INTERPOLATION OF FUNCTIONS OVER A MEASURE SPACE AND CONJECTURES ABOUT MEMORY

by

Andrzej Ehrenfeucht*
Jan Mycielski**

*Department of Computer Science
**Department of Mathematics
UNIVERSITY OF COLORADO, BOULDER, COLORADO

Report #CU-CS-009-72 December, 1972

Dr. Ehrenfeucht's work was supported by NSF Grant GJ-660.
Dr. Mycielski's work was supported in part by NSF Grant GP-19405.

1. Introduction

Let X be a set and f a bounded real valued function over X. Let $N \subseteq X$ be a finite set and suppose that f restricted to N (in symbols $f \restriction N$) is known. We will define an algorithm to estimate $f(x)$ for any $x \in X$. Let a collection $\underline{C}$ of subsets of X be given such that $\underline{C}$ covers X.

For any bounded non empty set S of real numbers, $\operatorname{mid} S = \frac{1}{2}(\sup S + \inf S)$ and $\operatorname{diam} S = \sup S - \inf S$.

Algorithm. Given $x \in X$ we choose $C \in \underline{C}$ such that $x \in C$, $C \cap N \neq \emptyset$ and $\operatorname{diam} f(C \cap N)$ is small and we estimate $f(x)$ as $\operatorname{mid} f(C \cap N)$.

This is a familiar procedure if X is a metric space, f is continuous and all sets $C \in \underline{C}$ have small diameters. But we are interested e.g. in the case $X = [0,1]^{30}$. In this case every covering $\underline{C}$ of X with sets of diameters $\leq 1/n$ contains more than n^{30} sets (in other words the entropy of X is high, see [8]) and the Algorithm will not work unless N has at least n^{30} elements (otherwise N could not intersect all sets of a subcovering of $\underline{C}$). Thus our assumption that $f \restriction N$ is known entails the storage of an enormous amount of information. It is the purpose of this paper to discuss stronger suppositions on f and $\underline{C}$ which imply that the Algorithm works and allow for smaller $\underline{C}$ and N.

In our case it is more natural to assume a probability measure μ over X and 'small measure' will play the rôle of 'small diameters' of the sets in $\underline{C}$.

This setting suggests other algorithms related to statistical estimation procedures. E.g. choose $C \in \underline{C}$ with $x \in C$ such that the estimated variance of f over C is small. Then estimate $f(x)$ as the estimated mean of f over C. One could also think of algorithms using several (or all) $C \in \underline{C}$ with $x \in C$ and estimate $f(x)$ as some weighted mean of the estimated means of f over the C's (weights could be functions of the estimated variances of f over the C's). (See Remark 5 in Section 3 for some references related to such ideas; see also [13] for Stone-Weierstrass-type approximations to measureable functions).

But in this paper we will consider only the simple Algorithm stated at the beginning. In Sections 2 and 3 we prove some theorems about it. The remaining section is a study of some finite functions which we call k-continuous and for which the Algorithm is efficient.

Our motivation for this work were attempts to imagine a mechanism having certain properties of the brain in particular its learning and recognition ability. In Section 3 Remark 3 we state a conjecture on the learning mechanism of the brain. This conjecture says that learning neurons use an interpolation algorithm as above.

We are indebted to R. McKenzie, W. Taylor and S. M. Ulam for many discussions concerning the subject of this paper. In particular the idea of the Algorithm is partly due to Taylor. The first part of Theorem 24 is due to Mrs. F. Yao.

Theorem 9 was announced in [12].

2. General Theorems

Let $\varepsilon \geq 0$ and let I_t^ε be the closed interval $[t - \varepsilon, t + \varepsilon]$; in particular $I_t^0 = \{t\}$.

<u>Lemma 1</u>. If $A \cap f^{-1}(I^{\varepsilon}_{f(x)}) \neq \emptyset$ then $|f(x) - \text{mid}\, f(A)| \leq \varepsilon + \frac{1}{2}\text{diam}\, f(A)$.

<u>Proof</u>. Choose $y \in A \cap f^{-1}(I^{\varepsilon}_{f(x)})$. Then

$$|f(x) - \text{mid}\, f(A)| \leq |f(x) - f(y)| + |f(y) - \text{mid}\, f(A)|$$
$$\leq \varepsilon + \tfrac{1}{2}\text{diam}\, f(A)\ .$$

Q.E.D.

Let μ be a probability measure over X and let f and all $C \in \underline{C}$ be μ-measurable.

Let the sequence $x_1, \ldots, x_n, x \in X$ be choosen at random. We put $N = \{x_1, \ldots, x_n\}$. Thus N is a random variable over the probability measure space $\langle X^n, \mu^n\rangle$.

Let K be a relation over, i.e. a subset of, the space $X^n \times R^n \times X \times \underline{C}$.

We define

$$P(f, \underline{C}, K, n, \varepsilon) = \text{Probability}\ \{C \cap N \cap f^{-1}(I^{\varepsilon}_{f(x)}) \neq \emptyset \ \text{ for all } C \in \underline{C} \text{ such that } (x_1, \ldots, x_n, f(x_1), \ldots, f(x_n), x, C) \in K\}\ .$$

By Lemma 1 we get immediately the following

<u>Theorem 2</u>. With probability not less than $P(f, \underline{C}, K, n, \varepsilon)$ the inequality

$$(1) \qquad |f(x) - \text{mid}\, f(C \cap N)| \leq \varepsilon + \tfrac{1}{2}\text{diam}\, f(C \cap N)$$

is true for all $C \in \underline{C}$ with $(x_1, \ldots, x_n, f(x_1), \ldots, f(x_n), x, C) \in K$.

This theorem is still too general to have practical importance since P may be close to 1 by the mere fact that the probability of the existence of any $C \in \underline{C}$ such that $(x_1, \ldots, x_n, f(x_1), \ldots, f(x_n), x, C) \in K$

is very small. On the other hand one may have some K's free from this defect. In fact the only K considered in this paper is as follows $(\ldots,x,C) \in K$ iff $x \in C$. Thus, since $\underline{C}$ covers X, the above objection does not apply. (It is possible however that other K's are interesting, especially K's involving a condition $\operatorname{card}(C \cap N) \geq s$.)

Let

$$P_o(f,\underline{C},n,\varepsilon) = \text{Probability}\ \{C \cap N \cap f^{-1}(I^{\varepsilon}_{f(x)}) \neq \emptyset \quad \text{for all } C \in \underline{C} \text{ such that } x \in C\}.$$

By Theorem 2 (or directly from Lemma 1) we get the following information on the algorithm.

