Probabilistic existence theorems in group testing

Anatoly Zhigljavsky,
School of Mathematics, Cardiff University,
Senghennydd Road, Cardiff CF24 4YH, UK
e-mail ZhigljavskyAA@cardiff.ac.uk

Abstract

For a wide range of combinatorial group testing problems including additive, binary and multiaccess channel models, a probabilistic method is developed to derive upper bounds for the length of optimal nonsequential designs. A general result is proven allowing in many particular cases to compute the asymptotic bounds. The existence theorems are also derived for the situation when several errors in the test results can occur (searching with lies) and for the group testing problem in the binomial sample.

Key words: Probabilistic method, Existence theorems, Group testing, Factor screening, Optimum experimental design, Significant factors, Multiaccess channel, Searching with lies.

AMS classification: 62C10, 90Cxx, 28Dxx.

1 Introduction

Group testing, known also as factor screening and search for defective (significant) factors, is a vast area with many papers developing both theoretical and applied aspects.

Group testing algorithms aim at finding defective factors with a relatively small number of tests. To give a rough idea about the activity in the area, assume that the observations are error–free, the total number of factors, denoted by n, is large and t, the number of defective factors, is small, relative to n. Let us also fix the function defining the output of the tests. It is then typically possible to construct group testing designs, or at least to prove existence of such designs, that provide exact determination of all the defective factors in

$$N = N(t,n) \sim \; C_t \log n + O(1) \quad (n \to \infty) \tag{1}$$

tests, for some constant C_t. For large n this number of tests is much smaller than n. The optimal design can be defined as a design with the minimal design length in a certain class of designs. Asymptotic problems deal with minimization of the constant C_t and, possibly, the constant in $O(1)$ in (1).

Three different kinds of problem can then be distinguished:

(a) Construction of designs with minimal or at least reasonably small length $N(t,n)$;

(b) Derivation of lower bounds for the length of optimal algorithms;

(c) Derivation of upper bounds for the length of optimal algorithms (existence theorems).

In the present paper we deal exclusively with the problem (c) above, that is, with establishing existence theorems. These theorems guarantee that there exist designs with a length smaller than the derived upper bound; they also give an idea about the design set where designs with good or even optimal design length can be expected. However, they do not provide construction schemes for such designs. Moreover, in all nontrivial cases, for large n and $t > 2$, the author has failed to find design schemes providing the designs with lengths within the bounds derived below. This can perhaps be considered as an indication of the fact that the optimum group testing design problem is rather complex.

A general statement of the group testing problem assumes that n factors (elements, items, variables, etc.) $x_1, \ldots, x_n$ are given and some of them are *defective* (significant, important, etc.). The problem is to determine which factors are defective by testing a certain number of factor groups from a design set $\mathcal{X}$, which contains some subsets of the set $\mathbf{X} = \{x_1, \ldots, x_n\}$.

The problems differ in the following aspects:
(i) assumptions concerning the occurrence of defective factors;
(ii) assumptions on admissible designs;
(iii) forms of the test function which provides observation results;
(iv) definitions of the problem solution.

The group testing problems considered below are specified by the following properties.

(i) Let t be some integer, $0 < t \leq n$. We assume that either there are exactly t defective factors or the number of defective factors is smaller than or equal to t (see also Section 6.2, where we modify some results to the problem of finding defectives in the binomial sample, when each factor has a prior probability to be defective).

(ii) We only consider nonsequential designs. (Obviously, the upper bounds for nonsequential designs can also serve as upper bounds in the sequential case.) Also, we only consider the design sets $\mathcal{X}$, which contain the factor groups consisting of either exactly s or $\leq s$ factors for some $s \leq n$. We show that the former case is the principal one. In the asymptotic considerations, when $n \to \infty$, we assume that $s = s(n) = \lambda n + o(n)$, where λ is some number, $0 < \lambda < 1$.

(iii) Let T be the unknown collection of defective factors and $X \in \mathcal{X}$ be a test factor group. In the models considered below the test function f is

$$f(X,T)=f_K(X,T)=\min\{K, |X \cap T|\}. \tag{2}$$

where $|\cdot|$ stands for the number of elements in a discrete set and K is some integer. This is a rather general form of the output and corresponds to the K-channel model, see e.g. Du and Hwang (2000), Section 10.4 and Tsybakov et al (1983).

The three well-known special cases are the following ones. In the *binary* (alternatively, disjunctive) model $K = 1$. (In this model by inspecting a group $X \subset \mathbf{X}$ we receive 1 if there is at least one defective factor in X and 0 otherwise.) In the *additive* model $K = n$ (alternatively, any number between n and ∞ can be chosen as K) and therefore $f(X,T) = |X \cap T|$. (In the additive model, after inspecting a group X we receive the number of defectives in X.) In the *multiaccess channel* model $K = 2$ and therefore $f(X,T) = \min\{2, |X \cap T|\}$. The binary model is by far the most popular in the group testing theory.

We mostly consider error–free models which assume that there are no observation errors. However, in Section 6.1 we demonstrate that many results can be generalized to the case when a few mistakes in the test results are possible; this corresponds to *searching with lies.*

(iv) Concerning the requirements for the solution, we consider the problem of exact determination of the collection of unknown defectives, *strong separation*, and the problem of finding this collection in most cases, *weak separation.*

Group testing is a well established area attracting attention of specialists in optimum design, combinatorics, information theory and discrete search. The paper Dorfman (1943), devoted to sequential procedures of blood testing for detection of syphilitic men, is usually considered as the first work on the theory of group testing. A state-of-art in the field is well presented in the monograph Du and Hwang (2000). We do not consider an important problem of finding efficient designs. We only refer to Du and Hwang (2000), Ghosh and Avila (1985), Katona (1979), Katona and Srivastava (1983), Macula and Reuter (1998), Macula (1997b) and Patel (1987) as a sample of works dealing with design construction schemes in important specific cases including the case of the binary model with two and, more generally, t defectives.

The relationship between combinatorial group testing and discrete search is a key ingredient of the present methodology. We refer to Ahlswede (1987), O'Geran et al (1991) and Srivastava (1975), where different aspects of this relationship are thoroughly described. In this respect we mention that the well-known paper Srivastava (1975) considers also some schemes of design construction and demonstrates many properties of design matrices for combinatorial group testing problems in a general setup.

Existence theorems constitute an important part of the theory of discrete mathematics, see e.g. Ahlswede (1987), Alon, Spencer and Erdős (1992). In the field of group testing, the corresponding activity has been originated in the seminal work Rényi (1965) and has been successfully continued by many authors, numerous examples are given in Du and Hwang (2000). We especially mention a great input in the field of M.B. Malutov, A.G. Dyachkov, V.V. Rykov and other representatives of the Moscow probability school, see e.g. Dyachkov and Rykov (1983), Tsybakov et al (1983). The problem of using the probabilistic methods for constructing upper bounds of the length of optimal strongly separating designs in the binary group testing model has attracted a reasonable attention in literature, see Du and Hwang (2000) for a survey of results. For a fixed number of defectives t and $n \to \infty$, the best known upper bound for the length of strongly separating designs in the binary group testing problem has been derived in Dyachkov, Rykov and Rashad (1989), see also Theorem 7.2.15 in Du and Hwang (2000): $N \leq \frac{t}{A_t}(1+o(1))\log_2 n$, where

$$A_t = \max_{0\leq q\leq 1}\max_{0\leq Q\leq 1}\left\{-(1-Q)\log_2(1-q^t)+t\left[Q\log_2\frac{q}{Q}+(1-Q)\log_2\frac{1-q}{1-Q}\right]\right\}$$

and $A_t = \frac{2}{t\log_2 e}(1+o(1))$ when $t \to \infty$. The derivation of this bound heavily uses specific features of the binary group testing problem and in this sense could not be considered as a general methodology. Asymptotically, when both n and t are large, this is a marginally better bound than the asymptotic bound (68) of Section 5.2, that is,

$$N \leq N_*(n,t) \sim \frac{e}{2}t^2\log n\,,\quad n\to\infty,\ t=t(n)\to\infty,\ t(n)/n\to 0.$$

Note that the latter double asymptotic bound has been also derived in Dyachkov and Rykov (1983) with the help of another version of the probabilistic method which is defined in Section 2.4.

The case of strongly separating designs for the additive model, the simplest model among considered in this paper, has been thoroughly studied in Zhigljavsky and Zabalkanskaya (1996). To some extent, the present work can be considered as a development of the technique of that paper.

In the asymptotic considerations we assume that the number of defective factors is small relative to the total number of factors n. This makes a big difference between the asymptotic results of the present paper and the results obtained in a series of papers, see e.g. Erdős and Rényi (1963), Lindstrem (1975), where the nonsequential group testing problem for the additive model is considered with no constraints on both the test groups and the number of defective factors. The above papers yield the result $N_{\min} \sim 2n/\log_2 n$, $n \to \infty$, for the minimal length of the nonsequential strategies that guarantee detection of all defectives.

The group testing problem in the binomial sample is very popular in the group testing literature. It has been introduced and thoroughly investigated in Sobel and Groll (1959). For modern generalizations of this problem see e.g. Bruno et al (1988).

In recent years a number of papers have been published on the problem of constructing optimal algorithms for finding one, two or three defectives in search with lies, see e.g. Hill and Karim (1992), Hill (1995), Macula (1997a), DeBonis et al (1997).

The author is not aware of the existence theorems for the group testing designs in the binomial sample problem and in the presence of lies.

The paper is organized as follows.

In Section 2 we consider the group testing problems from a general point of view of discrete search and provide two general existence theorems. These theorems can be applied as soon as suitable terms, called *Rényi coefficients*, can be computed for all the pairs (T, T') of distinct possible groups of defective factors.

In Section 3 we demonstrate that in many interesting group testing problems the set of these pairs can be partitioned into certain subsets, $\mathcal{T}_i$, where the Rényi coefficients are equal and have closed-form representations through the binomial coefficients. The number of elements in the sets $\mathcal{T}_i$ can also be expressed through the binomial coefficients. We thus show that the upper bounds can typically be written in the form $N(n) = \min \left\{ k = 1, 2, \ldots \,:\, \sum_i q_{i,n} r_{i,n}^k < 1 \right\}$ for suitable numbers $q_{i,n}$ and $r_{i,n}$.

In Section 4 we use general results of Sections 2 and 3 to derive specific forms of the upper bounds for the length of optimal designs in specific group testing models.

In Section 5 we prove a general result, which in many particular cases allows to obtain the asymptotic laws for $N(n), n \to \infty$, and apply this result to derive asymptotic upper bounds for optimal design length in specific group testing problems.

In Section 6 we demonstrate how the main existence theorems can be modified for the group testing problems in the binomial sample and generalized for the case when several lies in the test results can occur.

The Appendix contains several tables providing numerical illustrations.

2 General Existence Theorems

2.1 Discrete search problems

We consider the group testing problems from the general point of view of discrete search. Let us first give several definitions.

Following O'Geran et al (1991) a discrete *search problem* can often be determined as a quadruple $\{\mathcal{T}, \mathcal{X}, f, \mathcal{Y}\}$ where $\mathcal{T} = \{T\}$ is a target field, that is an ordered collection of all possible *target elements* T, $\mathcal{X} = \{X\}$ is a test field, that is a collection of all possible *test elements*, and $f : \mathcal{X} \times \mathcal{T} \to \mathcal{Y}$ is a *test function* mapping $\mathcal{X} \times \mathcal{T}$ to some space $\mathcal{Y}$. A value $f(X, T)$ for fixed $X \in \mathcal{X}$ and $T \in \mathcal{T}$ is regarded as test or experimental result at a test element X when the unknown target is T. We only consider *solvable* search problems where each $T \in \mathcal{T}$ can be separated (found) by means of the test results at all $X \in \mathcal{X}$.

A *nonsequential design* $\mathcal{D}_N$ of length N is a collection $\mathcal{D}_N = \{X_1, \ldots, X_N\}$ of test elements, which are chosen before the tests start. We shall not consider sequential (adaptive) designs and will omit the word 'nonsequential' while referring to a nonsequential design.

For a pair of targets $T, T' \in \mathcal{T}$, we say that $X \in \mathcal{X}$ separates T and T' if $f(X,T) \neq f(X,T')$. We say that a design $\mathcal{D}_N = \{X_1, \ldots, X_N\}$ separates T in $\mathcal{T}$ if for any $T' \in \mathcal{T}$, such that $T' \neq T$, there exists a test element $X \in \mathcal{D}_N$ separating the pair (T, T'). A design $\mathcal{D}_N$ is *strongly separating* if it separates all T in $\mathcal{T}$. A design $\mathcal{D}_N$ is γ*-separating* if

$$\frac{|\{T \in \mathcal{T}: \text{ design } \mathcal{D}_N \text{ separates } T \text{ in } \mathcal{T}\}|}{|\mathcal{T}|} \geq 1 - \gamma\,, \tag{3}$$

where γ is a fixed number, $0 \leq \gamma \leq 1$.

0-Separating designs are strongly separating; γ-separating designs with $\gamma > 0$ are called *weakly separating.*

Strongly separating designs are used when it is required to find the unknown target T whatever the target $T \in \mathcal{T}$ is. Alternatively, weakly separating designs can be used when it is sufficient to separate the target in the majority of cases. As we shall see later, one can typically guarantee the existence of weakly separating designs with much smaller length than for the strongly separating designs.

2.2 Existence theorems

The purpose of this section is formulation of two key theorems (Theorems 2.1 and 2.2). For the sake of completeness these results are accompanied with short proofs. Note that certain versions of these results are known in literature, see Rényi (1965), O'Geran et al (1991), Zhigljavsky and Zabalkanskaya (1996), Dyachkov and Rykov (1983).

For the strongly separating designs we have the following *existence principle*, see e.g. O'Geran et al (1991).

Let $\mathcal{R}$ be a probability distribution on $\mathcal{X}$, $\mathcal{A}_{ij}(N)$ be the event that the targets T_i and T_j are not separated in N independent $\mathcal{R}$-distributed random tests and let

$$p_{ij} = \Pr\{\mathcal{A}_{ij}(1)\} = \Pr\{f(X,T_i) = f(X,T_j)\}$$

be the probability that the targets T_i and T_j are not separated by one test at random $X \in \mathcal{X}$, which is distributed according to $\mathcal{R}$.

Then $\Pr\{\mathcal{A}_{ij}(N)\} = p_{ij}^N$ and therefore we have for the probability that at least one pair T_i and T_j is not separated after performing N random tests:

$$\Pr\left\{\bigcup_{i<j} \mathcal{A}_{ij}(N)\right\} \leq \sum_{i<j} \Pr\{\mathcal{A}_{ij}(N)\} = \sum_{i<j} p_{ij}^N\,. \tag{4}$$

Hence, the probability that all pairs of distinct targets are separated after performing N random tests is

$$1 - \Pr\left\{\bigcup_{i<j} \mathcal{A}_{ij}(N)\right\} \geq 1 - \sum_{i<j} p_{ij}^N \tag{5}$$

Assume that N is large enough to provide the positivity of the right-hand side in (5), then the discreteness of $\mathcal{T}$ immediately implies the existence of a deterministic design of length N separating all the targets $T_i \in \mathcal{T}$. We thus have the following upper bound for the length of optimal designs.

