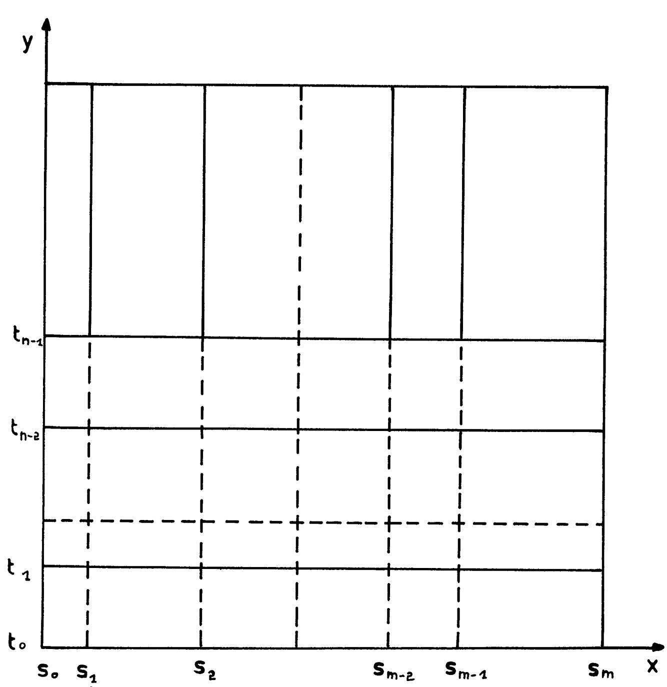

FIG. 1. *Partition* α *of the square* Q.

x (see Figure 1) one generates a decomposition $\pi(F, m, n)$ of the square Q into mn rectangles. This cartesian decomposition is a refinement of α, hence

$$H[\pi(F, m, n)] \geq \log(m + n - 1),$$

which yields (2).

The following section is devoted to the proof of the main result.

3. Mutual dependence and maximum entropy. Under the same assumptions of Theorem 1, that is

(i) $x \to F_1(x)$, $y \to F_2(y)$ continuous on $[0, 1]$,
(ii) $x \to F_1(x)$, $y \to F_2(y)$ strictly increasing on $[0, 1]$,

we have

THEOREM 2. *If for each $m, n \in \mathbb{N}$*

$$\mathrm{Max}_{\pi(F,m,n)} H[\pi(F, m, n)] = \log(m + n - 1) \tag{3}$$

then there exists a continuous and invertible function ϕ: $[0, 1] \to_{\text{onto}} [0, 1]$ such that $Y = \phi(X)$. The converse is also true.

In other words (3) is the necessary and sufficient condition for Y to be functionally dependent on X in the class of random vectors (X, Y) that verify (i) and (ii).

PROOF. Let $\bar{\gamma}$ be the set of all those points (x_0, y_0) in Q, such that for every rectangle

$$R(\varepsilon, \eta) := \{(x, y): |x - x_0| < \varepsilon, |y - y_0| < \eta\}$$

the probability of $R(\varepsilon, \eta) \cap Q$ is positive. The set $\bar{\gamma}$ is clearly closed. To prove the first part of the theorem we have just to prove that $\bar{\gamma}$ is the graph on a continuous and invertible function ϕ: $[0, 1] \to_{\text{onto}} [0, 1]$.

Since $x \to F_1(x)$ is strictly increasing and $\bar{\gamma}$ is closed, it is easy to recognize that for each $\bar{x}$ in $[0, 1]$ there exists one $\bar{y}$ in $[0, 1]$ such that $(\bar{x}, \bar{y}) \in \bar{\gamma}$. Suppose that there are two such numbers $\bar{y}$: $\bar{y}_1 < \bar{y}_2$. $F_2(y)$ being strictly increasing, at least one of the following inequalities holds true

$$F(\bar{x}, \bar{y}_2) - F(\bar{x}, \bar{y}_1) > 0 \tag{4}$$

$$F_2(\bar{y}_2) + F(\bar{x}, \bar{y}_1) - F(\bar{x}, \bar{y}_2) - F_2(\bar{y}_1) > 0. \tag{5}$$