<u>Corollary 3</u>. With probability not less than $P_o(f,\underline{C},n,\varepsilon)$ the inequality (1) is true for every $C \in \underline{C}$ with $x \in C$.

The following basic Lemma will be used in our estimates of P_o.

Let $\underline{D}$ be a finite collection of μ-measurable subsets of X. We put

$$d = \operatorname{card}(\underline{D})$$

and

$$\mu_o = \min\{\mu(D) : D \in \underline{D}\} .$$

<u>Lemma 4</u>. The probability that $N \cap D \neq \emptyset$ for every $D \in \underline{D}$ is not less than

$$1 - d(1 - \mu_o)^n .$$

<u>Proof</u>. Let $s(N)$ be the number of sets $D \in \underline{D}$ which are not intersected by N. Clearly the expected value of $s(N)$ is $\leq d(1 - \mu_o)^n$. Since $s(N) = 0$ or $s(N) \geq 1$ therefore the probability that $s(N) = 0$ is $\geq 1 - d(1 - \mu_o)^n$. Q.E.D.

Let now $X_o \subseteq X$ be μ-measurable and for every $x \in X_o$ let $\underline{D}(x)$ be a collection of μ-measurable subsets of X such that for every $C \in \underline{C}$ with $x \in C$ there exists a $D \in \underline{D}(x)$ with $D \subseteq C \cap f^{-1}(I^{\varepsilon}_{f(x)})$. We put

$$d_o = \max\{\mathrm{card}(\underline{D}(x)) : x \in X_o\}$$

and

$$\mu_o = \inf\{\mu(D) : D \in \underline{D}(x), x \in X_o\}.$$

<u>Theorem 5</u>. $P_o(f,\underline{C},n,\varepsilon) \geq \mu(X_o)(1 - d_o(1-\mu_o)^n)$.

<u>Proof</u>. Clearly

$$\begin{aligned} P_o(f,\underline{C},n,\varepsilon) &\geq \text{Probability } \{x \in X_o \text{ and} \\ &\qquad N \cap D \neq \emptyset \text{ for every } D \in \underline{D}(x)\} \\ &\geq \mu(X_o)(1 - d_o(1-\mu_o)^n), \end{aligned}$$

the last inequality following from Lemma 4. Q.E.D.

In the next Section we shall consider a more concrete situation, with $\varepsilon = 0$, and define $\underline{D}(x)$ so that Corollary 3 and Theorem 5 will yield interesting estimates.

Let now

$$\begin{aligned} Q_o(f,\underline{C},n,\varepsilon) = \text{Probability } \{&C \cap N \cap f^{-1}(I^{\varepsilon}_{f(x)}) \neq \emptyset \\ &\text{for all } x \in X \text{ and all } C \in \underline{C} \text{ with } x \in C\} \end{aligned}$$

The following theorem is analogous to Corollary 3 (a similar analog of Theorem 2 would be also possible) and follows immediately from Lemma 1.

<u>Theorem 6</u>. With probability not less then $Q_o(f,\underline{C},n,\varepsilon)$ the inequality (1) is true for all $x \in X$ and all $C \in \underline{C}$ with $x \in C$.

Let now $\underline{D}(x)$ and μ_o be as in Theorem 5. We put

$$d_1 = \operatorname{card}(\bigcup_{x \in X} \underline{D}(x)).$$

Theorem 7. $Q_o(f,\underline{C},n,\varepsilon) \geq \mu(X_o)(1 - d_1(1-\mu_o)^n)$.

The proof is similar to that of Theorem 5.

3. Interpolation over $\{0,1\}^m$.

$\{0,1\}^m$ denotes the set of all sequences of 0's and 1's of length m. Let $k \leq m$.

A k-cylinder in $\{0,1\}^m$ is any set $C \subseteq \{0,1\}^m$ which is of the form

$$C = \{(\xi_1,\ldots,\xi_m) : (\xi_{i_1},\ldots,\xi_{i_k}) = (c_1,\ldots,c_k)\},$$

where $1 \leq i_1 < \ldots < i_k \leq m$ and $(c_1,\ldots,c_k) \in \{0,1\}^k$. We put also $B(C) = \{i_1,\ldots,i_k\}$; $\underline{C}_k$ denotes the family of all k-cylinders.

Let f be a function with domain $X \subseteq \{0,1\}^m$ and μ a probability measure over X. We shall say that f is k-continuous if X can be covered with a collection $\underline{C}$ of k-cylinders such that $f \restriction (C \cap X)$ is a constant for every $C \in \underline{C}$. (See Section 4 for examples of such functions.)

We put

$$\mu_f = \min\{\mu(C \cap f^{-1}\{f(x)\}) : C \in \underline{C}_k \text{ and } x \in C \cap X\}$$

Theorem 8. If $x_1,\ldots,x_n,x \in X$ are chosen at random then, with probability not less than $1 - \binom{m}{k}(1-\mu_f)^n$,

$$f(x) = v$$

for every C and v such that $x \in C \in \underline{C}_k$ and $f(x_i) = v$ for all $x_i \in C \cap \{x_1, \ldots, x_n\}$.

<u>Proof</u>. For all $x \in X$ we put $\underline{D}(x) = \{C \cap f^{-1}\{f(x)\} : x \in C \in \underline{C}_k\}$. Then $\mathrm{card}(\underline{D}(x)) \leq \binom{m}{k}$. Hence Theorem 8 follows from Corollary 3 and Theorem 5 for $X_o = X$ and $\varepsilon = 0$.

<u>Remark 1</u>. Although Theorem 8 is valid without any assumptions on f, it is more interesting for k-continuous f's since for such f's there are $C \in \underline{C}_k$ with $x \in C$ and $f \upharpoonright C$ being a constant. Moreover the probability that $C \cap N \neq \emptyset$ for any such C is $(1 - \mu(X \cap C))^n$.

f will be called <u>regular</u> <u>k-continuous</u> if, for every r in the range of f, $f^{-1}\{r\}$ is a union of k-cylinders. (See Section 4 for examples of such functions.) Let μ be the probability measure over X defined by

(2) $\mu(Y) = \mathrm{card}(Y)/\mathrm{card}(X)$, for all $Y \subseteq X$.

