On some combinatorical problems.

In memoriam Tibor Szele.

By P. ERDŐS and A. RÉNYI in Budapest.

Introduction.

Let $C_k(n)$ denote the least number of such combinations of order k of n different elements, that any two elements are contained in at least one combination ($k, n = 2, 3, \ldots$). Such a system of combinations will be called a (k, n)-system. Clearly we have $\binom{k}{2} C_k(n) \geqq \binom{n}{2}$, as there are $\binom{k}{2}$ pairs in any combination of order k and each of the $\binom{n}{2}$ possible pairs must be contained in one of the $C_k(n)$ combinations. Thus we have

$$C_k(n) \geqq \frac{n(n-1)}{k(k-1)}. \tag{1}$$

If for some values of k and n there is equality in (1), we say that an optimal (k, n)-system exists. It is well-known, that if $k = P+1$ and $n = P^r + P^{r-1} + \cdots + P + 1$ where P is a power of a prime and $r \geqq 1$ an arbitrary integer, there exists an optimal (k, n)-system. This has been proved — according to our knowledge — first by TH. SKOLEM (see [1]). There exist also optimal (P, P^r) systems, if P is a power of a prime and $r \geqq 1$. These facts are nowadays utilized in constructing balanced incomplete block designs (see [2]). An optimal (k, n)-system is clearly a balanced incomplete block design of n varieties into $\frac{n(n-1)}{k(k-1)}$ blocks of k plots each, such that every variety occurs with every other variety exactly once in the same block. It seems that up to now interest was focused on optimal (k, n)-systems and the asymptotic behaviour of $C_k(n)$ for $n \to \infty$ has not been investigated. In § 1. of the present paper we prove that if $k = P$ is fixed, where P is a power of a prime, we have

$$\lim_{n \to \infty} \frac{C_k(n)}{n(n-1)} = \frac{1}{k(k-1)}, \tag{2}$$

i. e. there exists a sequence of asymptotically optimal (k, n)-systems for each fixed $k=P$ where P is a prime power. (2) is valid also for $k=P+1$ where P is a prime power. The proof is analogous to that given in the present paper for $k=P$, only Lemma 1 is used instead of Lemma 3. It can be proved by the same method that the limit $\lim\limits_{n\to\infty}\frac{C_k(n)}{n(n-1)}=\gamma_k$ exists for any $k\geqq 1$, but we do not know the value of γ_k for other values of k than mentioned above. However it can be proved that

$$\lim_{k\to\infty} k(k-1)\lim_{n\to\infty}\frac{C_k(n)}{n(n-1)}=1 \tag{3}$$

i. e. that $\lim\limits_{k\to\infty} k(k-1)\gamma_k=1$. These results together with a simple but ingenious method of proof, which has been formulated and used by T. SZELE in his thesis [3], are applied in § 2. to prove a conjecture which has been recently proposed by the second named author [4]. Let $D_k(n)$ denote the length of the shortest sequence formed from the digits $1, 2, \ldots, n$ in which any two digits i and j $(1\leqq i<j\leqq n)$ are at least once to be found in such position, that they are separated by at most k numbers. It has been proved, in [4], that

$$\frac{1}{2k-2}\leqq \varliminf_{n\to\infty}\frac{D_k(n)}{n^2}\leqq \varlimsup_{n\to\infty}\frac{D_k(n)}{n^2}\leqq\frac{1}{k}$$

and it has been conjectured, that

$$\lim_{n\to\infty}\frac{D_k(n)}{n^2}=\mu_k \tag{4}$$

exists for $k=2, 3, \ldots$; however the existence of (4) is proved only for $k=2$ and $k=3$, the proof for $k=3$ being due to N. G. DE BRUIJN; in these two cases the limit is $\frac{1}{2k-2}$. We prove that the limit (4) exists for all $k\geqq 2$; however our method does not lead to the determination of the value of μ_k.

§ 1. The asymptotic behaviour of $C_k(n)$.