If (5) is satisfied, let

$$\begin{aligned} a &:= F_2(\bar{y}_2) + F(\bar{x}, \bar{y}_1) - F(\bar{x}, \bar{y}_2) - F_2(\bar{y}_1) \\ &= \mathrm{Prob}\{(x, y): \bar{x} \leq x \leq 1, \bar{y}_1 \leq y \leq \bar{y}_2\} \end{aligned}$$

and choose $n \in \mathbb{N}$, so that $n^{-1} < a/3$.

Then choose $m \in \mathbb{N}$, $m \geq 2$, to satisfy

$$(6) \quad \begin{array}{ll} (m-2)(m+n-1)^{-1} < F_1(\bar{x}) < (m-1)(m+n-1)^{-1} & \text{for} \quad \bar{x} > 0 \\ m = 2 & \text{if} \quad \bar{x} = 0. \end{array}$$

This is always possible since for all $m \geq 2$, $n \in \mathbb{N}$

$$F_1(\bar{x}) < 1 - n^{-1} < 1 - (m + n - 1)^{-1}.$$

According to the two diagrams in Figure 2, construct the partition β of Q, by exchanging horizontals with verticals in the rules we followed to construct the partition α.

At least for one of the numbers t_i we have

$$\bar{y}_1 < t_i < \bar{y}_2,$$

since $a = \text{Prob}(A) > 2n^{-1} \geq 2(m + n - 1)^{-1}$. The entropy associated to partition β equals $\log(m + n - 1)$. As before, the cartesian decomposition $\pi(F, m, n)$ generated by β (see dotted lines in diagram (b) of Figure 2) has entropy not less than $\log(m + n - 1)$ since it is obtained by further partitioning the rectangles of β. But in the present case we are sure that $H[\pi(F, m, n)]$ is strictly greater than $\log(m + n - 1)$. In fact the points $(\bar{x}, \bar{y}_1)$ and $(\bar{x}, \bar{y}_2)$ belong to two separate rectangles (of positive probability) in $\pi(F, m, n)$. Hence by going from β to $\pi(F, m, n)$ we have divided an atom of β into at least two atoms in $\pi(F, m, n)$ of probability strictly greater than zero, thus increasing the entropy. But

$$H[\pi(F, m, n)] > \log(m + n - 1)$$

contradicts the hypothesis of the theorem, hence $(\bar{x}, \bar{y}_1) \in \bar{\gamma}$ and $(\bar{x}, \bar{y}_2) \in \bar{\gamma}$ imply $\bar{y}_1 = \bar{y}_2$.

If the inequality (4) is the one which is satisfied, the procedure is quite similar. In constructing the partition β one goes left to right instead of going right to left. The special case $\bar{x} = 0$ is now the case $\bar{x} = 1$.

In the same manner, by interchanging x with y one can prove that for each $\bar{y} \in [0, 1]$ there exists at least one $\bar{x} \in [0, 1]$ such that $(\bar{x}, \bar{y}) \in \bar{\gamma}$, and $(\bar{x}_1, \bar{y})$, $(\bar{x}_2, \bar{y}) \in \bar{\gamma}$ imply $\bar{x}_1 = \bar{x}_2$. Thus there exists an invertible function ϕ: $[0, 1] \rightarrow_{\text{onto}} [0, 1]$ such that $Y = \phi(X)$.