<u>Theorem 9</u>. If f is regular k-continuous, μ is defined by (2), C and v are as in Theorem 8 then $f(x) = v$ with probability not less than

$$1 - \binom{m-k}{k}(1 - 4^{-k})^n.$$

<u>Proof</u>. For every $x \in X$ let $x \in C_x \in \underline{C}_k$, $C_x \subseteq f^{-1}\{f(x)\}$, and $\underline{D}(x) = \{C \cap C_x : x \in C \in \underline{C}_k$ and $B(C) \cap B(C_x) = \emptyset\}$. Clearly $\underline{D}(x)$ satisfies the condition preceding Theorem 5 and $\mathrm{card}(\underline{D}(x)) = \binom{m-k}{k}$.

Also $\mu(D) = 4^{-k}$ for every $D \in \underline{D}(x)$. Thus Theorem 9 follows from Corollary 3 and Theorem 5 with $X_o = X$ and $\varepsilon = 0$.

In practice it may be more useful to formulate Theorems 8 and 9 as follows

Corollary 10. (i) Under the suppositions of Theorem 8 the probability that $f(x) \neq v$ is $\leq p$ if

(3) $$n \geq \frac{\log\binom{m}{k} - \log p}{-\log(1-\mu_f)} .$$

(ii) Under the suppositions of Theorem 9 the probability that $f(x) \neq v$ is $\leq p$ if

(4) $$n \geq \frac{\log\binom{m-k}{k} - \log p}{-\log(1-4^{-k})} .$$

Remark 2. We think that Corollary 10 and Theorem 13 (see below) indicate that the Algorithm is applicable in some situations (a difficulty is pointed out in Remark 10 at the end of this paper). Although the estimates (3) and (4) depend very much on μ_f and k respectively (since $-\log(1-\alpha) \approx \alpha$ for small α) still for some f it may happen that the true values of n which secure the required p are much smaller than the above estimates.

Let $n(m,p,k)$ be the least integer n which satisfies (4). We have the following table of values of $n(m,p,k)$.

m	k \ p	1/20	1/100	1/1000
200	1	29	35	43
	2	200	225	261
	3	1082	1185	1331
	4	5340	5751	6340
	5	25098	26746	29102
	6	114452	121043	130473
	7	511155	537523	575248
500	1	33	38	46
	2	229	254	289
	3	1259	1361	1508
	4	6294	6705	7293
	5	29898	31545	33902
	6	137614	144206	153636
	7	619810	646179	683903
1000	1	35	41	49
	2	250	275	311
	3	1392	1494	1640
	4	7008	7419	8008
	5	33481	35128	37485
	6	154859	161450	170880
	7	700469	726837	764561

3. Perhaps the learning neurons in the brain learn in fact k-continuous Boolean (i.e. two-valued) functions f with small k (or functions of some related class). They store a sequence $x_1,\ldots,x_n$, $f(x_1),\ldots,f(x_n)$ or some information extracted from this sequence (where $x_i \in \{0,1\}^m$ and m is the number of inputs of the neuron) and then estimate $f(x)$ using the Algorithm with $\underline{C} = \underline{C}_k$ or some related algorithm. It is not clear how the values $f(x_i)$ are taught to the neuron but one can imagine various mechanisms for such self-teaching of the brain. All this suggests studying nets built from k-continuous

Boolean functions. For some information on such nets see [3] and [9], but learning nets of this sort have not yet been studied.

Is it so that some neurons in the central nervous system are k-continuous Boolean functions with small k (say $k < 10$)? (Neurons usually have hundreds of inputs and probably depend on most of them). In theory one could try to prove this checking the predictability of the activity of a neuron, from its past activity, applying our Algorithm.

4. It is not clear, although it seems probable, that k-continuous and regular k-continuous functions constitute the natural domain of applications of the Algorithm. But those are the only interesting (simple enough) classes of functions related to the Algorithm which we know. We shall study them in the following sections of this paper.

5. There exist other functions (different from k-continuous ones) depending on may variables for which efficient interpolation algorithms are known. It seems that these algorithms are all closely related to linear approximation theory, like the least squares method, the Monte Carlo methods (see [6], Chapter 12, [15] and [16]), the perceptron learning theorem and equalizing algorithms (see [10] and [11]). Some of them yield small mean square errors rather than uniform approximations like the Algorithm of this paper.

6. Lemma 4 implies the following

<u>Proposition 11</u>. If the elements $x_1, \ldots, x_n \in \{0,1\}^m$ are chosen at random then, with probability not less than $1 - 2^k \binom{m}{k} (1 - 2^{-k})^n$ $\{x_1, \ldots, x_n\}$ intersects every k-cylinder.

Let $n(m,k)$ be the minimal number n such that there exists a set $\{x_1, \ldots, x_n\} \subseteq \{0,1\}^m$ intersecting every k-cylinder. Clearly

Proposition 11 implies that $n(m,k) \le n$ if $2^k \binom{m}{k} (1 - 2^{-k})^n < 1$. This was proved by J. H. Spencer [14], Theorem 2.3.1. We do not know any sharper estimate of $n(m,k)$ unless $k = 2$ or $m - 1$. Of course $n(2,2) = 4$, and if $m > 2$ then $n(m,2)$ is the least integer n such that

[n/2] $\binom{n-1}{[\frac{n}{2}] - 1} \ge m$. Ralph McKenzie remarked that this follows from Erdös, Ko and Rado [4], Theorem 1, see also [7], if one uses the following obvious lemma: If M is a 01-matrix with m columns which are characteristic functions of a collection of m sets such that no two are included in one another, each two intersect and the complements of each two intersect, then the set of rows of M intersects every 2-cylinder in $\{0,1\}^m$. He noticed also that $n(m,m-1) = 2^{m-1}$.

7. For other applications of probability to combinatorics, see [5] and [14]. Another application of Lemma 4 is the following

<u>Proposition 12</u>. If $f_i : \{1,\ldots,m\} \to \{1,\ldots,k\}$ are functions chosen at random for $i = 1,\ldots,n$ then, with probability not less than $1 - \binom{m}{k}\left(1 - \frac{k!}{k^k}\right)^n$, we have

> (*) for every set $A \subseteq \{1,\ldots,m\}$ with k elements there is an $i \in \{1,\ldots,n\}$ such that f_i restricted to A is one-to-one.

Let $n(m,k)$ be the minimal n such that there exists $f_1,\ldots,f_n$ as in Proposition 12 satisfying (*). Clearly Proposition 12 implies that $n(m,k) \le n$ if $\binom{m}{k}(1 - \frac{k!}{k^k})^n < 1$. Again (as in Remark 6) we do not know any sharper estimate of $n(m,k)$ unless $k = 2$. It is easy to check that $n(m,2)$ is the least integer not less than $\frac{\log m}{\log 2}$.