Theorem 2.1 (Existence theorem for strongly separating designs.) *Let $\{\mathcal{T}, \mathcal{X}, f, \mathcal{Y}\}$ be a solvable discrete search problem and $\mathcal{R}$ be any randomisation scheme. Then there exists a strongly separating design with sample size*

$$N \; \leq \; N^* = \min \left\{ k = 1, 2, \ldots \text{ such that } \sum_{i=1}^{|\mathcal{T}|} \sum_{j=1}^{i-1} p_{ij}^k \; < 1 \right\}. \tag{6}$$

The inequality (4) and therefore the upper bound (6) seem to be crude. To a certain extent, this is true in simple problems, when either N is small or all p_{ij} are equal or approximately equal. For example, in group testing problems with one defective factor the upper bound (6) gives the asymptotic upper bound $2\log_2 n$ for the number of tests but the optimal value is $\lceil \log_2 n \rceil$, see Ahlswede(1987), Du and Hwang (2000), Rényi (1965), Zhigljavsky and Zabalkanskaya (1996) for details. However, the bound (6) seems to be reasonably sharp in many difficult problems including the problems discussed below. That can be explained by the fact that the values of p_{ij} for most of the pairs (T_i, T_j) are relatively small and the value of the sum in the right–hand side of (4) is basically determined by a very small number of probabilities p_{ij}. This is the situation occurring in the group testing problems discussed below.

Note that the events $\mathcal{A}_{ij}(N)$ are typically dependent (which is certainly true for all group testing problems with $n \geq 4$) and therefore we actually have the strict inequality in (4) and therefore non-strict inequality in the upper bound (6). Another minor refinement to the bound (6) can be obtained if we choose X_i using the random sampling without replacement rather than the random sampling with replacement, see Zhigljavsky and Zabalkanskaya (1996) for details. These refinements can be of importance only when both $|\mathcal{T}|$ and $|\mathcal{X}|$ are small.

Construction of upper bounds is similar in the case of weakly separating designs. To derive these bounds, only a minor modification of the arguments leading to Theorem 2.1 is required. Indeed, by analogy with (4), for a fixed $T_i \in \mathcal{T}$ the probability that T_i is not separated from at least one $T_j \in \mathcal{T}$ $(T_j \neq T_i)$ after N random tests is less than or equal to $\sum_{j \neq i} (p_{ij})^N$ and we thus have $1 - \sum_{j \neq i} (p_{ij})^N$ as a lower bound for the probability that T_i is separated from all other $T_j \in \mathcal{T}$. Summation over i and the use of (3) yields the upper bound

$$N \; \leq \; N_\gamma = \min \left\{ k = 1, 2, \ldots \text{ such that } \frac{2}{|\mathcal{T}|\gamma} \sum_{i=1}^{|\mathcal{T}|} \sum_{j=1}^{i-1} p_{ij}^k \; < 1 \right\} \tag{7}$$

for the length of the optimal γ-separating design. We thus have the following theorem.

Theorem 2.2 (Existence theorem for weakly separating designs.) *Let* $\{\mathcal{T}, \mathcal{X}, f, \mathcal{Y}\}$ *be a solvable search problem,* $|\mathcal{T}| \geq 2$, $0 < \gamma < 1$. *Then there exists a non-sequential* γ*-separating design with the sample size* (7).

If we rewrite the upper bound (7) in the form

$$N_\gamma = \min\left\{ k = 1, 2, \ldots \text{ such that } \sum_{i=1}^{|\mathcal{T}|} \sum_{j=1}^{i-1} p_{ij}^k < \frac{|\mathcal{T}|\gamma}{2} \right\}$$

then the resemblance with (6) becomes obvious: for given $\mathcal{X}$ and $\mathcal{T}$, the difference between the two formulae is that 1 in the right–hand side of the inequality inside the curly brackets in the formula for N^* should be replaced with $\frac{|\mathcal{T}|\gamma}{2}$ in the formula for N_γ. Therefore, if we would find a closed-form expression for the upper bound in the case of strong separability, we automatically get a related expression for the upper bounds for the weakly separating designs. Note that this only concerns nonasymptotic expressions. Asymptotic considerations, when $|\mathcal{T}| \to \infty$, are usually different.

Note also that values of γ in (7) such that $\gamma < \frac{2}{|\mathcal{T}|}$ do not have much sense since γ-separation with $\gamma < \frac{2}{|\mathcal{T}|}$ implies strong separation.

2.3 Group testing as a discrete search problem

In the group testing problems both $\mathcal{X}$ and $\mathcal{T}$ are certain collections of factor groups. In the main body of the paper (until Section 6) we consider the problems when $\mathcal{T}$ is either $\mathcal{P}_n^t$ or $\mathcal{P}_n^{\leq t}$ and analogously $\mathcal{X}$ is either $\mathcal{P}_n^s$ or $\mathcal{P}_n^{\leq s}$. Here n is the total number of factors, $1 \leq t \leq n$, $1 \leq s \leq n$,

$$\mathcal{P}_n^k = \{\{x_{i_1}, \ldots, x_{i_k}\}, \quad 1 \leq i_1 < \ldots i_k \leq n\} \tag{8}$$

is the collection of all factor groups containing exactly k factors, and

$$\mathcal{P}_n^{\leq k} = \bigcup_{j=0}^{k} \mathcal{P}_n^j \tag{9}$$

is the collection of factor groups containing not more than k factors. In an important particular case, $\mathcal{P}_n^{\leq n} = 2^{\mathbf{X}}$ is the set of all possible 2^n subsets of $\mathbf{X} = \{x_1, \ldots, x_n\}$.

All the test functions f considered below belong to the general class (2).

2.4 Randomisation schemes

The randomisation scheme $\mathcal{R}' = \mathcal{R}'(\lambda)$ where $\mathcal{X} = 2^{\mathbf{X}} = \mathcal{P}_n^{\leq n}$ and every element of $\mathbf{X}$ is included into random X with a fixed probability λ (to be optimised at a later stage) is known to work quite well in some group testing problems, see Du and Hwang (2000), Dyachkov and Rykov (1983). We consider a different scheme, called $\mathcal{R}(s)$, where $\mathcal{X} = \mathcal{P}_n^s$ for some s and $\mathcal{R}$ is the uniform distribution on $\mathcal{X}$. (Sometimes it might also be worthwhile to consider $\mathcal{X}$ in the form $\mathcal{P}_n^{\leq s}$ or even $\cup_{s \in S} \mathcal{P}_n^s$, where S is a subset of $\{1, \ldots, n\}$; below these cases are covered as well.)

The advantages of the scheme $\mathcal{R}(s)$ over $\mathcal{R}'(\lambda)$ are as follows:

(i) numerical comparison shows that the optimal s in $\mathcal{R}(s)$ scheme always, for any n and t and any group testing problem among considered, allows to achieve better (or sometimes the same) upper bound than the optimal λ in the corresponding $\mathcal{R}'(\lambda)$ scheme;

(ii) the use of the scheme $\mathcal{R}(s)$ with subsequent optimisation over s gives more information about the structure of the design, where the upper bounds are achievable;

(iii) restriction on the number of elements in the test sets X in the form either $X \in \mathcal{P}_n^s$ or $X \in \mathcal{P}_n^{\leq s}$ for some s is obviously important in many practical situations (imagine blood testing, finding active compounds in drug development, detecting false computer chips, weighing problems, etc.)

We should however mention that the leading terms in the asymptotic expressions (when $n \to \infty$) for the upper bounds coincides for both schemes, $\mathcal{R}(s)$ and $\mathcal{R}'(\lambda)$ with $s = \lfloor n\lambda \rfloor$, in all the cases when the upper bounds based on the use of the scheme $\mathcal{R}'(\lambda)$ are known; this follows, roughly speaking, from the fact that the major part of the randomly selected test sets X_i in the scheme $\mathcal{R}'(\lambda)$ contain approximately λn factors.

We restrict our attention to the case when the distribution $\mathcal{R}$ is uniform on $\mathcal{X}$. In this case many probabilistic statements can be reduced to equivalent combinatorial ones. For example, the probabilities p_{ij} in (6) and (7) can be written as

$$p_{ij} = k_{ij}/|\mathcal{X}| \tag{10}$$

where $k_{ij} = k(T_i, T_j)$ is the number of $X \in \mathcal{X}$ such that $f(X, T_i) = f(X, T_j)$; that is,

$$k_{ij} = |\{X \in \mathcal{X} : \; f(X, T_i) = f(X, T_j)\}| \quad \text{for} \quad T_i, T_j \in \mathcal{T}\,. \tag{11}$$

The coefficients (11) in accordance with O'Geran et al (1991) will be called *Rényi coefficients*. Their derivation implies derivation of the upper bounds (6), (7) and constitutes one of the main objectives of the present work.

3 Computation of the Rényi coefficients

To write down the upper bounds (6) and (7), we need to find a way of computing the whole set of the Rényi coefficients (11), that is the set $\{k_{ij},\ (T_i, T_j) \in \mathcal{T} \times \mathcal{T}\}$. It happens that in many interesting group testing problems the set $\mathcal{T} \times \mathcal{T}$ can be partitioned into a few subsets, they are (13), where the Rényi coefficients are equal and have a closed-form representation through the binomial coefficients, see (28).

3.1 Branching of the target field

Let

$$\binom{n}{n_1\, n_2 \ldots n_k} = \frac{n!}{n_1! n_2! \ldots n_k!} \quad \text{for } n_r \geq 0,\ \sum_{r=1}^{k} n_r = n$$

be the multinomial coefficient and let us use the following convention

$$\binom{n}{n_1\; n_2 \ldots n_k} = 0 \quad \text{if} \quad \min\{n_1, \ldots, n_k\} < 0\,. \tag{12}$$

Let $0 \le p \le m \le l \le n$, $p < l$. Denote

$$\mathcal{T}(n,l,m,p) = \{(T,T') \in \mathcal{P}_n^{\le n} \times \mathcal{P}_n^{\le n} :\; |T| = m,\; |T'| = l,\; |T \cap T'| = p\}\,. \tag{13}$$

Note that the condition $p < l$ guarantees that $T \neq T'$ for all pairs $(T,T') \in \mathcal{T}(n,l,m,p)$.

The following theorem allows computing the number of elements in the sets (13).

Theorem 3.1 *The number of different non-ordered pairs in $\mathcal{T}(n,l,m,p)$ equals*

$$Q(n,l,m,p) = \begin{cases} \binom{n}{p\; m-p\; l-p\; n-l-m+p} & \text{if } m < l \\ \frac{1}{2}\binom{n}{p\; m-p\; m-p\; n-2m+p} & \text{if } m = l \end{cases} \tag{14}$$

The proof contains only easy counting arguments, details see in Zhigljavsky and Zabalkanskaya (1996).

3.2 Balanced test fields

Let $\mathcal{X} \subseteq \mathcal{P}_n^{\le n}$ be a test field and (T,T') be any pair in $\mathcal{P}_n^{\le n} \times \mathcal{P}_n^{\le n}$ such that $T \neq T'$. Define the sets

$$\mathcal{X}_{uvr}(T,T') = \{X \in \mathcal{X} :\; |X \cap (T \backslash T')| = u,\; |X \cap (T' \backslash T)| = v,\; |X \cap T \cap T'| = r\} \tag{15}$$

where u, v, r are some nonnegative integers.

Note that $(T,T') \in \mathcal{T}(n,l,m,p)$ for some l,m,p such that $0 \le p \le m \le l \le n$ and $p < l$. Observe also that the sets $\mathcal{X}_{uvr}(T,T')$ may be non-empty only if

$$0 \le u \le l-p,\; 0 \le v \le m-p,\; 0 \le r \le p\,. \tag{16}$$

Joining these restrictions on the parameters u, v, r with the restrictions on p, m and l in the definition of the sets $\mathcal{T}(n,l,m,p)$, we obtain the combined parameter restriction

$$0 \le p \le m \le l \le n,\; p < l,\; 0 \le u \le l-p,\; 0 \le v \le m-p,\; 0 \le r \le p\,. \tag{17}$$

Another consequence of (16) is the fact that for any pair (T,T'), such that $T \neq T'$, the test field $\mathcal{X}$ can be partitioned as follows

$$\mathcal{X} = \bigcup_{r=0}^{p} \bigcup_{u=0}^{l-p} \bigcup_{v=0}^{m-p} \mathcal{X}_{uvr}(T,T')\,. \tag{18}$$

Definition. *We shall call the test field $\mathcal{X}$ balanced if the number $|\mathcal{X}_{uvr}(T,T')|$ does not depend on the choice of the pair $(T,T') \in \mathcal{T}(n,l,m,p)$ for any set of integers u, v, r, p, m, l satisfying* (17).

For balanced test fields $\mathcal{X}$ the number $|\mathcal{X}_{uvr}(T,T')|$ for $(T,T') \in \mathcal{T}(n,l,m,p)$ will be denoted $R(n,l,m,p,u,v,r)$:

$$R(n,l,m,p,u,v,r) = |\mathcal{X}_{uvr}(T,T')| \quad \text{for } (T,T') \in \mathcal{T}(n,l,m,p)\,. \tag{19}$$

Note that the number $R(n,l,m,p,u,v,r)$ depends on $\mathcal{X}$.

Below we consider some examples of balanced test fields and derive explicit formulae for $R(n,l,m,p,u,v,r)$. Theorem 3.2 below demonstrating that the test field $\mathcal{X} = \mathcal{P}_n^s$ is balanced, is basic in this respect.

3.3 The case $\mathcal{X} = \mathcal{P}_n^s$

The case $\mathcal{X} = \mathcal{P}_n^s$ is the main example of the test fields we consider. It deserves special attention for the following reasons:

- the test field $\mathcal{X} = \mathcal{P}_n^s$ is balanced, see Theorem 3.2;
- this case is the most important from practical point of view;
- the formulae for the Rényi coefficients and the upper bounds can often be simplified using suitable combinatorial identities;
- as a rule, in the case $\mathcal{X} = \mathcal{P}_n^s$ with optimal s one can achieve better bounds than, say, in the case $\mathcal{X} = \mathcal{P}_n^{\le s'}$, with optimal s'.

Theorem 3.2 *Let* $\mathbf{X} = \{x_1, \ldots, x_n\}$ *be a collection of* $n \ge 2$ *factors and let the integers* u, v, r, p, m, l *satisfy* (17). *Then the test field* $\mathcal{X} = \mathcal{P}_n^s = \{X = \{x_{i_1}, \ldots, x_{i_s}\}\}$ *is balanced and*

$$R(n,l,m,p,u,v,r) = \binom{p}{r}\binom{l-p}{u}\binom{m-p}{v}\binom{n-l-m+p}{s-r-u-v} \tag{20}$$

where, in accordance with (12),

$$\binom{b}{a} = 0 \text{ for } a < 0 \quad \text{and } a > b. \tag{21}$$

Proof. Let $\mathcal{X} = \mathcal{P}_n^s$, (T,T') be any pair in $\mathcal{T}(n,l,m,p)$ and integers n, m, l, p, u, v, r be fixed and satisfy (17). Then $R(n,l,m,p,u,v,r)$, the number of the test groups $X \in \mathcal{X}_{uvr}$, can be computed as follows. To construct a factor group $X \in \mathcal{X}_{uvr}$ we make the following four sequential selection steps:

(1) select r factors from the set $T \cap T'$, there are $\binom{p}{r}$ possibilities to do this;

(2) select u factors ($0 \le u \le l-p$) from the set $T \backslash T'$, there are $\binom{l-p}{u}$ possibilities for doing this;

(3) select v ($0 \le v \le m-p$) factors from the set $T' \backslash T$, we can do this by $\binom{m-p}{v}$ choices;

(4) include $s-r-u-v$ factors into X from the set $\mathcal{X} \backslash (T \cup T')$ containing $n-l-m+p$ elements (obviously, we have $\binom{n-l-m+p}{s-r-u-v}$ possibilities to do this).