Let us put

$$c_k(n)=\frac{k(k-1)C_k(n)}{n(n-1)}. \tag{5}$$

We shall prove

Theorem 1.

$$\lim_{n\to\infty} c_k(n)=1 \qquad \textit{for } k=P$$

where P is a power of a prime.

The proof requires a number of lemmas, some of which are well-known and are stated only for convenience.

Lemma 1. (Skolem): *If $k=P+1$, where P is a power of a prime and $n=P^r+P^{r-1}+\cdots+P+1$, where r is arbitrary, we have $c_k(n)=1$, i.e. there exists an optimal (k,n)-system; this system can be chosen in such a way that it contains a subsystem which is an optimal $\left(k, \frac{n-1}{k-1}\right)$-system.*

For the proof see [1] or [2] p. 109—111.

Lemma 2.

$$c_k(n) \leqq c_k(q)c_q(n) \qquad (k<q<n).$$

Proof of Lemma 2. Let us form from the numbers $1, 2, \ldots, n$ a (q,n)-system consisting of $C_q(n)$ combinations of order q. From each such combination let us form a (k,q) system consisting of $C_k(q)$ combinations of order k. Thus a (k,n) system is obtained. Thus we have $C_k(n) \leqq C_k(q)C_q(n)$. Multiplying this inequality by $\frac{k(k-1)}{n(n-1)}$ we obtain the assertion of Lemma 2.

Lemma 3. *If P is a power of a prime, and $r \geqq 1$, we have $c_P(P^r)=1$, i.e. there exists an optimal (P, P^r)-system.*

Lemma 3 can be deduced from Lemma 1 (see [2] p. 112). Let us consider an optimal (k,n)-system, for $k=P+1$, $n=P^r+P^{r-1}+\cdots+P+1$, which exists according to Lemma 1, and which contains as a subsystem an optimal $\left(k, \frac{n-1}{k-1}\right)$-system. Let us omit from the given (k,n)-system the mentioned $\left(k, \frac{n-1}{k-1}\right)$-system; it is easy to see that any one of the remaining combinations contains exactly one element of the omitted subsystem. Omitting the mentioned element from each of these combinations, we obtain an optimal $\left(k-1, n-\frac{n-1}{k-1}\right)$-system, i.e. an optimal (P,P^r)-system, as $k-1=P$ and $n-\frac{n-1}{k-1}=P^r$.

Optimal (P,P^r)-systems can also be constructed directly, without using optimal $(P+1, P^r+P^{r-1}+\cdots+P+1)$-systems. We give here only the construction of optimal (p,p^2)-systems if p is prime. Let us represent the p^2 elements by all pairs (i,j) of residue classes $\bmod p$ $(i,j=0,1,\ldots,p-1)$. Let us consider to any two residues h and k $\bmod p$ the combinations

$$C_{hk}:(0,h),(1,h+k),\ldots,(r,h+rk),\ldots,(p-1,h+(p-1)k).$$

Thus we obtain p^2 combinations; let us consider besides these for any residue h $\bmod p$ the combination

$$C_h:(h,0),(h,1),\ldots,(h,p-1).$$

Thus we obtain altogether p^2+p combinations of order p, which together form an optimal (p, p^2)-system. As a matter of fact if (x, y) and (u, v) are two pairs of residue classes $\bmod p$ then if $x=u$ these two pairs of residue classes are both contained in C_x; if $x \neq u$ the two pairs of residue classes are both contained in the combinations C_{hk}, where h and k are determined by the congruences $h+xk \equiv y \pmod{p}$, $h+uk \equiv v$, $(\bmod p)$, which have a solution owing to $x \not\equiv u \pmod{p}$ namely $k \equiv \dfrac{y-v}{x-u} \pmod{p}$ and $h \equiv \dfrac{xv-yu}{x-u} \pmod{p}$.

In what follows P shall denote a *fixed* number which is the power of a prime; $a_1, a_2, \ldots$ will denote positive constants, depending eventually on P or ε but not on n.