The following Theorem follows from Theorems 6 and 7 in the same way in which Theorems 8 and 9 followed from Corollary 3 and Theorem 5.

Theorem 13. If f is regular k-continuous, μ is defined by (2) and $x_1,\dots,x_n \in X$ are chosen at random then with probability not less than

$$1 - 4^k\binom{m}{2k}(1 - 4^{-k})^n$$

$f(x) = v$ for every $x \in X$ and every v such that there exists a $C \in \underline{C}_k$ with $x \in C$ and $f(x_i) = v$ for all $x_i \in C \cap \{x_1,\dots,x_n\}$.

Proof. Let, for every $x \in X$, $\underline{D}(x) = \{C : x \in C \in \underline{C}_{2k}\}$. Hence for every $x \in X$ and every $C \in \underline{C}_k$ with $x \in C$ there exists a $D \in \underline{D}(x)$ such that $D \subseteq C \cap f^{-1}\{f(x)\}$. Clearly for every $D \in \underline{D}(x)$, $\mu(D) \geq 4^{-k}$ and

$$\operatorname{card}\Big(\bigcup_{x \in X} \underline{D}(x)\Big) \leq \operatorname{card}(\underline{C}_{2k}) = 4^k\binom{m}{2k}.$$

Hence Theorem 13 follows from Theorems 6 and 7 with $X_o = X$ and $\varepsilon = 0$.

Remark 8. The estimate

$$(5) \qquad n \geq \frac{\log\binom{m}{2k} + k\log 4 - \log p}{-\log(1 - 4^{-k})},$$

similar to Corollary 10 (ii), which follows from Theorem 13 is not much worse than (4). E.g. if $n(m,k,p)$ is the smallest integer satisfying (5) we have the following table of values of $n(m,k,p)$

m	k \ p	1/20	1/100	1/1000
200	1	50	56	64
	2	369	393	429
	3	2051	2153	2299
	4	10267	10678	11266
	5	48696	50343	52700
	6	223490	230082	239512
	7	1002989	1029357	1067081
500	1	57	62	70
	2	426	451	486
	3	2403	2505	2651
	4	12161	12537	13161
	5	58215	59863	62219
	6	269356	275947	285377
	7	1217777	1244145	1281869
1000	1	61	67	75
	2	469	494	529
	3	2668	2770	2916
	4	13585	13997	14585
	5	65356	67004	69360
	6	303694	310286	319716
	7	1378274	1404643	1442367

4. k-continuous functions

k-continuous and regular k-continuous functions are defined prior to Theorem 8 and Theorem 9 respectively. We shall change the notation in this respect that, for any $x \in \{0,1\}^m$, x_i will be the i-th coordinate of x, thus $x = (x_1,\dots,x_m)$.

If $X \subseteq \{0,1\}^m$ and f is a function with domain X we shall say that f depends on the variable x_i if there are $x,y \in X$ such that $x_j = y_j$ for all $j \neq i$ but $f(x) \neq f(y)$.

Our main interest will be in the question on how many variables can a k-continuous or regular k-continuous function depend.

<u>Example</u>. The following function $f : \{0,1\}^7 \to \{0,1\}$ is regular 3-continuous

$$f(x_1,\dots,x_7) = \begin{cases} x_4 & \text{if } x_1 = 0 \text{ and } x_2 = 0, \\ x_5 & \text{if } x_1 = 0 \text{ and } x_2 = 1, \\ x_6 & \text{if } x_1 = 1 \text{ and } x_3 = 0. \\ x_7 & \text{if } x_1 = 1 \text{ and } x_3 = 1. \end{cases}$$

<u>Proposition 14</u>. For every integer $m > 1$ there are 2-continuous functions $f : X \to \{0,1\}$ where $X \subseteq \{0,1\}^m$ depending on all m variables.

<u>Proof</u> (due to D.B. Thompson). Let X be the set of all sequences $(\underbrace{0,0,\dots,0}_{i},\underbrace{1,1,\dots,1}_{m-i})$, where $i \in \{0,1,\dots,m\}$, and $f(x) \equiv i \pmod 2$. It is easy to see that f is 2-continuous and depends on all its m variables.

Let $\varphi(k)$ be the maximum number of variables on which a regular k-continuous Boolean (i.e., two-valued) function may depend and $\varphi_0(k)$ the maximum m for which there are k-continuous functions $f: \{0,1\}^m \to \{0,1\}$ depending on all m variables.

<u>Theorem 15</u>. $2k + \binom{2k}{k} \le \varphi_0(k+1) \le \varphi(k+1) \le (2k+1)4^k$. *

This theorem follows from Propositions 16 and 17 and Theorems 23 and 24 proved below. It shows that regular k-continuity is a much stronger condition than k-continuity.

* See Theorems 17A and 25A added in proof.

Proposition 16. If f is a k-continuous function with domain $\{0,1\}^m$ then f is regular k-continuous.

Proposition 17. There are $(k+1)$-continuous functions $f : \{0,1\}^m \to \{0,1\}$, where $m = 2k + \binom{2k}{k}$, depending on all m variables.

Proof. Let $M = K \cup \{A : A \subseteq K,\ \mathrm{card}(A) = k\}$, where $\mathrm{card}(K) = 2k$. Hence $\mathrm{card}(M) = m$. Let $x \in \{0,1\}^M$ i.e. $x : M \to \{0,1\}$. Now we define $f : \{0,1\}^M \to \{0,1\}$ as follows: 1^o if $\mathrm{card}\{i \in K : x(i) = 0\} > k$ then $f(x) = 0$; 2^o if $\mathrm{card}\{i \in K : x(i) = 1\} > k$ then $f(x) = 1$; 3^o if $A = \{i \in K : x(i) = 0\}$ and $\mathrm{card}(A) = k$ then $f(x) = x(A)$. It is not hard to check that f is $(k+1)$-continuous and depends on all m variables.

Problem. We do not know if the number $2k + \binom{2k}{k}$ in Proposition 17 is maximal nor if $\varphi_0(k) < \varphi(k)$ for some k. (See Theorem 17A).

Proposition 18. A function $f : \{0,1\}^m \to \{0,1\}$ is k-continuous iff f can be represented as a disjunction of conjunctions of variables and negations of variables each conjunction having no more than k terms and also as a conjunction of disjunctions of variables and negations of variables each disjunction having no more than k terms.