If s is either too small (i.e. $s < l + m - p$) or large ($s > n - l - m + p$) then the last step in the selection procedure is impossible and the corresponding value of R equals zero, which is in agreement with the convention (21).

To finish the proof we only mention that the above calculation of $R(n, l, m, p, u, v, r)$ does not depend on the choice of the pair $(T_i, T_j) \in \mathcal{T}(n, l, m, p)$. □

3.4 Auxiliary statements

In this section we provide several lemmas that will be frequently used to simplify expressions for the Rényi coefficients and the upper bounds in specific setups.

Lemma 3.1 *For any balanced test field $\mathcal{X}$ and all integers n, l, m, p such that $0 \le p \le m \le l \le n$, $p < l$, the following relationship holds:*

$$|\mathcal{X}| = \sum_{r=0}^{p} \sum_{u=0}^{l-p} \sum_{v=0}^{m-p} R(n, l, m, p, u, v, r) \,. \tag{22}$$

Proof. Since for any pair (T, T') in $\mathcal{P}_n^{\le n} \times \mathcal{P}_n^{\le n}$ such that $T \neq T'$, the sets in the collection $\{\mathcal{X}_{uvr}(T, T')\}_{u,v,r}$ are disjoint and provide a partition of $\mathcal{X}$, (22) follows from (18). □

The following simple statement not requiring a proof shows that a union of two disjoint balanced sets is also balanced. This and Theorem 3.2 imply, for instance, that the set $\mathcal{P}_n^{\le s}$ is balanced, as is the set $\mathcal{P}_n^{s}$. This also yields that the sets

$$\mathcal{X} = \mathcal{X}_S = \cup_{s \in S} \mathcal{P}_n^{s} \,, \tag{23}$$

where $S \subset \{0, 1, \ldots, n\}$, are balanced; it also provides the way of computing values $R(n, l, m, p, u, v, r)$ for these sets. (Note that in particular cases $S = \{s\}$ and $S = \{0, 1, \ldots, s\}$ we have $\mathcal{X}_S = \mathcal{P}_n^{s}$ and $\mathcal{X}_S = \mathcal{P}_n^{\le s}$, respectively.)

Lemma 3.2 *Let the target field $\mathcal{T}$ be fixed, $\mathcal{X}' \cap \mathcal{X}'' = \emptyset$, both $\mathcal{X}'$ and $\mathcal{X}''$ be balanced sets with families of R-values $\{R'(n, l, m, p, u, v, r)\}_{l,m,p,u,v,r}$ and $\{R''(n, l, m, p, u, v, r)\}_{l,m,p,u,v,r}$ respectively. Then the set $\mathcal{X} = \mathcal{X}' \cup \mathcal{X}''$ is also balanced and*

$$R(n, l, m, p, u, v, r) = R'(n, l, m, p, u, v, r) + R''(n, l, m, p, u, v, r) \tag{24}$$

for all admissible values of l, m, p, u, v, r.

As a corollary of Lemma 3.2 we obtain the following additivity property of the Rényi coefficients.

Corollary 3.1 *Let the target field $\mathcal{T}$ be fixed, $\mathcal{X} = \mathcal{X}' \cup \mathcal{X}''$ and $\mathcal{X}' \cap \mathcal{X}'' = \emptyset$. Let also $\{k'_{ij};\, T_i, T_j \in \mathcal{T}\}$, $\{k''_{ij};\, T_i, T_j \in \mathcal{T}\}$ be two sets of the Rényi coefficients (11) for the test fields $\mathcal{X}'$ and $\mathcal{X}''$, respectively. Then for any pair T_i, T_j of distinct elements in $\mathcal{T}$ we have*

$$k_{ij} = |\{X \in \mathcal{X} \text{ such that } |X \cap T_i| = |X \cap T_j|\}| = k'_{ij} + k''_{ij} \,. \tag{25}$$

The following classic combinatorial identity will help to simplify several expressions involving the binomial coefficients.

Lemma 3.3 (Vandermonde convolution formula) *For all positive integers a, b, c such that $c < a$ and $b < a$*

$$\binom{a}{b} = \sum_{k=0}^{c} \binom{c}{k} \binom{a-c}{b-k}, \tag{26}$$

where for certain values of k the convention (21) *is perhaps used.*

For a proof see, for instance, Riordan (1968), p.8.

For the ease of references let us also state two obvious relationships between the binomial coefficients:

$$\binom{a-1}{b} = \frac{a-b}{a}\binom{a}{b}, \quad \binom{a}{b-1} = \frac{b}{a-b+1}\binom{a}{b} \quad \text{for } 1 \le b < a. \tag{27}$$

3.5 Computation of the Rényi coefficients

As the most general form of the test function $f(\cdot,\cdot)$ we consider (2). That is, we assume that $f(X,T)=\min\{K, |X \cap T|\}$, where K is some positive integer. The following theorem provides a closed-form expression for the Rényi coefficients in this case and represents the major input into the non-asymptotic expressions of the upper bounds in specific cases.

Theorem 3.3 *Let the test function be defined by* (2), $0 \le p \le m \le l \le n$, $p < l$, $\mathcal{X}$ *be a balanced test field and* $(T_i, T_j) \in \mathcal{T}(n,l,m,p)$. *Then the value of the Rényi coefficient* k_{ij} *does not depend on the choice of the pair* $(T_i, T_j) \in \mathcal{T}(n,l,m,p)$ *and equals* $k_{ij} = K(\mathcal{X}, n, l, m, p)$, *where*

$$K(\mathcal{X}, n, l, m, p) = \sum_{r=0}^{p} \sum_{u=0}^{m-p} R(n,l,m,p,u,u,r)$$
$$+ \sum_{r=0}^{p} \sum_{u=q}^{l-p} \sum_{v=u+1}^{m-p} R(n,l,m,p,u,v,r) + \sum_{r=0}^{p} \sum_{v=q}^{m-p} \sum_{u=v+1}^{l-p} R(n,l,m,p,u,v,r), \tag{28}$$

with $q = \max\{0, K-r\}$.

Proof. Let $(T_i, T_j) \in \mathcal{T}(n,l,m,p)$ and a be some integer. Introduce the sets

$$\mathcal{X}^{a,a} = \{X \in \mathcal{X} : |X \cap T_i| = a, |X \cap T_j| = a\},$$
$$\mathcal{X}^{a,>a} = \{X \in \mathcal{X} : |X \cap T_i| = a, |X \cap T_j| > a\},$$
$$\mathcal{X}^{>a,a} = \{X \in \mathcal{X} : |X \cap T_i| > a, |X \cap T_j| = a\}.$$

Remind that $k_{ij} = |\{X \in \mathcal{X} : f(X,T_i) = f(X,T_j)\}|$ and $f(X,T)=\min\{K, |X \cap T|\}$.

We have the equality $f(X,T_i) = f(X,T_j)$ if and only if one of the three following cases occurs:

(i) $X \in \mathcal{X}^{a,a}$ for some $a \geq 0$;
(ii) $X \in \mathcal{X}^{a,>a}$ for some $a \geq K$;
(iii) $X \in \mathcal{X}^{>a,a}$ for some $a \geq K$.
Therefore,

$$k_{ij} = \sum_{a \geq 0} |\mathcal{X}^{a,a}| + \sum_{a \geq K} |\mathcal{X}^{a,>a}| + \sum_{a \geq K} |\mathcal{X}^{>a,a}|. \tag{29}$$

By definition, $X \in \mathcal{X}_{uvr}$ for some u, v and r such that the sets $\mathcal{X}_{uvr} = \mathcal{X}_{uvr}(T_i, T_j)$ are defined, see (15). The set of integers n, m, l, p, u, v and r satisfy then the constraints (17). Using these constraints and the definition of the coefficients $R(\cdot)$, see (19), we can re-express the sums in the right-hand side of (29) as follows:

$$\sum_{a \geq 0} |\mathcal{X}^{a,a}| = \sum_{r=0}^{p} \sum_{u=0}^{m-p} |\mathcal{X}_{uur}| = \sum_{r=0}^{p} \sum_{u=0}^{m-p} R(n,l,m,p,u,u,r)\,,$$

$$\sum_{a \geq K} |\mathcal{X}^{a,>a}| = \sum_{r=0}^{p} \sum_{u=q}^{l-p} \sum_{v=u+1}^{m-p} |\mathcal{X}_{uvr}| = \sum_{r=0}^{p} \sum_{u=q}^{l-p} \sum_{v=u+1}^{m-p} R(n,l,m,p,u,v,r)\,,$$

where $q = \max\{0, K-r\}$, and analogously

$$\sum_{a \geq K} |\mathcal{X}^{>a,a}| = \sum_{r=0}^{p} \sum_{v=q}^{m-p} \sum_{u=v+1}^{l-p} |\mathcal{X}_{uvr}| = \sum_{r=0}^{p} \sum_{v=q}^{m-p} \sum_{u=v+1}^{l-p} R(n,l,m,p,u,v,r)\,.$$

By substituting this into (29) we get (28). To finish the proof we just need to mention that the above calculation does not depend on the choice of the pair $(T_i, T_j) \in \mathcal{T}(n,l,m,p)$. □

4 Upper bounds for the length of optimal designs

In this section we use Theorems 2.1, 2.2, 3.1, 3.2 and 3.3 to formulate the upper bounds for the length of optimal designs for the general case of K-channel model and general balanced test fields $\mathcal{X}$. We then make suitable simplifications in specific cases of the additive, binary and multiaccess channel models.

4.1 K–Channel model

Theorem 4.1 *Let the test field $\mathcal{X} \subseteq \mathcal{P}_n^{\leq n}$ be balanced, $1 \leq t \leq n$, the target field $\mathcal{T}$ be either $\mathcal{P}_n^t$ or $\mathcal{P}_n^{\leq t}$, for $p < l$ the numbers $K(\mathcal{X}, n, l, m, p)$ be defined according to (28) and $K(\mathcal{X}, n, l, l, l) = 0$ for all $l = 0, \ldots, n$. Then the upper bound N^* for the length of optimal strongly separating design for the model (2) is*

$$N^* = \min\left\{k : \sum_{l,m} \sum_{p \leq m} Q(n,l,m,p) \left(\frac{K(\mathcal{X},n,l,m,p)}{|\mathcal{X}|}\right)^k < 1\right\}. \tag{30}$$

Analogously, for the γ-separating designs the upper bound can be written as

$$N_\gamma = \min\left\{k : \sum_{l,m} \sum_{p \leq m} Q(n,l,m,p) \left(\frac{K(\mathcal{X},n,l,m,p)}{|\mathcal{X}|}\right)^k < \frac{\gamma|\mathcal{T}|}{2}\right\}. \tag{31}$$

In (30) *and* (31)*, the first summation is taken over* m, l *such that* $0 \le m \le l \le t$ *in the case* $\mathcal{T} = \mathcal{P}_n^{\le t}$*; in the case* $\mathcal{T} = \mathcal{P}_n^t$ *the first summation disappears and* $m = l = t$.

Theorem 4.1 is a direct consequence of Theorems 2.1, 2.2, 3.1, 3.2 and 3.3.

4.2 Additive model

In the additive group testing problem $f(X,T) = |X \cap T|$ and we therefore can take $K = \infty$ in (2) and (28). This removes two terms in (28) and simplifies the expression.

Theorem 4.2 *Let the test function be* $f(X,T) = |X \cap T|$, $0 \le p \le m \le l \le n$, $p < l$, *and* $\mathcal{X}$ *be balanced. Then*

$$K(\mathcal{X}, n, l, m, p) = \sum_{r=0}^{p} \sum_{u=0}^{m-p} R(n, l, m, p, u, u, r)\,. \tag{32}$$

For $\mathcal{X} = \mathcal{P}_n^s$ *the formula* (32) *can be simplified to*

$$K(\mathcal{P}_n^s, n, l, m, p) = \sum_{u=0}^{p} \binom{l-p}{u} \binom{m-p}{u} \binom{n-l-m+2p}{s-2u}\,. \tag{33}$$

Note that to get (33) from (32) and (20) we need to use the Vandermonde convolution formula (26).

As simple consequences, Theorems 4.1 and 4.2 and expressions for $R(n,l,m,p,u,v,r)$ of Theorem 3.2 and Lemma 3.1 imply specific expressions, see Zhigljavsky and Zabalkanskaya (1996), for the upper bounds N^* and N_γ of the additive group testing designs in the cases when $\mathcal{T}$ is either $\mathcal{P}_n^t$ or $\mathcal{P}_n^{\le t}$ and $\mathcal{X}$ is either $\mathcal{P}_n^s$ or $\mathcal{P}_n^{\le s}$.

4.3 Binary model

In the binary group testing $K = 1$ and thus

$$f(X,T) = \begin{cases} 0 & \text{if } |X \cap T| = \emptyset, \\ 1 & \text{otherwise.} \end{cases} \tag{34}$$

Theorem 4.3 *Let the test function be* (34), $0 \le p \le m \le l \le n$, $p < l$ *and* $\mathcal{X}$ *be a balanced test field. Then*

$$K(\mathcal{X}, n, l, m, p) = |\mathcal{X}| - \left(\sum_{u=1}^{l-p} R(n,l,m,p,u,0,0) + \sum_{v=1}^{m-p} R(n,l,m,p,0,v,0) \right)\,. \tag{35}$$

For $\mathcal{X} = \mathcal{P}_n^s$ *the formula* (35) *can be simplified to*

$$K(\mathcal{P}_n^s, n, l, m, p) = \binom{n}{s} - \binom{n-l}{s} - \binom{n-m}{s} + 2\binom{n-l-m+p}{s}\,. \tag{36}$$

Proof. Rewriting (28) for $K = 1$ we obtain

$$K(\mathcal{X}, n, l, m, p) = \sum_{r=0}^{p} \sum_{u=0}^{m-p} R(n, l, m, p, u, u, r) + \sum_{r=1}^{p} \sum_{u=0}^{l-p} \sum_{v=u+1}^{m-p} R(n, l, m, p, u, v, r) +$$

$$\sum_{r=1}^{p} \sum_{u=0}^{m-p} \sum_{v=u+1}^{l-p} R(n,l,m,p,v,u,r) + \sum_{u=1}^{l-p} \sum_{v=u+1}^{m-p} R(n,l,m,p,u,v,0) + \sum_{u=1}^{m-p} \sum_{v=u+1}^{l-p} R(n,l,m,p,v,u,0)$$

$$= \sum_{r=1}^{p} \sum_{u=0}^{l-p} \sum_{v=0}^{m-p} R(n, l, m, p, u, v, r) + \sum_{u=1}^{l-p} \sum_{v=1}^{m-p} R(n, l, m, p, u, v, 0) + R(n, l, m, p, 0, 0, 0).$$

The use of (22) yields (35).