Lemma 4. *If* $0<\varepsilon<\dfrac{1}{4}$ *we have*

$$c_k(n) \leqq c_k(N)(1+8\varepsilon)$$

if

$$N(1-\varepsilon) \leqq n \leqq N \qquad (N \geqq 2).$$

PROOF. We have $C_k(n) \leqq C_k(N)$ for $n \leqq N$ and thus

$$c_k(n) \leqq \frac{N(N-1)}{n(n-1)} c_k(N) \leqq \frac{1}{(1-2\varepsilon)^2} c_k(N)$$

and if $0<\varepsilon<\dfrac{1}{4}$ we have $\dfrac{1}{(1-2\varepsilon)^2} \leqq 1+8\varepsilon$ which proves Lemma 4.

Lemma 5. (INGHAM): *If p_n denotes the sequence of primes,* $(n=1, 2, \ldots)$, *we have* $p_{n+1}-p_n<p_n^{11/16}$ *for* $p_n>a_1$, (See [5]).

Lemma 6. *If p_k denotes the k-th prime number and* $A_2>A_1 \geqq a_1$ *we have*

$$\operatorname*{Max}_{A_1 \leqq p_k \leqq A_2} \frac{p_{k+1}}{p_k} \leqq 1+\frac{1}{A_1^{5/16}}.$$

Lemma 6. follows easily from Lemma 5.

Lemma 7. *There exists to any ε with* $0<\varepsilon<\dfrac{1}{4}$ *arbitrary large numbers B for which*

$$c_P(n) \leqq 1+32\varepsilon \quad \textit{for} \quad B \leqq n \leqq 2B^2.$$

PROOF. Let $0<\varepsilon<\dfrac{1}{4}$ and an integer a_3 be given. Let us choose an integer r such that $A=P^r \geqq a_3$. By Lemma 3 $c_P(A)=1$ and thus by Lemma 4 $c_P(n) \leqq 1+8\varepsilon$ for $A(1-\varepsilon) \leqq n \leqq A$. If a_3 is sufficiently large, there is at least one prime in the interval $(A(1-\varepsilon), A)$. Let $q_1<q_2<\cdots<q_s$ denote the primes in the interval $(A(1-\varepsilon), A)$. It follows from Lemmas 2 and 3 that if m is any fixed integer, $m \geqq 2$, we have $c_P(q_i^m) \leqq 1+8\varepsilon$ as $c_P(q_i) \leqq 1+8\varepsilon$

$(i=1,2,\dots,s)$; thus by Lemma 4

$$c_P(n) \leqq (1+8\varepsilon)^2 \leqq 1+32\varepsilon \quad \text{for} \quad q_i^m(1-\varepsilon) \leqq n < q_i^m .$$

Now, if we choose m in such a way that the intervals $(q_i^m(1-\varepsilon), q_i^m)$ are not disjoint, i.e. $q_{i+1}^m(1-\varepsilon) < q_i^m$, it follows, that $c_P(n) \leqq 1+32\varepsilon$ for $q_1^m(1-\varepsilon) \leqq$ $\leqq n \leqq q_s^m$. Now $q_{i+1}^m(1-\varepsilon) < q_i^m$ $(i=1,\dots,s)$ is satisfied if $\frac{q_{i+1}}{q_i} < \left(\frac{1}{1-\varepsilon}\right)^{\frac{1}{m}}$ $(1 \leqq i \leqq s)$. As by Lemma 6. $\frac{q_{i+1}}{q_i} < 1+\frac{1}{(A(1-\varepsilon))^{5/16}}$ if a_3 is sufficiently large, this is true if $1+\frac{1}{(A(1-\varepsilon))^{5/16}} < \left(\frac{1}{1-\varepsilon}\right)^{1/m}$ and thus if a_3 is sufficiently large, this is true if $m < A^{1/4}$. Thus if $a_3 = a_3(\varepsilon)$ is sufficiently large, we have $c_k(n) \leqq 1+32\varepsilon$ for $(A(1-\varepsilon)+A^{11/16})^m \leqq n \leqq (A-A^{11/16})^m$. Now the intervals $[(A(1-\varepsilon)+A^{11/16})^m, (A-A^{11/16})^m]$ are not disjoint, if $m > (\log \mathrm{A})^2$, if a_3 is sufficiently large. Thus it follows that

$$c_P(n) \leqq 1+32\varepsilon \quad \text{for} \quad (A(1-\varepsilon)+A^{11/16})^{(\log A)^2} \leqq n \leqq (A-A^{11/16})^{A^{1/4}}$$

and thus a fortiori for $e^{2(\log A)^3} \leqq n \leqq e^{A^{1/4}}$. As the interval $(e^{2(\log A)^3}, e^{A^{1/4}})$ contains an interval $(B, 2B^2)$, provided that a_3 is sufficiently large, Lemma 7 is proved.