Proposition 19. If f_i is k_i-continuous with domain $X_i \subseteq \{0,1\}^m$ and range R_i for $i = 1,\dots,n$ and g is any function with domain $\underset{i=1}{\overset{n}{P}} R_i$ then $f(x) = g(f_1(x),\dots,f_n(x))$ is a $(k_1 + \dots + k_n)$-continuous function with domain $\bigcap_{i=1}^{n} X_i$.

Proposition 20. If f is a (regular) k-continuous function with domain $X \subseteq \{0,1\}^m$ then $g(x) = f(\pi(x) + c)$ is (regular) k-continuous with domain $\pi^{-1}(X - c)$ where π is any permutation of coordinates, + denotes vector addition in $\{0,1\}^m$ treated as a vector space over the Galois field GF(2), and c is any vector in $\{0,1\}^m$.

Proposition 21. If f and g are regular k-continuous and ℓ-continuous functions respectively with the same domain X, $N \subseteq X$, N intersects every $(k + \ell)$-cylinder included in X and $f \upharpoonright N = g \upharpoonright N$ then $f = g$.

Proof. Let $x \in X$. Choose a k-cylinder C_1 and an ℓ-cylinder C_2 such that $x \in C_1 \subseteq X$, $x \in C_2 \subseteq X$, and $f \upharpoonright C_1$ and $g \upharpoonright C_2$ are constants. Since $C_1 \cap C_2$ includes a $(k + \ell)$-cylinder it contains an element $y \in N$. Since $f(y) = g(y)$ it follows that $f(x) = g(x)$.

Theorem 22. Given a set $X \subseteq \{0,1\}^m$ which is a union of k-cylinders such that X includes exactely d 2k-cylinders, and a set R, there are no more than $d^{4^k \log \operatorname{card}(R)}$ regular k-continuous functions $f : X \to R$.

Proof. By Lemma 4 if $x_1, \ldots, x_n$ are chosen at random in X then with probability not less than $1 - d(1 - 4^{-k})^n$ the set $N = \{x_1, \ldots, x_n\}$ intersects every 2k-cylinder included in X. Hence if $d(1 - 4^{-k})^n < 1$ i.e.

$$n > \frac{\log d}{-\log(1 - 4^{-k})}$$

then there exists a set N with n elements at most which intersects every 2k-cylinder included in X. Therefore, since $4^k \log d >$

$(\log d)/(-\log(1-4^{-k}))$, and by Proposition 21, to define a regular k-continuous function $f : X \to R$ it is enough to fix the values of f over a set N with no more than $4^k \log d$ elements. This can be done in no more than $(\mathrm{card}(R))^{4^k \log d} = d^{4^k \log \mathrm{card}(R)}$ ways. Q.E.D.

<u>Problem</u>. Improve the bound given in Theorem 22. (Cf. Theorem 15.)

Proving a conjecture of K. Kuratowski, Mrs. Calczyńska-Karlowicz [2] found the following lemma.

(6) For every positive integer k there exists a positive integer $\varkappa$ such that if $\underline{A}$ and $\underline{B}$ are two collections of k-element sets, such that $A \cap B \neq \emptyset$ for every $A \in \underline{A}$ and $B \in \underline{B}$, then there exists a set M with $\varkappa$ elements at most such that $M \cap A \cap B \neq \emptyset$ for every $A \in \underline{A}$ and $B \in \underline{B}$.

Theorem 24 proved below is a refinement of (6).

Let $\varkappa(k)$ be the smallest $\varkappa$ satisfying (6) and $\varphi(k)$ as defined prior to Theorem 15.

<u>Theorem 23</u>. $\varphi(k) = \varkappa(k)$.

<u>Proof</u>. $\varphi(k) \geq \varkappa(k)$. Let $\underline{A}$ and $\underline{B}$ be two collections of k-element sets and M a $\varkappa(k)$-element set which is minimal such that $M \cap A \cap B \neq \emptyset$ for every $A \in \underline{A}$ and $B \in \underline{B}$. We define two unions of k-cylinders

$$F_0 = \bigcup_{A \in \underline{A}} \{x \in \{0,1\}^M : x(j) = 0 \quad \text{for all} \quad j \in M \cap A\},$$

$$F_1 = \bigcup_{B \in \underline{B}} \{x \in \{0,1\}^M : x(j) = 1 \quad \text{for all} \quad j \in M \cap B\}.$$

It is clear that $F_0 \cap F_1 = \emptyset$. We put $X = F_0 \cup F_1$ and define $f : X \to \{0,1\}$ putting $f^{-1}(0) = F_0$ and $f^{-1}(1) = F_1$.

Thus f is regular k-continuous.

To see that f depends on all its $\varkappa(k)$ variables let $i \in M$. Hence since M is minimal there are $A \in \underline{A}$, $B \in \underline{B}$ such that $M \cap A \cap B = \{i\}$. Let

$$x(j) = \begin{cases} 0 & \text{for } j \in M \cap A, \\ 1 & \text{for } j \in M - A, \end{cases}$$

and $y(j) = x(j)$ for $j \neq i$ and $y(i) = 1$. Hence $y(j) = 1$ for all $j \in M \cap B$. Thus $f(x) = 0$ and $f(y) = 1$ but x and y differ only at the i-th coordinate.

Therefore $\varphi(k) \geq \varkappa(k)$.

$\varphi(k) \leq \varkappa(k)$. Let $f : X \to \{0,1\}$ be regular k-continuous and X be a union of k-cylinders in $\{0,1\}^{\varphi(k)+t}$ and let f depend on $\varphi(k)$ variables $x_1, \ldots, x_{\varphi(k)}$. For each k-cylinder C in $\{0,1\}^{\varphi(k)+t}$ we put

(7) $$F(C) = a,$$

where $a : B(C) \to \{0,1\}$ is such that

$$C = \{x \in \{0,1\}^{\varphi(k)+t} : x \upharpoonright B(C) = a\}$$

(hence a is a function and is a set of k ordered pairs). Let $1 - a : B(C) \to \{0,1\}$ be defined by $(1-a)(i) = 1 - a(i)$ for all $i \in B(C)$. We put

$$\underline{A} = \{F(C) : C \subseteq X,\ C \text{ is a k-cylinder},\ f(C) = \{0\}\}.$$

$$\underline{B} = \{1 - F(C) : C \subseteq X, \quad C \text{ is a k-cylinder}, \quad f(C) = \{1\}\}.$$

We have $A \cap B \neq \emptyset$ for each $A \in \underline{A}$ and $B \in \underline{B}$ since otherwise there would be a k-cylinder $C_0 \subseteq X$ with $f(C_0) = \{0\}$ and a k-cylinder $C_1 \subseteq X$ with $f(C_1) = \{1\}$ such that $F(C_0) \cup F(C_1)$ is a function. But then $C_0 \cap C_1 \neq \emptyset$, which is a contradiction.