Assume now that $\mathcal{X} = \mathcal{P}_n^s$. Then we can apply (20) and rewrite (35) in the form

$$K(\mathcal{X}, n, l, m, p) = \binom{n}{s} - \sum_{u=1}^{l-p} \binom{l-p}{u} \binom{n-l-m+p}{s-u} - \sum_{v=1}^{m-p} \binom{m-p}{v} \binom{n-l-m+p}{s-v}.$$

Applying now the Vandermonde convolution formula (26) we obtain (36). □

Corollary 4.1 *Assume either* $\mathcal{T} = \mathcal{P}_n^t$ *or* $\mathcal{T} = \mathcal{P}_n^{\le t}$, $t \le n$, *and let the test field* $\mathcal{X}$ *be balanced. Then the upper bounds* N^* *and* N_γ *can be expressed as follows*

$$N^* = \min\Bigg\{ k = 1, 2, \ldots : \sum_{l,m} \sum_{\substack{0 \le p \le m \\ p < l}} Q(n, l, m, p) \times$$

$$\times \left(1 - \frac{1}{|\mathcal{X}|} \left(\sum_{u=1}^{l-p} R(n, l, m, p, u, 0, 0) + \sum_{v=1}^{m-p} R(n, l, m, p, 0, v, 0) \right)\right)^k < 1 \Bigg\}, \tag{37}$$

$$N_\gamma = \min\Bigg\{ k : \sum_{l,m} \sum_{\substack{0 \le p \le m \\ p < l}} Q(n, l, m, p) \times$$

$$\times \left(1 - \frac{1}{|\mathcal{X}|} \left(\sum_{u=1}^{l-p} R(n, l, m, p, u, 0, 0) + \sum_{v=1}^{m-p} R(n, l, m, p, 0, v, 0) \right)\right)^k < \frac{|\mathcal{T}|\gamma}{2} \Bigg\} \tag{38}$$

where the first summation in the above formulae is over $0 \le m \le l \le t$ *for the case* $\mathcal{T} = \mathcal{P}_n^{\le t}$ *and* $l = m = t$; *that is, the first summation disappears in* (37) *and* (38), *in the case* $\mathcal{T} = \mathcal{P}_n^t$.

The binary group testing model is the most popular in literature. In view of its importance we provide below several specific existence theorems for this model. Of course, all these theorems are easy corollaries of (36), (37), (38) and Theorems 3.1, 3.2.

Corollary 4.2 *Let $\mathcal{T} = \mathcal{P}_n^t$ and $\mathcal{X} = \mathcal{P}_n^s$, where $n \geq 2$, $1 \leq t < n$, $1 \leq s < n$. Then there exists a nonsequential group testing design with the sample size $N \leq N^* = N^*(n,t,s)$, where*

$$N^* = \min\left\{k = 1, 2, \ldots : \frac{1}{2}\sum_{p=0}^{t-1} \binom{n}{p\;\; t-p\;\; t-p\;\; n-2t+p} \left(1 - 2 \cdot \frac{\binom{n-t}{s} - \binom{n-2t+p}{s}}{\binom{n}{s}}\right)^k < 1\right\} \tag{39}$$

Corollary 4.3 *Let $\mathcal{T} = \mathcal{P}_n^{\leq t}$ and $\mathcal{X} = \mathcal{P}_n^s$, where $n \leq 2$, $1 \leq t < n$, $1 \leq s < n$. Then there exists a nonsequential group testing design with the sample size $N \leq N^* = N^*(n, \leq t, s) =$*

$$\min\left\{k = 1, 2, \ldots : \frac{1}{2}\sum_{m=1}^{t}\sum_{p=0}^{m-1} \binom{n}{p\;\, m-p\;\, m-p\;\, n-2m+p} \left(1 - 2\frac{\binom{n-m}{s} - \binom{n-2m+p}{s}}{\binom{n}{s}}\right)^k \right.$$
$$\left. + \sum_{l=1}^{t}\sum_{m=1}^{l-1}\sum_{p=0}^{m} \binom{n}{p\;\, l-p\;\, m-p\;\, n-l-m+p} \left(1 - \frac{\binom{n-l}{s} + \binom{n-m}{s} - 2\binom{n-l-m+p}{s}}{\binom{n}{s}}\right)^k < 1\right\}$$

As illustrations, Tables 1 and 2 in the Appendix provide values of $N^*(n,t,s)$ and $N^*(n, \leq t, s)$ for $t = 3$, $s = \lceil \lambda n \rceil$, various n and λ.

The additivity property of the Rényi coefficients (24) implies the following statements.

Corollary 4.4 *Let $\mathcal{T} = \mathcal{P}_n^t$ and $\mathcal{X} = \mathcal{P}_n^{\leq s}$, where $n \leq 2$, $1 \leq t < n$, $1 \leq s \leq n$. Then there exists a nonsequential group testing design with the sample size $N \leq N^* = N^*(n, t, \leq s) =$*

$$\min\left\{k = 1, 2, \ldots : \frac{1}{2}\sum_{p=0}^{t-1} \binom{n}{p\;\, t-p\;\, t-p\;\, n-2t+p} \left(1 - \frac{2\sum_{w=0}^{s}\left(\binom{n-t}{w} - \binom{n-2t+p}{w}\right)}{\sum_{w=0}^{s}\binom{n}{w}}\right)^k < 1\right\}.$$

Corollary 4.5 *Let $n \geq 2$, $1 \leq t < n$, $\mathcal{T} = \mathcal{P}_n^t$ and $\mathcal{X} = \mathcal{P}_n^{\leq n}$. Then there exists a nonsequential group testing design with the sample size $N \leq N^* = N^*(n, t, \leq n) =$*

$$\min\left\{k = 1, 2, \ldots : \frac{1}{2}\sum_{p=0}^{t-1} \binom{n}{p\;\; t-p\;\; t-p\;\; n-2t+p} \left(1 - 2^{1-t} + 2^{p+1-2t}\right)^k < 1\right\}.$$

Corollary 4.6 *Let $n \geq 2$, $1 \leq t < n$, $\mathcal{T} = \mathcal{P}_n^{\leq t}$ and $\mathcal{X} = \mathcal{P}_n^{\leq n}$. Then there exists a nonsequential group testing design with the sample size $N \leq N^* = N^*(n, \leq t, \leq n) =$*

$$\min\left\{k = 1, 2, \ldots : \frac{1}{2}\sum_{m=1}^{t}\sum_{p=0}^{m-1} \binom{n}{p\;\; m-p\;\; m-p\;\; n-2m+p} \left(1 - 2^{1-m} + 2^{p+1-2m}\right)^k \right.$$
$$\left. + \sum_{l=1}^{t}\sum_{m=1}^{l-1}\sum_{p=0}^{m} \binom{n}{p\;\; l-p\;\; m-p\;\; n-l-m+p} \left(1 - 2^{-l} - 2^{-m} + 2^{p+1-l-m}\right)^k < 1\right\}.$$

Obviously, Corollary 4.5 is a special, although important, case of Corollary 4.4. We shall not provide formulae for N^* in other particular cases, they can be derived in similar manner.

As was mentioned earlier, the formulae for N_γ have the form analogous to the formulae for N^* and can be obtained by substituting $\frac{|\mathcal{T}|\gamma}{2}$ for 1 in the right–hand side of the inequality inside the curly brackets in the formulae for N^*. As an example, we formulate the analogue of Corollary 4.2.

Corollary 4.7 *Let $0 < \gamma < 1$, $\mathcal{T} = \mathcal{P}_n^t$ and $\mathcal{X} = \mathcal{P}_n^s$, where $n \geq 2$, $1 \leq t < n$, $1 \leq s < n$. Then there exists a nonsequential group testing design with the sample size $N \leq N_\gamma = N_\gamma(n,t,s) =$*

$$\min\left\{k{=}1,2,\ldots: \frac{1}{2}\sum_{p=0}^{t-1}\binom{n}{p\ t{-}p\ t{-}p\ n{-}2t{+}p}\left(1{-}2\cdot\frac{\binom{n-t}{s}-\binom{n-2t+p}{s}}{\binom{n}{s}}\right)^k < \frac{\gamma}{2}\binom{n}{t}\right\}. \quad (40)$$

4.4 Multiaccess channel

Consider the multiaccess channel problem, where the test function is

$$f(X,T) = \min\{2, |X \cap T|\} = \begin{cases} 0 & \text{if } |X \cap T| = 0, \\ 1 & \text{if } |X \cap T| = 1, \\ 2 & \text{if } |X \cap T| \geq 2. \end{cases} \quad (41)$$

The first step is to deduce from Theorem 3.3 simpler expressions for the Rényi coefficients.

Theorem 4.4 *Let $0 \leq p \leq m \leq l \leq n$, $p < l$ and $\mathcal{X}$ be a balanced test field. Then $K(\mathcal{X}, n, l, m, p) =$*

$$\begin{aligned} &|\mathcal{X}| - \sum_{u=1}^{l-p} R(n,l,m,p,u,0,0) - \sum_{u=1}^{l-p} R(n,l,m,p,u,0,1) - \sum_{u=2}^{l-p} R(n,l,m,p,u,1,0) \\ &- \sum_{v=1}^{m-p} R(n,l,m,p,0,v,0) - \sum_{v=1}^{m-p} R(n,l,m,p,0,v,1) - \sum_{v=2}^{m-p} R(n,l,m,p,1,v,0). \end{aligned} \quad (42)$$

For $\mathcal{X} = \mathcal{P}_n^s$ the formula (42) can be simplified to

$$\begin{aligned} K(\mathcal{P}_n^s, n, l, m, p) &= \binom{n}{s} - \binom{n-l}{s} - \binom{n-m}{s} - l\binom{n-l}{s-1} - m\binom{n-m}{s-1} \\ +2\binom{n-l-m+p}{s} &+ (l+m)\binom{n-l-m+p}{s-1} + 2(l-p)(m-p)\binom{n-l-m+p}{s-2}. \end{aligned} \quad (43)$$

Proof. Let $(T_i, T_j) \in \mathcal{T}(n,l,m,p)$ and $X \in \mathcal{X}_{uvr}(T_i, T_j)$. Applying Theorem 3.3 we get

$$K(\mathcal{X}, n, l, m, p) = \sum_{r=0}^{p}\sum_{u=0}^{m-p} R(n,l,m,p,u,u,r) + \sum_{r=0}^{p}\sum_{u=q}^{l-p}\sum_{v=u+1}^{m-p} R(n,l,m,p,u,v,r)$$

$$+\sum_{r=0}^{p}\sum_{u=q}^{m-p}\sum_{v=u+1}^{l-p} R(n,l,m,p,v,u,r),$$

where $q=\max\{0,2-r\}$. This implies

$$K(\mathcal{X},n,l,m,p)=\sum_{r=0}^{p}\sum_{u=0}^{m-p} R(n,l,m,p,u,u,r)+\sum_{r=2}^{p}\sum_{u=0}^{l-p}\sum_{v=u+1}^{m-p} R(n,l,m,p,u,v,r)$$
$$+\sum_{r=2}^{p}\sum_{u=0}^{m-p}\sum_{v=u+1}^{l-p} R(n,l,m,p,v,u,r)+\sum_{u=1}^{l-p}\sum_{v=u+1}^{m-p} R(n,l,m,p,u,v,1)$$
$$+\sum_{u=1}^{m-p}\sum_{v=u+1}^{l-p} R(n,l,m,p,v,u,1)+\sum_{u=2}^{l-p}\sum_{v=u+1}^{m-p} R(n,l,m,p,u,v,0)$$
$$+\sum_{u=2}^{m-p}\sum_{v=u+1}^{l-p} R(n,l,m,p,v,u,0)=\sum_{r=2}^{p}\sum_{u=0}^{l-p}\sum_{v=0}^{m-p} R(n,l,m,p,u,v,r)$$
$$+\sum_{u=1}^{l-p}\sum_{v=1}^{m-p} R(n,l,m,p,u,v,1)+\sum_{u=2}^{l-p}\sum_{v=2}^{m-p} R(n,l,m,p,u,v,0)$$
$$+R(n,l,m,p,0,0,1)+R(n,l,m,p,1,1,0)+R(n,l,m,p,0,0,0).$$

The use of (22) leads to (42).

Note that (42) can be rewritten in the form

$$K(\mathcal{X},n,l,m,p)=|\mathcal{X}|-\sum_{u=0}^{l-p}[R(n,l,m,p,u,0,0)+R(n,l,m,p,u,0,1)+R(n,l,m,p,u,1,0)]$$
$$-\sum_{v=0}^{m-p}[R(n,l,m,p,0,v,0)+R(n,l,m,p,0,v,1)+R(n,l,m,p,1,v,0)]+2R(n,l,m,p,0,0,0)$$
$$+2R(n,l,m,p,0,0,1)+2R(n,l,m,p,1,1,0)+R(n,l,m,p,0,1,0)+R(n,l,m,p,1,0,0)$$

Assume now that $\mathcal{X}=\mathcal{P}_n^s$. Applying (20), the expression for $K(\cdot)$ in case $\mathcal{X}=\mathcal{P}_n^s$, and six times the Vandermonde convolution formula (26), we obtain (43). □

As an example, let us formulate the existence theorem for the optimal multiaccess channel designs for $\mathcal{T}=\mathcal{P}_n^t$ and $\mathcal{X}=\mathcal{P}_n^s$.

Corollary 4.8 *Let the test function be* (41), $\mathcal{T}=\mathcal{P}_n^t$ *and* $\mathcal{X}=\mathcal{P}_n^s$, *where* $n\geq 2$, $1\leq t<n$, $1\leq s<n$. *Then there exists a strongly separating design with the length*

$$N\leq N^*=N^*(n,t,s)=\min\left\{k:\frac{1}{2}\sum_{p=0}^{t-1}\binom{n}{p\ \ t-p\ \ t-p\ \ n-2t+p}\times\right.$$
$$\left.\times\left(1-2\frac{\binom{n-t}{s}+t\binom{n-t}{s-1}-\binom{n-2t+p}{s}-t\binom{n-2t+p}{s-1}-(t-p)^2\binom{n-2t+p}{s-2}}{\binom{n}{s}}\right)^k<1\right\} \tag{44}$$

Obviously, formulae for N^* and N_γ in other particular cases can be written down in the same manner.

Tables 5 and 6 provide values of $N^*(n,t,s)$ for $\mathcal{T}=\mathcal{P}_n^3$ and $\mathcal{T}=\mathcal{P}_n^5$, $\mathcal{X}=\mathcal{P}_n^s$, $s=\lceil\lambda n\rceil$, various n and λ.

5 Asymptotic bounds

5.1 Auxiliary statements

In this section we present several auxiliary statements which will be used in deriving the asymptotic versions of the upper bounds of Section 4.