The set $\hat{\gamma}$ being closed, it contains each of the following points

$$(\bar{x}, \liminf_{x \to \bar{x}-} \phi(x)), \quad (\bar{x}, \limsup_{x \to \bar{x}-} \phi(x)) \quad \text{for all} \quad 0 < \bar{x} \leq 1$$

$$(\bar{x}, \liminf_{x \to \bar{x}+} \phi(x)), \quad (\bar{x}, \limsup_{x \to \bar{x}+} \phi(x)) \quad \text{for all} \quad 0 \leq \bar{x} < 1$$

and, of course $(\bar{x}, \phi(\bar{x}))$. But the point on $\bar{\gamma}$ with $x = \bar{x}$ is unique, hence

$$\begin{aligned} \liminf_{x \to \bar{x}-} \phi(x) &= \limsup_{x \to \bar{x}-} \phi(x) \\ &= \liminf_{x \to \bar{x}+} \phi(x) = \limsup_{x \to \bar{x}+} \phi(x) = \phi(\bar{x}) \end{aligned}$$

for all $\bar{x}$ in the open interval (0, 1), and

$$\liminf_{x \to 0+} \phi(x) = \limsup_{x \to 0+} \phi(x) = \phi(0)$$

$$\liminf_{x \to 1-} \phi(x) = \limsup_{x \to 1-} \phi(x) = \phi(1).$$

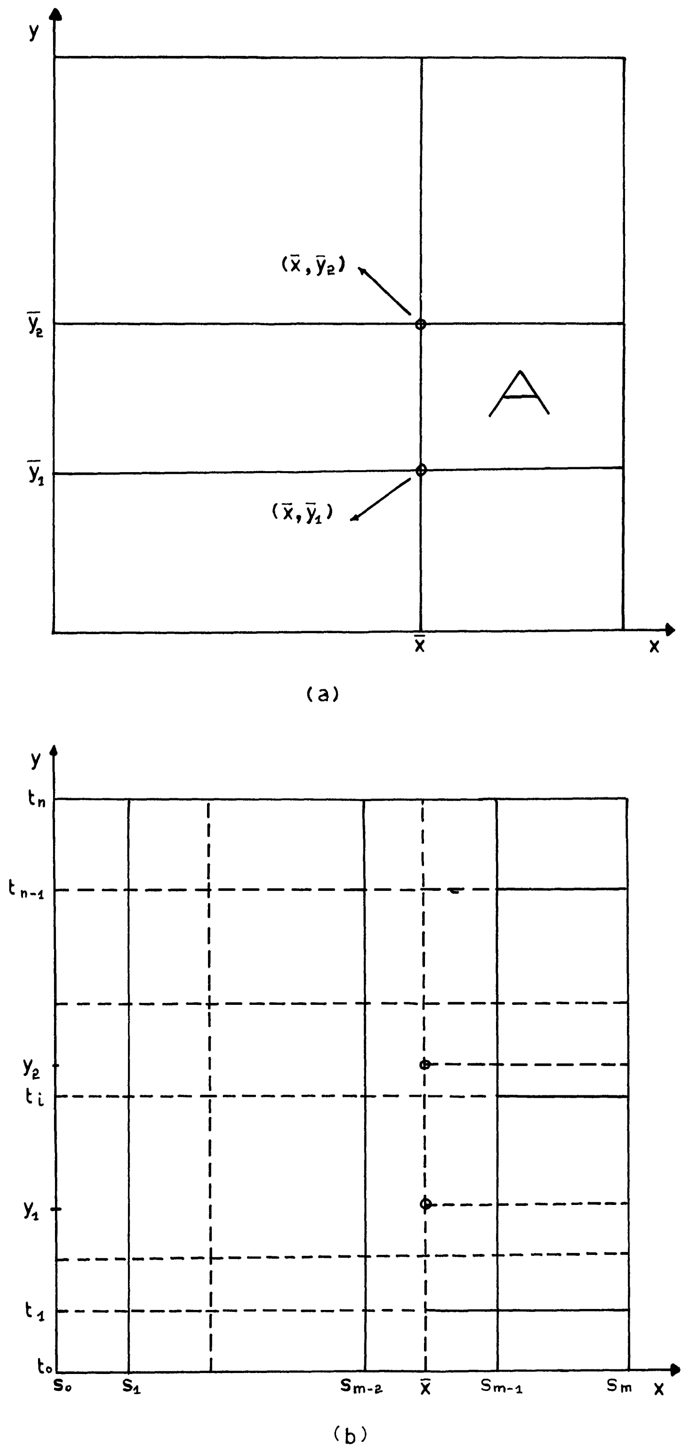

FIG. 2. *Partition β of the square Q.*

Hence, ϕ is continuous on the closed interval [0, 1].