Now we will show that if $M \cap A \cap B \neq \emptyset$ for each $A \in \underline{A}$ and $B \in \underline{B}$ then for every $i \in \{1, \ldots, \varphi(k)\}$ there is a pair $\langle i, b \rangle$, where $b \in \{0,1\}$, which belongs to M. This will finish the proof since it implies that M has at least $\varphi(k)$ elements and hence $\varkappa(k) \geq \varphi(k)$.

Since f is k-continuous and depends on x_i for every $i \in \{1, \ldots, \varphi(k)\}$ it follows that for each such i there are two disjoint k-cylinders C_0 and C_1 such that $i \in B(C_0) \cap B(C_1)$, [redacted] $F(C_0)(i) \neq F(C_1)(i)$ and $F(C_0)(j) = F(C_1)(j)$ for every $j \in B(C_0) \cap B(C_1) - \{i\}$. Hence $F(C_0) \cap (1 - F(C_1))$ is a singleton $\{\langle i, b \rangle\}$ and $\langle i, b \rangle \in M$ since $M \cap F(C_0) \cap (1 - F(C_1)) \neq 0$. Q.E.D.

<u>Theorem 24</u>. $2k + \binom{2k}{k} \leq \varkappa(k+1) \leq (2k+1)4^k$. *

<u>Proof</u>. The first inequality is due to Frances Yao. Her proof is the following. Let K be a set with $\text{card}(K) = 2k$. Let $\underline{A} = \{A \cup \{A\} : A \subseteq K \text{ and } \text{card}(A) = k\}$ and $\underline{B} = \{(K - A) \cup \{A\} : A \subseteq K \text{ and } \text{card}(A) = k\}$. Thus for each $A \in \underline{A}$ and $B \in \underline{B}$ we have $\text{card}(A) = \text{card}(B) = k + 1$ and $A \cap B \neq \emptyset$. Also it is clear that the minimal set which intersects all intersections $A \cap B$ is

* See Theorems 17A and 25A added in proof

$$K \cup \{\{A\} : A \subseteq K \text{ and } \mathrm{card}(A) = k\},$$

which has cardinality $2k + \binom{2k}{k}$ as desired. (An alternative proof follows from Theorem 23 and Proposition 17 ~~or Theorem 17a~~).

To prove the second inequality we need the following lemma.

<u>Lemma 25</u>. Let $A_1, \ldots, A_n$, $B_1, \ldots, B_n$ be k-element sets such that $A_i \cap B_j = \emptyset$ iff $i = j$. Then

$$n \le 4^k .$$

<u>Proof</u>. Let $n(k,m)$ be the maximal n as above such that A_i and B_i satisfy the additional condition $\mathrm{card}(\bigcup_{i=1}^{n} (A_i \cup B_i)) \le m$. Thus $n(k,m) \le \binom{m}{k}$. We need the following auxiliary facts

(8) $$n(k,m) \le n(k,m+1) ;$$

(9) $$n(k,2k) = \binom{2k}{k}$$

(10) $$n(k,2(k+\ell))\binom{2\ell}{\ell} \le n(k+\ell, 2(k+\ell)) .$$

(8) and (9) are obvious. (10) is proved as follows. Let $A_i, B_i \subseteq U$, $\mathrm{card}(U) = 2(k+\ell)$, $\mathrm{card}(A_i) = \mathrm{card}(B_i) = k$ and $A_i \cap B_j = \emptyset$ iff $i = j$ for $i,j = 1,\ldots,n(k,2(k+\ell))$. Let $U_i = U - (A_i \cup B_i)$. Hence $\mathrm{card}(U_i) = 2\ell$. Let C_r^i for $r = 1,\ldots,\binom{2\ell}{\ell}$ be the sequence of all subsets of U_i having ℓ elements. We put

$$A_{ir} = A_i \cup C_r^i \quad \text{and} \quad B_{ir} = B_i \cup (U_i - C_r^i) .$$

Hence $\mathrm{card}(A_{ir}) = \mathrm{card}(B_{ir}) = k + \ell$ for all i and r, $\mathrm{card}(\bigcup (A_{ir} \cup B_{ir})) \le \mathrm{card}(U) = 2(k+\ell)$ and $A_{ir} \cap B_{js} = \emptyset$ iff $(i,r) = (j,s)$, and (10) follows.

By (9) and (10)

$$n(k,2(k+\ell)) \le \binom{2(k+\ell)}{k+\ell} \Big/ \binom{2\ell}{\ell}.$$

Since $\lim_{\ell\to\infty} \binom{2(k+\ell)}{k+\ell} \Big/ \binom{2\ell}{\ell} = 4^k$ and by (8) we get Lemma 25.

Now we conclude the proof of Theorem 24. Let $\underline{A}$ and $\underline{B}$ be collections of sets such that for every $A \in \underline{A}$ and $B \in \underline{B}$ card(A) $\le k + 1$, card(B) $\le k + 1$ and $A \cap B \ne \emptyset$. We can assume without loss of generality that for every $u \in U$, where $U = \bigcup_{A\in\underline{A}, B\in\underline{B}} (A \cup B)$, there are $A \in \underline{A}$ and $B \in \underline{B}$ such that $A \cap B = \{u\}$. Thus the proof of Theorem 24 will be completed if we show

(11) $$\text{card}(U) \le (2k+1)4^k.$$

To show this let a set $F \subseteq U$ be called free if for every $u \in F$ there are $A \in \underline{A}$ and $B \in \underline{B}$ such that $A \cap B = \{u\}$ and $(A \cup B) \cap F = \{u\}$. We shall prove first that

(12) U is a union of no more than $2k + 1$ disjoint free sets.

We shall produce a sequence $F_1, \ldots, F_{2k+1}$ of disjoint free sets covering U by assigning one by one the elements of U to the F_i. Given $u \in U$ not yet assigned let $\{u\} = A \cap B$ for some $A \in \underline{A}$ and $B \in \underline{B}$. Thus $\text{card}(A \cup B - \{u\}) \le 2k$. We assign u to any of the sets F_i which is still disjoint with $A \cup B - \{u\}$ (such an F_i exists since there are $2k + 1$ of them). If the original set F_i was free then the extended set F_i is still free. Thus (12) is proved.