Theorem 5.1 *Let I and n_0 be some integers, c_i, r_i, α_i $(i = 1, \ldots, I)$ be some real numbers, $c_i > 0$, $0 < r_i < 1$, at least one of α_i be positive, $\{q_{i,n}\}$, $\{r_{i,n}\}$ be families of positive numbers $(i = 1, \ldots, I,\ \ n \geq n_0)$ such that $0 < r_{i,n} < 1$ for all i and $n \geq n_0$,*

$$q_{i,n} = c_i n^{\alpha_i}(1 + o(1)), \quad r_{i,n} = r_i + o\left(\frac{1}{\log n}\right) \quad \text{as } \ n \to \infty\,, \tag{45}$$

$$N = N(n) = \min\left\{k = 1, 2, \ldots \text{ such that } \ \sum_{i=1}^{I} q_{i,n} r_{i,n}^k < 1\right\}, \tag{46}$$

$L = L(n)$ be the solution of the equation

$$\sum_{i=1}^{I} q_{i,n} r_{i,n}^L = 1\,, \tag{47}$$

$$C = \max_{i=1,\ldots,I} \frac{\alpha_i}{-\log r_i}\,, \tag{48}$$

c be the solution of the equation $\sum_{j\in\mathcal{J}} c_j r_j^c = 1$ where $\mathcal{J}$ is the subset of the set $\{1, \ldots, I\}$ where the maximum in (48) is attained. Then $N(n) = \lfloor L(n)\rfloor + 1$ for all $n \geq n_0$ and

$$L(n) = C\log n + c + o(1) \quad \text{when} \quad n \to \infty\,. \tag{49}$$

Proof. The relation $N(n) = \lfloor L(n)\rfloor + 1$ for all n obviously follows from the definitions (46) and (47) for $N(n)$ and $L(n)$, respectively. The problem is in proving (49).

Let us define the sets

$$\mathcal{I} = \{1, \ldots, I\},\ \ \mathcal{I}_+ = \{i \in \mathcal{I} \text{ such that } \alpha_i > 0\ \},\ \ \mathcal{I}_- = \mathcal{I} \setminus \mathcal{I}_+ = \{i \in \mathcal{I} \text{ such that } \alpha_i \leq 0\}\,.$$

Also, for $i \in \mathcal{I}$ define

$$N(i, n) = \min\left\{k = 1, 2, \ldots \text{ such that } \ q_{i,n} r_{i,n}^k < 1\right\}. \tag{50}$$

Then for all $i \in \mathcal{I}$ and $n = 1, 2, \ldots$ we have $N(i, n) = \lfloor L(i, n)\rfloor + 1$, where $L(i, n)$ are the solutions of the equations $q_{i,n} r_{i,n}^L = 1$; that is,

$$L(i, n) = \frac{\log q_{i,n}}{-\log r_{i,n}}\,.$$

Obviously, $L(i, n) \leq L(n)$ for all i, n and $L(i, n)$ are bounded for $i \in \mathcal{I}_-$.

Consider the asymptotic behaviour, as $n \to \infty$, of $L(i,n)$ for $i \in \mathcal{I}_+$:

$$L(i,n) = -\frac{\log q_{i,n}}{\log r_{i,n}} = -\frac{\log c_i + \alpha_i \log n + o(1)}{\log(r_i + o(\frac{1}{\log n}))} = \frac{\alpha_i \log n + \log c_i}{-\log r_i} + o(1), \quad n \to \infty\,.$$

Let $j \in \mathcal{J}$; that is, j is one of the indices where the maximum in (48) is attained. Then the above asymptotic expressions for $L(i,n)$ imply that for $i \in \mathcal{I}_+ \setminus \mathcal{J}$

$$L(j,n) - L(i,n) = \left(C - \frac{\alpha_i}{-\log r_i} \right) \log n + O(1) \to \infty \quad \text{when} \;\; n \to \infty\,.$$

If $i \in \mathcal{I}_-$ then $L(i,n)$ is bounded and the fact $L(j,n) - L(i,n) \to \infty$ when $n \to \infty$ is obvious. The facts that $q_{i,n} r_{i,n}^{L(i,n)} = 1$ for all i and n (this is the definition of $L(i,n)$) and $L(j,n) - L(i,n) \to \infty$ as $n \to \infty$ for any $j \in \mathcal{J}$ and $i \in \mathcal{I} \setminus \mathcal{J}$ imply for all $i \in \mathcal{I} \setminus \mathcal{J}$

$$q_{n,i} r_{i,n}^{L(n)} \le q_{n,i} r_{i,n}^{L(j,n)} = q_{n,i} r_{i,n}^{L(i,n)} r_{i,n}^{L(j,n)-L(i,n)} = o(1) \quad \text{when} \quad n \to \infty\,,$$

where j is any index in $\mathcal{J}$.

This yields that for large n the equation for $L(n)$ is

$$\sum_{j \in \mathcal{J}} q_{j,n} r_{j,n}^{L(n)} = 1 + o(1) \quad \text{when} \quad n \to \infty\,.$$

This can asymptotically be rewritten as

$$\sum_{j \in \mathcal{J}} c_j n^{\alpha_j} r_j^{L(n)} = 1 + o(1) \quad \text{when} \quad n \to \infty\,. \tag{51}$$

It is straightforward to check that $l(n) = C \log n + c$ is the solution of the equation

$$\sum_{j \in \mathcal{J}} c_j n^{\alpha_j} r_j^{l(n)} = 1 \tag{52}$$

and that $L(n)$ and $l(n)$, the solutions of the equations (51) and (52), are such that $L(n) = l(n) + o(1)$, $n \to \infty$. This completes the proof. □

The following statement is a simple consequence of Theorem 5.1.

Corollary 5.1 *Let the conditions of Theorem* 5.1 *hold and the maximum in* (48) *is attained at a single index* j, $1 \le j \le I$; *that is,* $\mathcal{J} = \{j\}$. *Then*

$$C = -\frac{\alpha_j}{\log r_j}\,, \quad c = -\frac{\log c_j}{\log r_j} \tag{53}$$

and therefore

$$N(n) = \left\lceil \frac{\alpha_j \log n + \log c_j}{-\log r_j} + o(1) \right\rceil \quad \text{as} \quad n \to \infty\,. \tag{54}$$

All the upper bounds of Section 4 have the form (46); therefore, Theorem 5.1 and Corollary 5.1 can potentially be applied. To facilitate the applications, we present two auxiliary statements.

Lemma 5.1 *Let n, u, w be positive integers, $n \to \infty$, u and w be fixed, $s = s(n) = \lambda n + O(1)$ when $n \to \infty$, where $0 < \lambda < 1$. Then*

$$\frac{\binom{n-w}{s-u}}{\binom{n}{s}} = \lambda^u (1-\lambda)^{w-u} + O\left(\frac{1}{n}\right) \text{ when } n \to \infty .$$

The proof directly follows from (27).

Lemma 5.2 *Let the integers p, m, l, n and the coefficients $Q(n, l, m, p)$ be as in Theorem 3.1, p, m, l be fixed and $n \to \infty$. Then*

$$Q(n, l, m, p) = c_{l,m,p} \cdot n^{l+m-p} \left(1 + O\left(\frac{1}{n}\right)\right), \quad n \to \infty , \tag{55}$$

where

$$c_{l,m,p} = \begin{cases} \frac{1}{p!(m-p)!(l-p)!} & \text{if } m \neq l \, , \\ \frac{1}{2p!((m-p)!)^2} & \text{if } m = l \, . \end{cases} \tag{56}$$

Proof. Let $m < l$. Then applying the Stirling formula we obtain

$$Q(n, l, m, p) = \frac{n!}{p!(m-p)!(l-p)!(n-l-m+p)!}$$

$$= \frac{n^{n+1/2}}{p!(m-p)!(l-p)!(n-l-m+p)^{n-l-m+p+1/2} \cdot e^{l+m-p}} \left(1 + O\left(\frac{1}{n}\right)\right)$$

$$= c_{l,m,p} n^{l+m-p} \left(1 + O\left(\frac{1}{n}\right)\right), \quad n \to \infty .$$

The case $m = l$ differs from the case $m < l$ in the multiplier $\frac{1}{2}$ in the original and asymptotic expressions for $Q(n, l, m, p)$. □

5.2 Additive model

Consider the asymptotic behaviour of N^* when n is large. Let $\mathcal{X} = \mathcal{P}_n^s$, $\mathcal{T} = \mathcal{P}_n^t$ and

$$n \to \infty, \;\; t = t(n) \geq 2, \;\; \frac{t(n)}{n} \to 0, \;\; s = s(n) \to \infty, \;\; \frac{s(n)}{n} \to \lambda, \tag{57}$$

where λ, $0 < \lambda < 1$, is some number and $t = t(n)$ is either fixed or increasing with n in such a way that $t(n) = o(n)$, $n \to \infty$. We only consider the case $\mathcal{X} = \mathcal{P}_n^s$ and $\mathcal{T} = \mathcal{P}_n^t$. Note that if the number of defectives is apriori unknown, then it can often be received in one preliminary test by selecting the full set $\mathbf{X}$ as the test group.

Theorem 5.2 *Let* $\mathcal{X}=\mathcal{P}_n^s$, $\mathcal{T}=\mathcal{P}_n^t$, $t \geq 2$ *be fixed,* $n \to \infty$, $s = s(n) = \lambda n + O(1)$ *when* $n \to \infty$, $0 < \lambda \leq \frac{1}{2}$, $0 < \gamma < 1$. *Then* $N(n,t,s) = \lceil N^{(\text{as})}(n,t,\lambda) + o(1) \rceil$ *as* $n \to \infty$, *where*

$$N^*(n,t,\lambda n) \sim N_0^{(\text{as})}(n,t,\lambda) = \lceil g_\lambda \left((t+1)\log_2 n - \log_2 (t-1)! - 1\right) \rceil, \tag{58}$$

$$g_\lambda = \frac{1}{-\log_2(\lambda^2 + (1-\lambda)^2)}$$

and the minimal value of g_λ *is equal to 1 and is achieved at* $\lambda=\frac{1}{2}$. *Also, for the weakly separating designs,*

$$N_\gamma(n,t,n/2) = \frac{t\log_2 n - \log_2 t\gamma}{2t - \log_2 (2t)! + 2\log_2 t!} + o(1), \quad n \to \infty, \tag{59}$$

and the value $\lambda=\frac{1}{2}$ *is optimal in this case as well.*

The theorem can be deduced from Theorems 4.1 and 5.1. We refer to Zhigljavsky and Zabalkanskaya (1996) for its direct proof and a thorough discussion concerning this and similar results for the additive model. We only note that for the optimal value of λ, namely for $\lambda=\frac{1}{2}$, the leading terms in the asymptotics in (58) and (59), for any $0 < \gamma < 1$, are

$$N^*(n,t,n/2) \sim (t+1)\log_2 n \quad \text{and} \quad N_\gamma(n,t,n/2) \sim 2t \log n / \log(\pi t) \tag{60}$$

correspondingly, where $n \to \infty$ and also $t \to \infty$ in such a way that $t/\log n \to 0$. The relationship (60) demonstrates that the problem of weak separation can be resolved more efficiently.

5.3 Binary model

Lemma 5.3 *Let the test function be* (34), $0 \leq p \leq m \leq l \leq n$, $p < l$, *and the test field be* $\mathcal{X} = \mathcal{P}_n^s$. *Then*

$$\frac{K(\mathcal{X}, n, l, m, p)}{|\mathcal{X}|} = p_{n,l,m,p,s}$$

where

$$p_{n,l,m,p,s} = 1 - \frac{1}{\binom{n}{s}} \left[\binom{n-l}{s} + \binom{n-m}{s} - 2\binom{n-l-m+p}{s} \right]. \tag{61}$$

If moreover $s = s(n) = \lambda n + O(1)$, $n \to \infty$, *where* $0 < \lambda < 1$, *then*

$$p_{n,l,m,p,s} = r_{l,m,p}(\lambda) + O\left(\frac{1}{n}\right), \quad n \to \infty, \tag{62}$$

where

$$r_{l,m,p}(\lambda) = 1 - \left[(1-\lambda)^l + (1-\lambda)^m - 2(1-\lambda)^{l+m-p} \right], \quad n \to \infty. \tag{63}$$

The proof easily follows from (36) and the application of Lemma 5.1 three times, with $u = 0$ and $w = l$, $w = m$, $w = l + m - p$.

Strongly separating designs

Here we consider the cases $\mathcal{T} = \mathcal{P}_n^t$ and $\mathcal{T} = \mathcal{P}_n^{\leq t}$, where $t \geq 2$ is a fixed integer. We also assume that $\mathcal{X} = \mathcal{P}_n^s$, the total number of factors n tends to infinity, $s = s(n) = \lambda n + O(1)$, where $n \to \infty$, λ is some fixed number, $0 < \lambda < 1$. There are many different asymptotic formulae for the upper bounds depending on the relation between λ and t. We are mainly interested in the ranges for λ giving the smallest asymptotic rate for the upper bounds $N(n,t,s)$ and $N(n, \leq t, s)$ and we thus do not consider the whole interval (0,1) for λ. By a similar reason we do not consider the test fields $\mathcal{X} = \mathcal{P}_n^{\leq s}$: we did not find cases where these sets provide better upper bounds than the sets $\mathcal{X} = \mathcal{P}_n^s$.

Let us first consider the case when the number of significant factors is known.

Theorem 5.3 *Assume that the test function is* (34), $\mathcal{X} = \mathcal{P}_n^s$, $\mathcal{T} = \mathcal{P}_n^t$, $t \geq 2$ *is fixed,* $n \to \infty$, $s = s(n) = \lambda n + O(1)$ *as* $n \to \infty$, $0 < \lambda \leq \frac{1}{2}$. *Then* $N(n,t,s) = \lceil N^{(\mathrm{as})}(n,t,\lambda) + o(1) \rceil$ *as* $n \to \infty$, *where*

$$N^{(\mathrm{as})}(n,t,\lambda) = \frac{(t+1)\log n - \log(t-1)! - \log 2}{-\log(1 - 2\lambda(1-\lambda)^t)}\,. \tag{64}$$

Proof. Corollary 4.2 implies the following expression for $N(n,t,s)$:

$$N(n,t,s) = \min\left\{k : \sum_{p=0}^{t-1} Q(n,t,t,p) \cdot p_{n,t,t,p,s}^k < 1\right\} \tag{65}$$

where $p_{n,t,t,p,s}$ are defined in (61) with $l = m = t$ and satisfy (62).

The statement follows from Corollary 5.1, Lemma 5.2 and the inequality

$$\frac{2t - p}{-\log r_{t,t,p}(\lambda)} < \frac{t+1}{-\log r_{t,t,t-1}(\lambda)} \tag{66}$$

which holds for any $p = 0, 1, \ldots, t-2$ and $0 < \lambda \leq \frac{1}{2}$; it is equivalent to the fact that the maximum in (48) is attained at the unique value of the summation index: $p = t - 1$; the inequality (66) follows from the fact that for any λ the left-hand side in (66) is an increasing function of p in the interval $[0, t-1]$. $\square$

Corollary 5.2 *For fixed* t *and* n *the minimal value of* $N^{(\mathrm{as})}(n,t,\lambda)$ *defined in* (64) *is reached when* $\lambda = 1/(t+1)$ *and equals*

$$N_*(n,t) = \min_\lambda N^{(\mathrm{as})}(n,t,\lambda) = \frac{(t+1)\log n - \log(t-1)! - \log 2}{-\log\left(1 - 2t^t/(t+1)^{t+1}\right)}\,. \tag{67}$$

The proof is straightforward. The corollary immediately implies that for $n \to \infty$, $t = t(n) \to \infty$ and $t(n)/n \to 0$ we have

$$N_*(n,t) \sim \frac{e}{2} t^2 \log n\,. \tag{68}$$

Consider now the case $\mathcal{T} = \mathcal{P}_n^{\leq t}$.