Lemma 8. *If $c_P(n) \leqq \alpha$ for $B \leqq n \leqq 2B^2$ we have $c_P(n) \leqq \alpha\left(1+\frac{24}{B^{5/16}}\right)$ for $B^2 \leqq n < 2B^4$ if $B \geqq a_4$.*

PROOF. Let $\pi_1, \pi_2, \dots, \pi_t$ denote the primes in the interval $(B, 2B^2)$; then by Lemma 2 and 3 $c_P(\pi_i^2) \leqq c_P(\pi_i) \leqq \alpha$. By Lemma 6 we have

$$\frac{\pi_{i+1}^2}{\pi_i^2} \leqq \left(1+\frac{1}{B^{5/16}}\right)^2 \leqq 1+\frac{3}{B^{5/16}} \qquad \text{if } B \geqq a_1$$

and thus we have

$$\pi_{i+1}^2\left(1-\frac{3}{B^{5/16}}\right) < \pi_i^2 .$$

It follows by Lemma 4. that $c_P(n) \leqq \alpha\left(1+\frac{24}{B^{5/16}}\right)$ for

$$\pi_i^2\left(1-\frac{3}{B^{5/16}}\right) \leqq n \leqq \pi_i^2 .$$

As

$$\pi_1^2\left(1-\frac{3}{B^{5/16}}\right) \leqq (B+B^{5/16})^2\left(1-\frac{3}{B^{5/16}}\right) < B^2$$

and

$$\pi_t^2 \geqq (2B^2-2B^{11/8})^2 \geqq 2B^4$$

if a_4 is sufficiently large, Lemma 8 is proved.

Now we are in position to prove Theorem 1. It follows from Lemma 7 that for any ε $\left(0<\varepsilon\leqq\frac{1}{4}\right)$ we can find a number B, which can be chosen greater than an arbitrary given number, such that $c_P(n)\leqq 1+32\varepsilon$ for $B\leqq n\leqq 2B^2$. It follows by Lemma 8 that

$$c_P(n)\leqq(1+32\varepsilon)\prod_{l=0}^{\infty}\left(1+\frac{24}{B^{5.2^{l-4}}}\right)\qquad\text{for any } n\geqq B.$$

As

$$\prod_{l=0}^{\infty}\left(1+\frac{24}{B^{5.2^{l-4}}}\right)\leqq 1+\frac{a_5}{B^{5/16}}$$

it follows that

$$\overline{\lim_{n\to\infty}}\; c_P(n)\leqq(1+32\varepsilon)\left(1+\frac{a_5}{B^{5/16}}\right). \tag{6}$$

As $\varepsilon>0$ can be chosen arbitrarily small and B arbitrarily large, (1) and (6) implies

$$\lim_{n\to\infty} c_P(n)=1.$$

Thus Theorem 1 is proved.

Now let k denote an arbitrary number and P the greatest prime power $\leqq k$. Clearly we have $C_k(n)\leqq C_P(n)$ as any (P,n)-system can be transformed into a (k,n) system, by adding arbitrary $k-P$ elements to each combination of the given (P,n)-system. Thus it follows from Lemma 3 that if k is arbitrary

$$\overline{\lim_{n\to\infty}}\; c_k(n)\leqq\frac{k(k-1)}{P(P-1)}$$

where P is the greatest prime power $\leqq k$. As $P>k-k^{11/16}$ for $k\geqq a_1$ it follows that

$$\overline{\lim_{n\to\infty}}\; c_k(n)\leqq 1+\frac{a_6}{k^{5/16}}$$

and thus

$$\lim_{k\to\infty}\left(\overline{\lim_{n\to\infty}}\, c_k(n)\right)=1. \tag{7}$$

It is not difficult to prove by the same method as applied in proving Theorem 1 that $\lim\limits_{n\to\infty} c_k(n)$ exists for every k, and thus $\overline{\lim}$ can be replaced by lim in (7).