(13) A free set has no more than 4^k elements.

Let F be a free set and for every $u \in F$ let $A_u \in \underline{A}$ and $B_u \in \underline{B}$ be such that $A_u \cap B_u = \{u\}$ and $(A_u \cup B_u) \cap F = \{u\}$. The systems $A_u - \{u\}$, $B_u - \{u\}$, where $u \in F$ satisfy the assumptions of Lemma 25 (except possibly that some of these sets may have less than k elements, but then they could be extended so to have exactly k). Hence $\mathrm{card}(F) \leq 4^k$ and (13) follow.

By (12) and (13) we get (11). Q.E.D.

Remarks. 9. Since $\binom{2k}{k} \sim 4^k/\sqrt{\pi k}$ it follows that the estimates of Theorem 24 are not too bad. Still in view of the next remark one would like to know more.

10. What is the best way to organize the computation of a k-continuous function f known on a sufficiently large set N? Sometimes it may be better to store the pairs $\langle F(C), b(C) \rangle$ (see formula (7)), where $f(C \cap X) = \{b(C)\}$, for a minimal set of k-cylinders C covering the domain X of f and such that $f(C \cap X) = \{0\}$ or $f(C \cap X) = \{1\}$. Then given $x \in X$, at which we want to evaluate f, we look for such $F(C)$ in this memory which satisfies $F(C) \subseteq x$, and the corresponding $b(C)$ is $f(x)$. But there may be large irredundant coverings of X with k-cylinders while very small ones exist too. How to find a small one (if it exists)? (See [1] for material somewhat related to this problem).

This question is important in view of the following difficulty of applying the Algorithm. Suppose that we have a table of $f \upharpoonright N$ for $N \subseteq \{0,1\}^{200}$, $\mathrm{card}(N) = 26{,}746$ and f is 2-valued and 5-continuous. Given $x \in \{0,1\}^{200}$, to apply the algorithm for estimating $f(x)$, we must find a 5-cylinder C containing x such that $f \upharpoonright (N \cap C)$ is a constant. But there are $\binom{200}{5} = 2,535,650,040$ 5-cylinders containing x and hence the search is rather prohibitive.

11. In view of Theorem 22 and Remark 10 it would be interesting to estimate the maximum number of k-cylinders in a minimal covering of $\{0,1\}^m$ or of any union of k-cylinders in $\{0,1\}^m$. In this respect we have the following observations made by D.B. Thompson and the referee. (1) For $m > 1$, $\{0,1\}^m$ has a minimal covering with $2m$ 2-cylinders $\{x: x_1 = x_m = v\}$ and $\{x: x_i = v, x_{i+1} = 1-v\}$, where $v = 0,1$ and $i = 1,\ldots,m-1$. (2) $\{0,1\}^m - \{(0,\ldots,0)\}$ has a minimal covering with m 1-cylinders and, if m is even, with $\frac{3}{2}m$ 2-cylinders $\{x: x_i = x_{i+m/2} = 1\}$ and $\{x: x_i = 1, x_{i+m/2} = 0\}$, where $i = 1,\ldots,m$ and + denotes addition mod m. (3) $\{x \in \{0,1\}^m: x_1 + \ldots + x_m \geq k\}$ has minimal covering with $\binom{m}{k}$ k-cylinders.

Notes added in September 1972.

1. J.H. Spencer (see [5]) proved the following theorem related to Lemma 4

Theorem. There exists a set $N \subseteq X$ such that $N \cap D \neq \emptyset$ for every $D \in \underline{D}$ and card(N) is the least integer not less than

$$\frac{\log d + 1 + \log(-\log(1-\mu_o))}{-\log(1-\mu_o)} .$$

This theorem permits to improve some estimates following Propositions 11 and 12. But his construction of this set N is not random as in Lemma 4, and hence it does not permit to improve our results, say Theorem 9.

2. A matrix similar to the 01-matrix in the proof of McKenzie in Remark 6 was used by J.H. Spencer, "Minimal completely separating systems", Journal of Combinatorial Theory, 8(1970), 446-447.

3. Proposition 17 and the first inequality of Theorems 15 and 24 can be improved as follows.

Theorem 17 A. There exist $(k+2)$-continuous functions $f\colon \{0,1\}^m \to \{0,1\}$, where $m = 2k + 4\binom{2k}{k}$, depending on all m variables.

Proof. Let K be a set with card(K) = 2k and $f_0\colon \{0,1\}^4 \to \{0,1\}$ be a 2-continuous function depending on all 4 variables (e.g. $f_0(x,y,u,v) = 0$ if $x = y = 0$ or $u = v = 0$ and $f_0(x,y,u,v) = 1$ if $1 \in \{x,y\} \cap \{u,v\}$). We put

$$M = K \cup (\{A\colon\ A \subseteq K \text{ and } \operatorname{card}(A) = k\} \times \{0,1,2,3\}) .$$

Hence card(M) = m . Let us define $f\colon \{0,1\}^M \to \{0,1\}$ as follows.

If $x \in \{0,1\}^M$ then 1° if $\mathrm{card}(\{i \in K: x(i) = 0\}) > k$ then $f(x) = 0$; 2° if $\mathrm{card}(\{i \in K: x(i) = 1\}) > k$ then $f(x) = 1$; 3° if $A_x = \{i \in K: x(i) = 0\}$ and $\mathrm{card}(A_x) = k$ then let $y(j) = x(\langle A_x, j\rangle)$ for $j = 0,1,2,3$ and let $f(x) = f_0(y)$.

To see that f thus defined is $(k+2)$-continuous notice that if case 1° or 2° applies then there exists a $(k+1)$-cylinder C with $x \in C$ and $f \restriction C$ is a constant. If case 3° applies and $f_o(y) = 0$ and $C_y \subseteq \{0,1\}^4$ is a 2-cylinder with $y \in C_y$ and $f_o \restriction C_y$ a constant, then the $(k+2)$-cylinder

$$C = \{z \in \{0,1\}^M: \ z(i) = x(i) \ \text{ for } \ i \in A_x \cup (\{A_x\} \times B(C_y))\}$$

contains x and $f \restriction C$ is a constant. While if $f_o(y) = 1$ and C_y is as above then the $(k+2)$-cylinder

$$C = \{z \in \{0,1\}^M: \ z(i) = x(i) \ \text{ for } \ i \in (K - A_x) \cup (\{A_x\} \times B(C_y))\}$$

also contains x and $f \restriction C$ is a constant. Thus f is $(k+2)$-continuous. It is also visible that f depends on all m variables. Q.E.D.