Theorem 5.4 *Let $\mathcal{X} = \mathcal{P}_n^s$, $\mathcal{T} = \mathcal{P}_n^{\leq t}$, $t \geq 2$ be fixed, $n \to \infty$, $s = s(n) = \lambda n + O(1)$ as $n \to \infty$, $0 < \lambda \leq \frac{1}{2}$. Then $N(n, \leq t, s) = \lceil N^{(\mathrm{as})}(n, \leq t, \lambda) + o(1) \rceil$ as $n \to \infty$, where*

$$N^{(\mathrm{as})}(n, \leq t, \lambda) = \frac{t \log n - \log(t-1)!}{-\log\left(1 + (1-\lambda)^t - (1-\lambda)^{t-1}\right)}\,. \tag{69}$$

Proof. The proof is analogous to the proof of Theorem 5.3, the difference is that we have the summation over m, l, p in the analogue of (65) rather than over p only and that the term (α_0, r_0), where the maximum in (48) is attained, corresponds to the case $(l, m, p) = (t, t-1, t-1)$ rather than to the case $(l, m, p) = (t, t, t-1)$ in the case $\mathcal{T} = \mathcal{P}_n^t$. □

Corollary 5.3 *For fixed t and n the minimal value of $N^{(\mathrm{as})}(n, \leq t, \lambda)$ defined via (69) is reached when $\lambda = 1/t$ and equals*

$$N_*(n, \leq t) = \min_\lambda N^{(\mathrm{as})}(n, \leq t, \lambda) = \frac{t \log n - \log(t-1)!}{-\log\left(1 - (t-1)^{t-1}/t^t\right)}\,. \tag{70}$$

For illustration, Table 3 in the Appendix provides values of $N_*(n, \leq t)$ for various t and n.

Corollary 5.3 implies that for $n \to \infty$, $t = t(n) \to \infty$ and $t(n)/n \to 0$ we obtain

$$N(n, \leq t) \sim et^2 \log n\,. \tag{71}$$

Comparing (68) with (71) we see that, roughly speaking, the binary group testing problem with large number of defectives is twice more complicated in the case $\mathcal{T} = \mathcal{P}_n^{\leq t}$ than in the case $\mathcal{T} = \mathcal{P}_n^t$.

As it was mentioned in the Introduction, the asymptotic bounds (68) and (71) follow also from the results in Dyachkov and Rykov (1983), which were derived by the application of the probabilistic method based on the randomisation scheme $\mathcal{R}'$, see Section 2.4 for a description.

Weakly separating designs

Theorem 5.5 *Let $\mathcal{T}$ be either $\mathcal{P}_n^t$ or $\mathcal{T} = \mathcal{P}_n^{\leq t}$, $t \geq 2$, γ $(0 < \gamma < 1)$ be fixed, $\mathcal{X} = \mathcal{P}_n^s$, $\lambda_t = 1 - 2^{-1/t}$, $s = s(n) = \lambda_t n + O(1)$ when $n \to \infty$. Then $N_\gamma(n, t, s) = \lceil N_\gamma^{(\mathrm{as})}(n, t) + o(1) \rceil$ and $N_\gamma(n, \leq t, s) = \lceil N_\gamma^{(\mathrm{as})}(n, t) + o(1) \rceil$ for $n \to \infty$, where*

$$N_\gamma^{(\mathrm{as})}(n, t) = t \log_2 n - \log_2 \gamma + c \tag{72}$$

and $c = c(t)$ is the solution of the equation

$$\sum_{p=0}^{t-1} 2^{-c(t-p)/t} \frac{t!}{p!(t-p)!^2} = 1\,. \tag{73}$$

Moreover, $\lambda_t = 1 - 2^{-1/t}$ is the optimal value of λ providing the smallest value of the rate of increase of both $N_\gamma(n, t, \lceil \lambda n \rceil)$ and $N_\gamma(n, \leq t, \lceil \lambda n \rceil)$ as $n \to \infty$.

Proof. For $\lambda = \lambda_t = 1 - 2^{-1/t}$ and every $p = 0, \ldots, t-1$ we have

$$\frac{t-p}{-\log_2 r_{t,t,p}(\lambda)} = \frac{t-p}{-\log_2(1 - 1/2 - 1/2 + 2^{1+(p-2t)/t})} = t\,, \tag{74}$$

where the quantities $r_{t,t,p}(\lambda)$ are defined in (63). The Stirling formula implies

$$|\mathcal{P}_n^t| = \binom{n}{t} \sim \frac{n^t}{t!} \quad \text{for } n \to \infty \quad \text{and fixed } t\,.$$

Using this and the equality (74) we obtain that in the sum inside (38) all the terms corresponding to $l = m = t$ and arbitrary $p \leq t-1$ have the same asymptotic rate as $n \to \infty$.

The main statement of the theorem follows now from Theorem 5.1 and the inequality

$$\frac{l+m-p-t}{-\log r_{l,m,p}(\lambda)} < \frac{t-p}{-\log r_{t,t,p}(\lambda)}\,, \tag{75}$$

which holds for every $l{=}1,\ldots,t{-}1$; $m{=}1,\ldots,t{-}1$; $p{=}0,1,\ldots,t{-}2$ and $\lambda = 1 - 2^{-1/t}$. Indeed, substituting this value for λ in (63), the asymptotic formula for the Rényi coefficients, we have (74) and

$$\frac{l+m-p-t}{-\log_2 r_{l,m,p}(\lambda)} = \frac{l+m-p-t}{-\log_2(1 - 2^{-m/t} - 2^{-l/t} + 2^{(p-m-l+t)/t})}.$$

Note that for $l{+}m{-}p{-}t \leq 0$ the inequality (75) holds since the left-hand side of it is non-positive and the right-hand side is positive.

Let $l{+}m{-}p{-}t{>}0$. Simple algebra implies

$$\frac{l{+}m{-}p{-}t}{-\log_2(1{-}2^{-m/t}{-}2^{-l/t}{+}2^{(p-m-l+t)/t})} {=} \frac{t}{1{+}\dfrac{t\log_2(1{+}2^{(l+m-t-p)/t}{-}2^{(m-t-p)/t}{-}2^{(l-t-p)/t})}{t{+}p{-}l{-}m}} {<} t$$

since

$$\frac{t\log_2(1{+}2^{(l+m-t-p)/t}{-}2^{(m-t-p)/t}{-}2^{(l-t-p)/t})}{t{+}p{-}l{-}m} > 0$$

for $l < t, m < t, p < t-2$.

Let us show now that $\lambda = 1 - 2^{-1/t}$ is the optimal value of $\lambda = \lambda_t$. Indeed, consider the term corresponding to $l{=}m{=}t, p{=}0$. Then $\lambda = 1 - 2^{-1/t}$ is the value where $t/(-\log_2 r_{t,t,0})$ achieves the minimum and therefore Theorem 5.1 implies that any other value of λ gives a worse asymptotic rate for N_γ. □

For illustration, Table 4 in the Appendix compares values of $N_*(n,t)$, defined in (67), and $N_\gamma^{(\mathrm{as})}(n,t)$, see (72), for $\mathcal{T} = \mathcal{P}_n^t$ and different t, γ and n.

5.4 Multiaccess channel

Lemma 5.4 *Let the test function be* (41), p, m, l *be integers such that* $0 \le p \le m \le l$, $p < l$, *and let* $\mathcal{X} = \mathcal{P}_n^s$, $0 < \lambda < 1$, $s = s(n) = \lambda n + O(1)$, $n \to \infty$. *Then*

$$\frac{K(\mathcal{X}, n, l, m, p)}{|\mathcal{X}|} = p_{l,m,p}(\lambda) + O\left(\frac{1}{n}\right), \quad n \to \infty\,,$$

where

$$p_{l,m,p}(\lambda) = 1 - (1 + (l-1)\lambda)(1-\lambda)^{l-1} - (1 + (m-1)\lambda)(1-\lambda)^{m-1} + \\ 2(1-\lambda)^{l+m-p} + (l+m)(1-\lambda)^{l+m-p-1} + 2(l-p)(m-p)(1-\lambda)^{l+m-p-2}\,. \tag{76}$$

Proof is analogous to the proof of Lemma 5.3 and relies on the use of the formula (43) for $K(\cdot)$ and multiple application of the combinatorial identity (27) and Lemma 5.1. □

This lemma will now be applied to deriving the asymptotic expressions for the upper bounds for the length of strongly separating designs in cases $\mathcal{T} = \mathcal{P}_n^t$ and $\mathcal{T} = \mathcal{P}_n^{\le t}$. We then optimize the choice of λ in both cases and consider the asymptotics when t is large.

Theorem 5.6 *Let the test function be* (41), $\mathcal{X} = \mathcal{P}_n^s$, $\mathcal{T} = \mathcal{P}_n^t$, $t \ge 2$ *be fixed,* $n \to \infty$, $s = s(n) = \lambda n + O(1)$ *when* $n \to \infty$, $0 < \lambda < 1$. *Then* $N(n,t,s) = \lceil N^{(\mathrm{as})}(n,t,\lambda) + o(1) \rceil$ *as* $n \to \infty$, *where*

$$N^{(\mathrm{as})}(n,t,\lambda) = \frac{(t+1)\log n - \log(t-1)! - \log 2}{-\log(1 - 2\lambda(1-\lambda)^{t-1}(1 + \lambda(t-2)))}\,. \tag{77}$$

Proof. Corollary 4.8 and Theorem 5.4 imply that $N(n,t,s)$ can be written as

$$N(n,t,s) = \min\left\{k = 1, 2, \ldots : \sum_{p=0}^{t-1} Q(n,t,t,p) \cdot r_{n,t,t,p,s}^k < 1\right\}, \tag{78}$$

where

$$r_{n,t,t,p,s} = r_{t,t,p}(\lambda) + O\left(\frac{1}{n}\right), \quad n \to \infty\,,$$

and, according to (76),

$$r_{t,t,p}(\lambda) = 1 - 2(1 + (t-1)\lambda)(1-\lambda)^{t-1} + 2(1-\lambda)^{2t-p-2}\left((1-\lambda)^2 + t(1-\lambda) + (t-p)^2\right)\,.$$

The statement of the theorem follows now from Corollary 5.1 and the inequality

$$\frac{2t-p}{-\log r_{t,t,p}(\lambda)} < \frac{t+1}{-\log r_{t,t,t-1}(\lambda)}\,,$$

which holds for every $p = 0, 1, \ldots, t-2$ and $0 < \lambda < 1$; this inequality is equivalent to the fact that the maximum (over p) in (48) is attained at $p = t-1$ and only at $p = t-1$. □

The optimal value of λ, which maximizes the denominator in the expression (77), is

$$\lambda_t = (t - 4 + \sqrt{5t^2 - 12t + 8})/(2t^2 - 2t - 4)\,. \tag{79}$$

Note that λ_t is a continuous function of t and is defined for all $t \geq 1$; for $t = 2$ we have $\lambda_t = \frac{1}{2}$ by continuity.

For illustration, Table 7 in the Appendix provides values of $N^{(\mathrm{as})}(n, t, \lambda)$, defined in (77), for $\mathcal{X} = \mathcal{P}_n^s$, $\mathcal{T} = \mathcal{P}_n^t$, $t = 4, 5, 6$, various n and $\lambda = \lambda_t$.

It is interesting to note the following bounds for λ_t in (79), expressed in terms of the Golden Section $\varphi = \frac{\sqrt{5}+1}{2} \simeq 1.618034$:

$$\frac{\varphi}{t} - \frac{\varphi}{t^2} \;<\; \lambda_t \;<\; \frac{\varphi}{t}\,, \quad 1 \leq t < \infty,$$

which implies, in particular, that $\lambda_t \sim \frac{\varphi}{t}$ as $t \to \infty$.

This yields the following corollary.

Corollary 5.4 *For fixed t, the minimal value of $\min_s N(n, t, s)$ defined in (77) is reached when $s = \lambda_t n + O(1)$, $n \to \infty$, and asymptotically, when $t = t(n) \to \infty$, $t/n \to 0$, equals*

$$\min_s N(n, t, s) \sim (\varphi - \frac{3}{2}) e^{\varphi} t^2 \log n\,,$$

where $(\varphi - \frac{3}{2}) e^{\varphi} \simeq 0.595265$.

Let us now consider the case $\mathcal{T} = \mathcal{P}_n^{\leq t}$.

Theorem 5.7 *Let the test function be (41), $\mathcal{X} = \mathcal{P}_n^s$, $\mathcal{T} = \mathcal{P}_n^{\leq t}$, $t > 1$ be fixed, $n \to \infty$, $s = s(n) = \lambda n + O(1)$ as $n \to \infty$, $0 < \lambda < 1$. Then for $n \to \infty$ the following holds:*

(i) *If either $t = 2$ or $t = 3$ then $N(n, \leq t, s) = \lceil N^{(\mathrm{as})}(n, t, \lambda) + o(1) \rceil$, where $N^{(\mathrm{as})}(n, t, \lambda)$ is defined in (77);*

(ii) *If $t > 3$ then $N(n, \leq t, s) = \lceil N^{(\mathrm{as})}(n, \leq t, \lambda) + o(1) \rceil$, where*

$$N^{(\mathrm{as})}(n, \leq t, \lambda) = \frac{t \log n - \log(t-1)!}{-\log(1 - \lambda(1-\lambda)^{t-2}(1 + \lambda(t-2))}\,. \tag{80}$$

Proof is analogous to the proof of the previous theorem, the difference is that for $t > 3$ the decisive term in the expression for $N(n, \leq t, s)$ corresponds to $l = t, m = p = t - 1$ rather than to $l = m = t, p = t - 1$; at the same time, for $t = 2$ and $t = 3$ this term corresponds to $l = m = t, p = t - 1$, as in Theorem 5.6. □

The optimal value of λ, which maximizes the denominator in the expression (80), is

$$\lambda_{\leq t} = (t - 3 + \sqrt{5t^2 - 14t + 9})/(2t^2 - 4t) \tag{81}$$

(note that the expression (80) can be applied only for $t \geq 4$).

For illustration, Table 8 in the Appendix provides values of $N^{(\mathrm{as})}(n, \leq t, \lambda)$ for $\mathcal{X} = \mathcal{P}_n^s$, $\mathcal{T} = \mathcal{P}_n^{\leq t}$, $t = 4, 5, 6$, $\lambda = \lambda_{\leq t}$ and various n.

Analogously to the bounds for λ_t, we have

$$\frac{\varphi}{t} < \lambda_{\leq t} \;<\; \frac{\varphi}{t} + \frac{\varphi}{6t^2} \quad \text{for } \; t \geq 4,$$

which again gives the asymptotics $\lambda_{\leq t} \sim \frac{\varphi}{t}$ when $t \to \infty$.

Corollary 5.5 *For fixed $t \geq 4$ the minimal value of* $\min_s N(n,\leq t,s)$ *defined in* (80) *is reached when $s=\lambda_{\leq t}\, n+O(1)$, $n \to \infty$, and asymptotically, when $t=t(n) \to \infty$, $t/n \to 0$, equals*

$$\min_s N(n, \leq t, s) \sim (2\varphi - 3)e^{\varphi} t^2 \log n$$

where $(2\varphi - 3)e^{\varphi} \simeq 1.19053$.

Corollaries 5.4 and 5.5 imply that, similarly to the binary model, for large t the case $\mathcal{T} = \mathcal{P}_n^{\leq t}$ can be thought as twice more difficult than the case $\mathcal{T} = \mathcal{P}_n^t$.

6 Two Generalizations

6.1 Search with lies

Let us return to the concept of a general discrete search problem considered in Section 2. Assuming that some test results may be wrong, we say that we have a search problem with lies (errors). In this case the test result $Y = Y(X, T)$ at $X \in \mathcal{X}$ may differ from $f(X, T)$.