The function ϕ, being continuous and bijective, is either strictly increasing or strictly decreasing.

Conversely, assume $Y = \phi(X)$ with ϕ: $[0, 1] \to_{\text{onto}} [0, 1]$, continuous and either

(a) strictly increasing, or
(b) strictly decreasing.

In the case (a) it is easy to recognize that

$$\bar{\gamma} = \gamma := \{(x, y) \in Q: F_1(x) = F_2(y)\}$$

and consequently $F(x, y) = \text{Inf}\{F_1(x), F_2(y)\}$. Then, by Lemma 2 in [1] we have

$$\text{Max}_{\pi(F,m,n)} H[\pi(F, m, n)] = \log(m + n - 1)$$

for all $m, n \in \mathbb{N}$.

In the case (b)

$$\bar{\gamma} = \gamma' := \{(x, y) \in Q: F_1(x) = 1 - F_2(y)\}.$$

Moreover

$$F(x, y) = \text{Sup}(0, F_1(x) + F_2(y) - 1).$$

Thus case (b) is symmetrical to case (a), namely reduces to case (a) by the change of variables $x' = x$, $y' = 1 - y$. Thus again

$$\text{Max}_{\pi(F,m,n)} H[\pi(F, m, n)] = \log(m + n - 1)$$

for all $m, n \in \mathbb{N}$.

4. Conclusion and final comments. We would like to point out that Theorem 1 still holds if we do not impose *strict* monotonicity to $F_1(x)$ and $F_2(y)$.

Note also that Theorem 2 can be immediately extended to a random p-vector $(X_1, X_2, \cdots, X_p)$, $p \in \mathbb{N}$, X_i real-valued random variables in the interval [0, 1]. With the said assumptions on the marginal distributions, X_i and X_j are mutually dependent $(i, j = 1, 2, \cdots, p)$ if and only if

$$\text{Max}_{\pi(F,m_1,\cdots,m_p)} H[\pi(F, m_1, \cdots, m_p)] = \log(m_1 + \cdots + m_p - p + 1)$$

for all $m_1, m_2, \cdots, m_p \in \mathbb{N}$.

However the p-dimensional case has a larger variety of dependence than the two-dimensional one. Further investigations are needed. A degree of dependence could be based on the values of $\text{Max}_{\pi(F,m,n)} H[\pi(F, m, n)]$. This will be the subject of future investigations.

REFERENCES

[1] FORTE, B., DE LASCURAIN, M. and WONG, A. K. C. (1984). The best lower bound of the maximum entropy for discretized two-dimensional probability distributions. Unpublished manuscript.

[2] DE LASCURAIN, M. (1983). On maximum entropy discretization and its applications in pattern recognition. Ph.D. dissertation, Faculty of Engineering, University of Waterloo, Ontario, Canada.

[3] FORTE, B., and PINTACUDA, N. (1984). Stochastic independence: from pattern recognition a different characterization. *Boll. Un. Mat. Ital.* **6** 3-A 119–123.

[4] HUNG, S. H. Y. and KASVAND, T. (1983). Linear approximation of quantized thin lines. NATO ASI series, Vol. 4, Pictorial Data Analysis 15–28, Springer-Verlag,

DIPARTIMENTO DI
INFORMATICA E SISTEMISTICA
UNIVERSITÀ DI PAVIA
CORSO STRADA NUOVA 65
27100 PAVIA, ITALY

DEPARTMENT OF APPLIED MATHEMATICS
FACULTY OF MATHEMATICS
UNIVERSITY OF WATERLOO
ONTARIO N2L 3G1, CANADA