Lemma 25 can be improved as follows

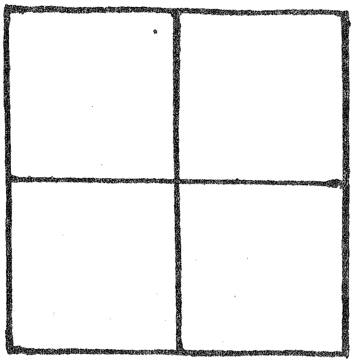
<u>Theorem 25A.</u> If $\mathrm{card}\,(A_1) = \ldots = \mathrm{card}\,(A_n) = a$, $\mathrm{card}\,(B_1) = \ldots = \mathrm{card}\,(B_n) = b$ and $A_i \cap B_j = \phi$ iff $i = j$ then $n \leq \binom{a+b}{a}$.

This theorem follows easily from a very strong Theorem 2 of B. Bollobås, On generalized graphs, Acta Math. Acad. Sci. Hung. 16(1965) 447-452.

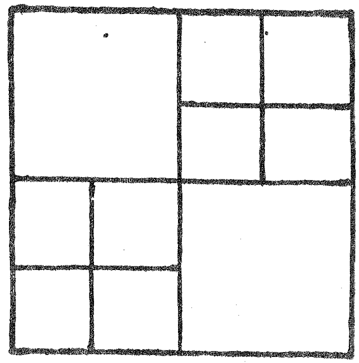
By Theorem 25A the upper estimate in Theorems 15 and 24 can be diminished to $(2k+1)\binom{2k}{k}$ (without changing our proof).

4. An example of regular k-continuous functions depending on $3 \cdot 2^{k-1} - 2$ variables.

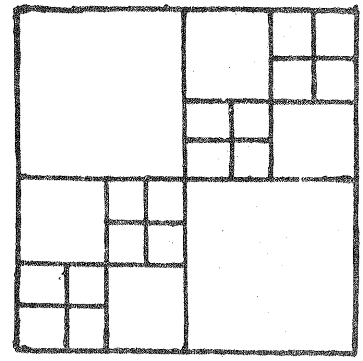
Consider the following partitions of a square into $3 \cdot 2^{k-1} - 2$ squares.



$k = 2$



$k = 3$



$k = 4$ ...

Let $\underline{A}_k$ be the collection of sets of squares of the k-th picture whose interiors can be intersected by one horizontal line and $\underline{B}_k$ be the collection of sets of squares of the k-th picture whose interiors can be intersected by one vertical line. Now if $M \cap A \cap B \neq \emptyset$ for all $A \in \underline{A}_k$ and $B \in \underline{B}_k$ then M consists of all the squares of the k-th picture. The regular k-continuous functions are constructed from $\underline{A}_k$ and $\underline{B}_k$ as in the first part of the proof of Theorem 23.

5. We give an example of a regular 3-continuous function $f: X \to \{0,1\}$, where $X \subseteq \{0,1\}^8$, such that f can not be extended to a 3-continuous function $f^*: \{0,1\}^8 \to \{0,1\}$. Let $+$ denote addition mod 8. We define two unions of 3-cylinders

$$X_0 = \{x \in \{0,1\}^8 : \exists i[(x_i, x_{i+1}, x_{i+2}) = (0,0,0)]\},$$

$$X_1 = \{x \in \{0,1\}^8 : \exists i[(x_i, x_{i+2}, x_{i+5}) = (1,1,1)]\}.$$

We put $X = X_0 \cup X_1$ and, since $X_0 \cap X_1 = \emptyset$, we can define f putting $f^{-1}(0) = X_0$, $f^{-1}(1) = X_1$. It is easy to check that every 3-cylinder containing the point $(0,0,1,1,0,0,1,1)$ intersects both X_0 and X_1. Hence no f^* can exist.

Problem. Under what condtions can a regular k-continuous function with domain included in $\{0,1\}^m$ be extended to a k-continuous function over $\{0,1\}^m$?

References

[1] A. Adám, Truth functions and the problem of their realization by two terminal graphs, Budapest 1968.

[2] M. Calczyńska-Karlowicz, Theorem on families of finite sets, Bull. Acad. Polon. Sci., ser. math astr. et phys. 12(1964), 87-89.

[3] A. Ehrenfeucht, Practical decidability (to appear).

[4] P. Erdös, Chao Ko and R. Rado, Intersection theorems for systems of finite sets, Quart. Journal of Math. 12(1961), 313-320.

[5] P. Erdös and J. H. Spencer, Probabilistic methods in combinatorics, (monograph to appear).

[6] J. M. Hammersley and D. C. Handscomb, Monte Carlo Methods, London 1964.

[7] G. O. H. Katona, A simple proof of the Erdös-Chao Ko-Rado theorem, (mimeographed).

[8] A. N. Kolmogorov and V. M. Tihomirov, ε-entropy and ε-capacity in function spaces, Uspehi Mat. Nauk 14(1959), no. 2(86), 3-86 or Amer. Math. Soc. Transl. (2) 17(1961), 277-364.

[9] R. McKenzie, J. Mycielski and D. Thompson, On Boolean functions and connected sets, Mathematical Systems Theory, 5(1971), 259-270.

[10] M. Minsky and S. Papert, Perceptrons, an introduction to computational geometry, MIT 1969.

[11] J. Mycielski, Review of [10], Bull. Amer. Math. Soc., 78(1972), 12-15.

[12] J. Mycielski, Monte Carlo interpolation of Boolean functions, Notices of the Amer. Math. Soc. 19(1972), A-594.

[13] L. Pukánszky and A. Renyi, On approximation of measurable functions, Publicationes Math. (Debrecen) 2(1951), 146-149.

[14] J. H. Spencer, Probabilistic methods in combinatorial theory, Thesis, Harvard 1970.

[15] T. Tsuda and H. Matsumoto, A note on linear extrapolation of multivariable functions by the Monte Carlo method, Journal of the ACM 13(1966), 143-150.

[16] T. Tsuda and Kozo Ichida, Non linear interpolation of multivariable functions by the Monte Carlo method, Journal of the ACM 17(1970), 420-425.