There are different search models with lies. We only consider the problems, called *L-lie search problems*, where the number of wrong answers is bounded by a given number $L \geq 0$. Also, we only consider the existence theorems for the strongly separating designs (the case of weakly separating designs can be considered analogously).

Let us mention that if a search problem is solvable then it is solvable as the L-lie search problem as well. Indeed, to provide a strongly separating design for the L-lie problem one may take a strongly separating design for the ordinary 0-lie search problem and repeat all the tests $2L+1$ times. Analogously to the 0-lie case, strong separating designs should provide unique identifiability of all $T \in \mathcal{T}$.

An important observation is that if a non-sequential design $\mathcal{D}_N = \{X_1, \ldots, X_N\}$ is applied in a general L-lie search problem, then one can guarantee that the target can be uniquely defined if and only if the two vectors $F_T = (f(X_1, T), \ldots, f(X_N, T))$ and $F_{T'} = (f(X_1, T'), \ldots, f(X_N, T'))$ differ in at least $2L+1$ components where (T, T') is any pair of different targets in $\mathcal{T}$. Let us formulate this observation as a proper statement.

Proposition 6.1. *Let $\{\mathcal{T}, \mathcal{X}, f, \mathcal{Y}\}$ be a general solvable L-lie search problem. A non-sequential design $\mathcal{D}_N = \{X_1, \ldots, X_N\}$ is strongly separating if and only if for any $T, T' \in \mathcal{T}$, $T \neq T'$*

$$d_H(F_T, F_{T'}) \geq 2L + 1\,, \qquad (82)$$

where $F_T = (f(X_1, T), \ldots, f(X_N, T))\,,\ F_{T'} = \left(f(X_1, T^{'}), \ldots, f(X_N, T^{'})\right)$ *and* $d_H(\cdot, \cdot)$ *is the Hamming distance in* $\mathcal{Y}^N$.

Recall that the Hamming distance between two vectors $F = (f_1, \ldots, f_N)$ and $F' = (f_1^{'}, \ldots, f_N^{'})$ is the number of components of F and F' that are different; that is,

$$d_H(F, F^{'}) = \text{ the number of } i \ \ (1 \leq i \leq N) \ \text{ such that } f_i \neq f_i^{'}\,.$$

The following statement is a generalization of Theorem 2.1 to the case of L-lie search problem.

Theorem 6.1 (Existence theorem for the L-lie search problem.) *Let $\{\mathcal{T},\mathcal{X},f,\mathcal{Y}\}$ be a solvable L-lie search problem and*

$$k_{ij}=k(T_i,T_j)=\mid\{X\in\mathcal{X}:\quad f(X,T_i)=f(X,T_j)\}\mid$$

be the Rényi coefficients. Then there exists a non-sequential strongly separating design with the length

$$N\le N^*(L)=\min\left\{k=1,2,\ldots:\ \sum_{i=1}^{|\mathcal{T}|}\sum_{j=1}^{i-1}\sum_{l=0}^{2L}\binom{k}{l}(p_{ij})^{k-l}(1-p_{ij})^l\le 1\right\}, \qquad (83)$$

where $p_{ij}=k_{ij}/|\mathcal{X}|$.

Proof. For a given design $\mathcal{D}_N=\{X_1,\ldots,X_N\}$, consider the matrix

$$\mathcal{A}_N=\|f(X_i,T_j)\|_{i,j=1}^{N,|\mathcal{T}|}$$

which rows correspond to the test sets X_i and the columns correspond to the targets T_j.

According to (82) the design $\mathcal{D}_N$ is strongly separating if all the pairs (a_i,a_j), $i\ne j$, of the columns of $\mathcal{A}_N$ have at least $2L+1$ different components; that is, $d_H(a_i,a_j)\ge 2L+1$.

Let $(X_1,X_2,\ldots,X_N)$ be a random sample from $\mathcal{X}$. Then for any fixed pair (i,j) such that $i\ne j$ $(i,j=1,\ldots,|\mathcal{T}|)$ and any integer l $(0\le l\le N)$ we have

$$\Pr\{d_H(a_i,a_j)=l\}=\binom{N}{l}(p_{ij})^{N-l}(1-p_{ij})^l$$

and therefore

$$\Pr\{d_H(a_i,a_j)\le 2L\}=\sum_{l=0}^{2L}\binom{N}{l}(p_{ij})^{N-l}(1-p_{ij})^l$$

This yields

$$\Pr\{\text{design }\mathcal{D}_N\text{ is strongly separating}\}$$

$$=\Pr\{d_H(a_i,a_j)\ge 2L+1\text{ for all }i,j=1,\ldots,|\mathcal{T}|,\ i\ne j\}$$

$$=1-\Pr\{d_H(a_i,a_j)\le 2L\text{ for at least one pair }(T_i,T_j)\in\mathcal{T}\times\mathcal{T},\ i\ne j\}$$

$$\ge 1-\sum_{i=1}^{|\mathcal{T}|}\sum_{j=1}^{i-1}\Pr\{d_H(a_i,a_j)\le 2L\}=1-\sum_{i=1}^{|\mathcal{T}|}\sum_{j=1}^{i-1}\sum_{l=0}^{2L}\binom{N}{l}(p_{ij})^{N-l}(1-p_{ij})^l$$

Applying Proposition 6.1 we obtain the required. $\square$

One can consider a version of the L-lie search problem where all wrong answers are the same; that is, the wrong results are equal to some $y \in \mathcal{Y}$, and this value y can be obtained by correct answers as well. This problem is a little simpler than the general L-lie problem and in this problem it is enough to ensure that

$$d_H(F_T, F_{T'}) \geq L + 1,$$

rather than (82), to guarantee the strong separability of a design. For this problem the upper bound (83) is reduced to

$$N \leq \min \left\{ k = 1, 2, \ldots : \sum_{i=1}^{|\mathcal{T}|} \sum_{j=1}^{i-1} \sum_{l=0}^{L} \binom{k}{l} (p_{ij})^{k-l} (1 - p_{ij})^l \leq 1 \right\} . \tag{84}$$

In Section 3 for several setups of the group testing problem we have derived the closed-form expressions for the Rényi coefficients k_{ij}; we therefore can easily compute the upper bounds (83) and (84) for the corresponding L-lie group testing problems as well. These bounds will look similar to the ones formulated in Section 4 (but, of course, there is an extra sum in the right-hand side of the corresponding expressions).

Let us now consider the modifications required to apply the asymptotic formulae of Section 5 to the general L-lie search problem considered in Theorem 6.1 and compare the L-lie and 0-lie problems in terms of the leading asymptotic terms. (In a simpler case leading to (84), in all related formulae $2L$ has to be substituted by L.)

Note first that the upper bounds $N^*(L)$ computed according to (83), form an increasing sequence; that is, $N^*(0) \leq N^*(1) \leq \ldots$ In all asymptotic considerations above $N^*(0) \to \infty$ and thus in all corresponding L-lie problems $N^*(L) \to \infty$ for every L.

For fixed i and j denote $p = p_{ij} = k_{ij}/|\mathcal{X}|$ and consider the terms

$$p_{(k,l)} = \binom{k}{l} p^{k-l}(1-p)^l \quad (l = 0, \ldots, 2L)$$

in the right-hand side of (83). Since we are interested in the values of k such that $k \geq N^*(0) \to \infty$, for each $l < 2L$ we have

$$\frac{p_{(k,l)}}{p_{(k,2L)}} = \frac{(k-2L)!(2L)!}{(k-l)!\, l!} \left(\frac{p}{1-p} \right)^{2L-l} = \frac{(2L)!}{l!} \left(\frac{p}{1-p} \right)^{2L-l} \frac{1}{k^{2L-l}} (1 + O(1)) = O\left(\frac{1}{k}\right)$$

as $k \to \infty$. This implies that the terms in the sum in the rigth-hand side of (83) corresponding to $l = 2L$ dominate all the other terms:

$$\frac{\sum_{i,j} \sum_{l=0}^{2L} \binom{k}{l} (p_{ij})^{k-l}(1-p_{ij})^l}{\sum_{i,j} \binom{k}{2L} (p_{ij})^{k-2L}(1-p_{ij})^{2L}} = 1 + O\left(\frac{1}{k}\right) \quad \text{as } k \to \infty\,.$$

This yields the following result.

Proposition 6.2. *For a fixed L, consider a family of solvable L-lie search problems $\{\mathcal{T}, \mathcal{X}, f, \mathcal{Y}\}_n$ such that $N^*(0) \to \infty$ as $n \to \infty$. Then*

$$N^*(L) \sim N_L = \min \left\{ k = 1, 2, \ldots : \sum_{i=1}^{|\mathcal{T}|} \sum_{j=1}^{i-1} \binom{k}{2L} (p_{ij})^{k-2L} (1 - p_{ij})^{2L} \leq 1 \right\} \tag{85}$$

as $n \to \infty$, where $N^*(L)$ is defined in (83) and both $N^*(L)$ and N_L depend on n.

For all $k \geq 6$ and $L = 0, 1, \ldots, \lfloor k/4 \rfloor$ we have

$$\binom{k}{2L} \leq (k-2L)^{2L}\,.$$

Using this inequality, the inequalities $1 - p_{ij} \leq 1$ for all i, j and introducing the new variable $m = k - 2L$ we obtain from the definition of N_L, see (85), that for fixed $L \geq 0$ and large enough n (here n is such that $N^*(0) \geq \max\{6, 4L\}$):

$$N_L \leq 2L + \min\left\{ m = 1, 2, \ldots \;\; \text{such that} \;\; m^{2L} \sum_{i=1}^{|\mathcal{T}|} \sum_{j=1}^{i-1} (p_{ij})^m \leq 1 \right\}\,. \tag{86}$$

Note that in the right-hand side of (86) all the terms p_{ij}^m decrease exponentially fast as $m \to \infty$, while the multiplier m^{2L} (which can be considered as a penalty for lies) increases only as some power of m. This implies that we should expect that the asymptotic behaviour of $N^*(L)$ is not much different from the asymptotic behaviour of $N^*(0)$ as $n \to \infty$.

Proposition 6.3. *Consider a family of L-lie search problems $\{\mathcal{T}, \mathcal{X}, f, \mathcal{Y}\}_n$ such that $p_{ij} \leq r < 1$ for some constant r and all n, i and j such that $i \neq j$. Assume also that $N^*(0) \sim C \log n$ as $n \to \infty$ for some constant C. Then for any $L \geq 0$ and $\varepsilon > 0$ there exists $n^* = n^*(L, \varepsilon)$ such that*

$$N^*(L) \leq (C + \varepsilon) \log n \tag{87}$$

for all $n \geq n^*$.

Proof. For every n and some $\varepsilon > 0$, let us consider $m = (1+\varepsilon)N^*(0)$ in the sum in the right-hand side of (86):

$$m^{2L} \sum_{i=1}^{|\mathcal{T}|} \sum_{j=1}^{i-1} (p_{ij})^m = m^{2L} \sum_{j<i} (p_{ij})^{N^*(0)} (p_{ij})^{\varepsilon N^*(0)} \leq m^{2L} r^{\varepsilon N^*(0)} \sum_{j<i} (p_{ij})^{N^*(0)}\,. \tag{88}$$

The definition of $N^*(0)$ implies that the sum in the right-hand side of (88) does not exceed 1; the multiplier m^{2L} has the order of growth of a power of $\log n$:

$$m^{2L} = (1+\varepsilon)^{2L} (N^*(0))^{2L} \sim \left(C(1+\varepsilon)\log n\right)^{2L}, \;\; n \to \infty;$$

at the same time, the multiplier $r^{\varepsilon N^*(0)}$ decreases as a power of n:

$$r^{\varepsilon N^*(0)} \sim r^{C\varepsilon \log n} = \frac{1}{n^b}, \;\; n \to \infty,$$

where $b = -C\varepsilon \log r > 0$. This implies that the right-hand side of the inequality (88) tends to 0 as $n \to \infty$ and will certainly be smaller than 1 for all large enough n. In view of the inequality (86) this yields

$$N_L \leq 2L + (1+\varepsilon) C \log n$$

for all n large enough. The last inequality and the asymptotic relation (85) immediately imply the statement of the proposition. □

All the group testing problems considered above satisfy the conditions of Proposition 6.3. Therefore, we can apply it and deduce that the asymptotic behaviour of the L-lie upper bounds $N^*(L)$ for all these group testing problems asymptotically almost coincide (in the sense defined in Proposition 6.3) with the asymptotic behaviour of the 0-lie upper bound $N^*(0)$.

6.2 Binomial sample group testing problems

In this section we show how to modify the results of the paper for the *binomial sample group testing problem*, that is for the problem of finding all defectives in a binomial sample. Specifically, assume that each element in the set $\mathbf{X} = \{x_1, \ldots, x_n\}$ has a prior probability q to be defective and the events 'element x_i is defective' are independent. We also assume that the test function is defined by (2) and the test field $\mathcal{X}$ is balanced, for example, $\mathcal{X} = \mathcal{P}_n^s$.

Strongly separating designs do not have much sense in this case. Unlike them, the weakly separating designs can be of interest.

First, let us use the notation of Section 2.1 and generalize the definition of γ-separating designs, see (3).

Let $\{\mathcal{T}, \mathcal{X}, f, \mathcal{Y}\}$ be a solvable search problem and μ be a prior probability measure on $\mathcal{T}$. A design $\mathcal{D}_N$ is called *γ-separating* with respect to μ if

$$\mu\{T \in \mathcal{T} :\ \text{design } \mathcal{D}_N \text{ separates } T \text{ in } \mathcal{T}\} \geq 1 - \gamma\,, \tag{89}$$

where γ is a fixed number, $0 \leq \gamma \leq 1$.

In the binomial sample group testing problems $\mathcal{T} = \mathcal{P}_n^{\leq n}$ (that is, $\mathcal{T}$ is the set of all possible subsets of $\mathbf{X}$) and μ is such that

$$\mu_t = \mu\{|T| = t\} = \binom{n}{t} q^t (1-q)^{n-t} \quad \text{for } \ t = 0, \ldots, n\,.$$

Set

$$q_{\mathcal{X},n,t,m,p} = \begin{cases} K(\mathcal{X}, n, t, m, p)/|\mathcal{X}| & \text{if } m \leq t \\ K(\mathcal{X}, n, m, t, p)/|\mathcal{X}| & \text{if } t \leq m\,, \end{cases}$$

where $K(\mathcal{X}, n, m, t, p)$ are the Rényi coefficients for the test function (2), see Theorem 3.3 (note that using the convention $K(\mathcal{X}, n, t, t, t) = 0$ of Theorem 4.1, we have $q_{\mathcal{X},n,t,t,t} = 0$ for all $t = 0, \ldots, n$).