§ 2. Application of a lemma of T. Szele.

In his paper [3] Szele has used the following simple but often very useful

Lemma 9. *If a_n is a sequence of real numbers, which is „almost monotonically decreasing", i. e. if $a_n\leqq a_m(1+\varepsilon)$ for any $\varepsilon>0$ and any $m\geqq m_0(\varepsilon)$, if $n\geqq n_0(\varepsilon,m)$, further a_n is bounded from below, then $\lim\limits_{n\to\infty} a_n=\alpha$ exist.*

Lemma 9 may be proved as follows. It follows from our supposition that for any $\varepsilon > 0$ and $m \geqq m_0(\varepsilon)$ we have

$$\overline{\lim_{n \to \infty}}\, a_n \leqq a_m(1+\varepsilon) \tag{8}$$

and thus

$$\overline{\lim_{n \to \infty}}\, a_n \leqq (1+\varepsilon) \underline{\lim_{m \to \infty}}\, a_m\,; \tag{9}$$

as $\varepsilon > 0$ is arbitrary, (9) implies that $\lim\limits_{n \to \infty} a_n$ exists.

Now let $D_k(n)$ denote the length of the shortest sequence, consisting of the digits $1, 2, \ldots, n$, which has the property that any two digits i and j $(1 \leqq i < j \leqq n)$ occur somewhere in the sequence in such a position that they are separated by not more than k elements of the sequence. We may restate the definition of $D_k(n)$ in the language of the theory of graphs. $D_k(n)$ is the length of the shortest directed path in the complete graph of n points, which has the property that from any point of the graph we may reach any other point in not more than $k+1$ steps, by going along the path always according to the given direction or always in the opposite direction. Then we have clearly

$$D_k(n) \leqq C_m(n) D_k(m) \qquad (k < m < n) \tag{10}$$

and thus

$$\frac{D_k(n)}{n^2} \leqq \frac{D_k(m)}{m^2}\, c_m(n) \frac{m}{m-1}\,. \tag{11}$$

Thus

$$\frac{D_k(n)}{n^2} \leqq \frac{D_k(m)}{m^2} \cdot \frac{m}{m-1}\left(1 + \frac{2a_6}{m^{5/16}}\right) \qquad \text{if } m \geqq a_1 \tag{12}$$

and $n \geqq n_0(\varepsilon, m)$, i. e.

$$\frac{D_k(n)}{n^2} \leqq (1+\varepsilon) \frac{D_k(m)}{m^2} \tag{13}$$

if $m \geqq m_0(\varepsilon)$ and $n \geqq n_0(\varepsilon, m)$. Applying the Lemma of Szele this implies that $\lim\limits_{n \to \infty} \dfrac{D_k(n)}{n^2}$ exists.

It should be mentioned that the authors of the present paper have applied the lemma of Szele with success to other combinatorical and number-theoretical questions too.

Bibliography.

[1] E. NETTO, Lehrbuch der Combinatorik, 2. Aufl. Noten von TH. SKOLEM p. 328—329.

[2] H. B. MANN, Analysis and design of experiments, *New-York*, Dover, 1949.

[3] T. SZELE, Kombinatorikai vizsgálatok az irányított teljes gráffal kapcsolatban, *Math. Phys. Lapok* **50** (1943), 223—256.

[4] A. RÉNYI, Néhány kombinatorikai problémáról, melyek a lucernanemesítéssel kapcsolatosak, *Mat. Lapok*, **6** (1955), 151—163.

[5] A. INGHAM, On the difference between consecutive primes, *Quart. J. Math. Oxford* **8** (1937), 255—266.

(Received October 20, 1955.)