Theorem 6.2 (Existence theorem for weakly separating designs in the binomial sample group testing problem.) *Let us consider the binomial sample group testing problem, where the test function is defined by* (2)*, the test field $\mathcal{X}$ is balanced, each element of $\mathbf{X}$ has a prior probability q to be defective and the events 'element x_i is defective' are independent.*

Then for any $0 < \gamma < 1$ *there exists a non-sequential* γ*-separating design with the sample size* $N \le N(n, q, \gamma) =$

$$\min\left\{k\colon \sum_{t=0}^{n} q^t(1-q)^{n-t} \min\left\{\binom{n}{t}, \sum_{m=0}^{n} \sum_{p=0}^{\min\{t,m\}} \binom{n}{p\ m-p\ t-p\ n-t-m+p} q^k_{\mathcal{X},n,t,m,p}\right\} < \gamma\right\}$$

Proof. Let us extend the arguments proving Theorem 2.2. Let, as above, $\mathcal{D}_N = \{X_1, \ldots, X_N\}$ be a random design. Then

$$\mathcal{P}_N = \Pr\{T \text{ is not uniquely determined after application of } \mathcal{D}_N\} = \sum_{t=0}^{n} \mu_t P_{N,n,t}(\mathcal{X}),$$

where $P_{N,n,t}(\mathcal{X})$ is the probability

$$P_{N,n,t}(\mathcal{X}) = \Pr\{T \text{ is not uniquely determined after application of } \mathcal{D}_N |\, |T| = t\}.$$

Note that $\mathcal{X}$ is a balanced test field and therefore the probability $P_{N,n,t}(\mathcal{X})$ is correctly defined; that is, it does not depend on the choice of a particular T such that $|T| = t$.

According to the definition of γ-separability, see (89), we need to choose N large enough to guarantee

$$\sum_{t=0}^{n} \mu_t P_{N,n,t}(\mathcal{X}) < \gamma. \tag{90}$$

For a pair (T, T') of different elements of $\mathcal{T}$ set $P(N, T, T')$ to be the probability of the event that T and T' are not separated after N random tests. If $T = T_i$ and $T' = T_j$ then, in the notation of Section 2.1, $P(1, T, T') = p_{ij} = k_{ij}/|\mathcal{X}|$, where k_{ij} are the Rényi coefficients and $P(N, T, T') = (P(1, T, T'))^N$.

Similarly to the proof of Theorem 2.2, for a fixed T, such that $|T| = t$, the probability $P_{N,n,t}(\mathcal{X})$ that after N random tests T is not separated from all $T' \ne T$, is less than or equal to

$$P_{N,n,t}(\mathcal{X}) \le Q_{N,n,t}(\mathcal{X})$$

where

$$Q_{N,n,t}(\mathcal{X}) = \min\{1, \sum_{T' \ne T} P(N, T, T')\} = \min\{1, S_1 + S_2 + S_3\}. \tag{91}$$

Here

$$S_1 = \sum_{T' : |T'| < t} P(N, T, T'), \quad S_2 = \sum_{T' \ne T, |T'| = t} P(N, T, T'), \quad S_3 = \sum_{T' : |T'| > t} P(N, T, T').$$

The arguments analogous to those leading to (31) give

$$S_1 = \frac{1}{\binom{n}{t}} \sum_{m=0}^{t-1} \sum_{p=0}^{m} Q(n, t, m, p) \left(\frac{K(\mathcal{X}, n, t, m, p)}{|\mathcal{X}|}\right)^N,$$

$$S_2 = \frac{2}{\binom{n}{t}} \sum_{p=0}^{t-1} Q(n,t,t,p) \left(\frac{K(\mathcal{X},n,t,t,p)}{|\mathcal{X}|} \right)^N$$

and

$$S_3 = \frac{1}{\binom{n}{t}} \sum_{m=t+1}^{n} \sum_{p=0}^{t} Q(n,t,m,p) \left(\frac{K(\mathcal{X},n,m,t,p)}{|\mathcal{X}|} \right)^N .$$

Using the definition of $q_{\mathcal{X},n,t,m,p}$ we obtain

$$S_1 + S_2 + S_3 = \frac{1}{\binom{n}{t}} \sum_{m=0}^{n} \sum_{p=0}^{\min\{t,m\}} \binom{n}{p\ m{-}p\ t{-}p\ n{-}t{-}m{+}p} q^N_{\mathcal{X},n,t,m,p} .$$

The inequality

$$\mathcal{P}_N = \sum_{t=0}^{n} \mu_t P_{N,n,t}(\mathcal{X}) \le \sum_{t=0}^{n} \mu_t Q_{N,n,t}(\mathcal{X}) = \sum_{t=0}^{n} \binom{n}{t} q^t (1-q)^{n-t} \min\{1, S_1+S_2+S_3\}$$

and the definition of γ-separability, see (90), implies the statement of the theorem. □

The binomial sample group testing problem is more difficult than the related group testing problem for $\mathcal{T} = \mathcal{P}_n^{\le \tau}$ with $\tau = nq$ (here τ is the average number of defective elements in $\mathbf{X}$) in the following two aspects: (i) we have to test the separation of the unknown target T from all the alternatives $T' \in \mathcal{P}_n^{\le n}$ rather than only from those with $|T'| \le |T|$, and (ii) the number of defective elements can be larger than τ.

The first difficulty is related to the inclusion of the term S_3 into the sum $S_1 + S_2 + S_3$ in the right-hand side of (91). This term may dominate the other two terms (S_1 and S_2 correspond to the alternatives T' with $|T'| < |T|$ and $|T'| = |T|$, respectively) and thus the derivation of the asymptotic formulae for the upper bounds may change.

The second difficulty can be dealt with by restricting the number of defective elements in T by the smallest $\tau_\delta = \tau(n,q,\delta)$ such that

$$\Pr\{|T| \le \tau_\delta\} = \sum_{t=0}^{\tau_\delta} \binom{n}{t} q^t (1-q)^{n-t} \ge 1 - \delta \tag{92}$$

and leaving the other part of γ for the probability of separating the elements of $\mathcal{P}_n^{\le \tau_\delta}$. Here δ can be chosen as any positive number such that $\delta < \gamma$.

The quantity τ_δ is just $(1-\delta)$-quantile of the Binomial distribution $\mathrm{Bin}(n,q)$. If n is large then either normal or Poisson (if the value of nq is small) approximations can be used to approximate τ_δ. Using the normal approximation, we obtain $\tau_\delta \simeq nq + t_\delta \sqrt{nq(1-q)}$, where t_δ is the $(1-\delta)$-quantile of the standard normal distribution. We can also use the Chebyschev inequality $\Pr\{\xi \ge E\xi + a\} \le \mathrm{var}(\xi)/a^2$, where $\xi \sim \mathrm{Bin}(n,q)$, $E\xi = nq$ and $\mathrm{var}(\xi) = nq(1-q)$. Solving the equation $\mathrm{var}(\xi)/a^2 = \delta$ with respect to a, we obtain $\tau_\delta \le nq + \sqrt{nq(1-q)/\delta}$. In either case, τ_δ can be expressed as $\tau_\delta = nq + c\sqrt{nq}$, where c is some constant, $c \le \sqrt{(1-q)/\delta}$.

We use these arguments to simplify the upper bound of Theorem 6.2 so that the results of Section 5 can be directly used to obtain the asymptotic versions of the upper bounds in the binomial sample group testing problems.

Proposition 6.4. *Consider the binomial sample group testing problem with the probability q for the elements of $\mathbf{X}$ to be defective. Let also δ and γ be such that $0 < \delta < \gamma < 1$ and τ_δ is the $(1-\delta)$-quantile of the Binomial distribution $\mathrm{Bin}(n, q)$. Then there exists a non-sequential γ-separating design with the sample size*

$$N \leq N_\theta(n, \leq \tau_\delta)\,, \tag{93}$$

where $\theta = (\gamma - \delta)/(1 - \delta)$ and $N_\theta(n, \leq t)$ is the length of the optimum θ-separating design for $\mathcal{T} = \mathcal{P}_n^{\leq t}$, the same test field and the same test function.

Proof. In the notation used in the proof of Theorem 6.2, for each $t = 1, \ldots, n$ we have the inequality

$$1 - \mathcal{P}_N = \Pr\{T \text{ is uniquely defined after } N \text{ random tests}\} \geq$$

$$\Pr\{T \in \mathcal{P}_n^{\leq t}\}\ \Pr\{T \text{ is uniquely defined after } N \text{ random tests} \,|\, T \in \mathcal{P}_n^{\leq t}\}\,.$$

Let us choose $t = \tau_\delta$ and apply the minimal length θ-separating design in $\mathcal{T} \in \mathcal{P}_n^{\leq t}$. We then obtain using (92) and (3) that the right-hand side of the last inequality is larger or equal than $(1 - \delta)(1 - \theta)$. Therefore, recalling the definition of θ, we obtain $1 - \mathcal{P}_N \geq (1 - \delta)(1 - \theta) = 1 - \gamma$. This implies the inequality (93). □

Note that in the asymptotic versions of the upper bounds, the value τ_δ typically comes as a multiplier in the constant C of the main term $C \log n$, while the value of the difference $\gamma - \delta$ does not affect the value of C; this effect is similar to that observed in Section 5, see, for example, (59) and (72). Therefore, the value of δ should be chosen so that $\gamma - \delta$ is much smaller than δ (at least, when n is large).

The inequality (93) does not seem as sharp as (90) but it is very convenient for deriving the asymptotic upper bounds: indeed, it estimates the upper bounds for a binomial sample group testing problem through the upper bounds for the related group testing problem with $\mathcal{T} = \mathcal{P}_n^{\leq t}$. The asymptotic versions of the upper bounds for $\mathcal{T} = \mathcal{P}_n^{\leq t}$ have been derived in Section 5 and can thus be applied to obtain upper bounds in the binomial sample case.

As an example, consider the most important case of the binary model, when the test function is defined by (34). In this case Theorem 5.5 and the inequality (93) imply that for any $0 < \gamma < 1$ there exists a non-sequential γ-separating design with length

$$N \leq \tau_\delta \log_2 n + O(1) \quad \text{as} \quad n \to \infty\,,$$

where (if the normal approximation is used) $\tau_\delta = nq + t_\delta \sqrt{nq(1-q)} + O(1)$ as $n \to \infty$, t_δ is the $(1-\delta)$-quantile of the standard normal distribution and δ is any number such that $0 < \delta < \gamma$.

Acknowledgments

The author has started his study of the probabilistic methods in discrete search in early eighties due to the influence of Prof. M.B.Malutov (North-Western University, Boston); the study and the influence still continue. The paper can be considered as an extension of a joint work of the author with his former student L.Zabalkanskaya (St.Petersburg), who has also participated at an early stage of the present work. The present work is a part of a broader research programme on probabilistic methods in search which the author is conducting in collaboration with Prof. H.P. Wynn (Warwick). The author is very grateful to these three colleagues.

The author is also extremely grateful to the referee for the very careful reading of the manuscript; this resulted in finding a number of important mistakes and misprints, which have been corrected in the revised version.

References

[1] Ahlswede R. and Wegener I. (1987). *Search Problems,* Wiley and Sons, N.Y.

[2] Alon, N., Spencer J. and Erdős P. (1992). *The Probabilistic Method*, Wiley and Sons, N.Y.

[3] DeBonis, A., Gargano, L., and Vaccaro, U. (1997) Group testing with unreliable tests *Information Sciences,* 96, 1–14.

[4] Bruno, W.J, Sun, F, Torney, DC (1988). Optimizing nonadaptive group tests for objects with heterogeneous priors, *SIAM J. on Applied Mathematics,* 58, 1043–1059.

[5] Dorfman, R. (1943). The detection of defective numbers of large population, *Ann. Math. Statist.* 14, 436–440.

[6] Du, D.Z. and Hwang, F.K. (2000) *Combinatorial Group Testing,* 2nd edition, World Scientific, Singapore.

[7] Dyachkov A.G. and Rykov V.V. (1983) A survey of superimposed code theory, *Problems Control Inform. Thy.* 12, 229–242.

[8] Dyachkov A.G., Rykov V.V. and Rashad A.M. (1989) Superimposed distance codes, *Problems Control Inform. Thy.* 18, 237–250.

[9] Erdős, P. and Rényi A. (1963) On two problems of information theory, *Magyar Tud. Akad. Mat. Kutato Int. Kozl.*, 8A, 229–243.

[10] Ghosh, S. and Avila, D. (1985). Some new factor screening designs using the search linear model, *J. Statist. Planning and Inference* 11, 259–266.

[11] Hill, R. and Karim J.P. (1992) Searching with lies: the Ulam problem, Discrete Mathematics, 106/107, 273–283.

[12] Hill, R. (1995) Searching with lies, *Surveys in Combinatorics*, London Math. Society Lecture Notes Series, 218, 41–70.

[13] Katona, G.O.H. (1979) On separating systems whose elements are sets of at most k elements, *Discrete Math.*, 28, 219–222.

[14] Katona, G. and Srivastava J.N. (1983). Minimal 2-coverings of a finite affine space of GF(2), *J. Statist. Planning and Inference* 8, 375–388.

[15] Lindstrem, B. (1975) Determining subsets by unramified experiments, *A Survey of Statistical Design and Linear Models*, (ed. Srivastava J.N.), North–Holland, Amsterdam, 407–418.

[16] Macula, A.J. and Reuter G.R. (1998) Simplified searching for two defects *J. Statist. Planning and Inference*, 66, 77–82.

[17] Macula A.J. (1997a) A nonadaptive version of Ulam's problem with one lie *J. of Statistical Planning and Inference*, 61, 175–180.

[18] Macula, A.J. (1997b) Error-correcting nonadaptive group testing with d(e)-disjunct matrices *Discrete Applied Mathematics*, 80, 217–222.

[19] O'Geran, J.H., Wynn, H.P. and Zhigljavsky, A.A. (1991), Search, *Acta Applicandae Mathematicae*, 25, 241–276.

[20] Patel, M.S. (ed.) (1987), Experiments in factor screening, *Commun. Stat. – Th. and Meth.*, 16, No. 10.

[21] Rényi, A. (1965). On theory of random search, *Bull. Amer. Math. Soc.* 71, 809–828.

[22] Riordan J. (1968) *Combinatorial Identities,* Wiley and Sons,N.Y.

[23] Sobel, M. and Groll, P.A. (1959). Group testing to eliminate efficiently all defectives in a binomial sample, *Bell System Techn. J.*, 38, 1179–1252.

[24] Srivastava J.N. (1975). Designs for searching nonnegligible effects, *A Survey of Statistical Design and Linear Models*, (ed. Srivastava J.N.), North–Holland, Amsterdam, 507–519.

[25] Tsybakov A.S., Mikhailov V.A. and Lihanov N.B. (1983) Bounds for packet transmission rate in a random multiple-access system, *Prob. Inform. Transm.*, 19, 61-81.

[26] Zhigljavsky, A. and Zabalkanskaya, L. (1996). Existence theorems for some group testing strategies, *J. of Statistical Planning and Inference*, 55, 151–173.

Appendix: Tables

Binary model

λ	$n = 10$	$n = 20$	$n = 50$	$n = 100$	$n = 200$	$n = 500$
0.05	32	98	127	184	218	261
0.10	32	56	85	105	124	149
0.15	24	44	64	82	96	115
0.20	24	40	59	73	86	102
0.25	25	39	58	71	83	99
0.30	25	41	59	73	86	102
0.35	34	45	66	79	92	110
0.40	34	52	73	89	104	124
0.45	59	62	90	105	122	145
0.50	59	79	107	128	149	176

Table 1: Values of $N^*(n, t, s)$, see (39), for $t = 3$, $s = \lceil \lambda n \rceil$, various n and λ.

λ	$n = 10$	$n = 20$	$n = 50$	$n = 100$	$n = 200$	$n = 500$
0.05	59	161	195	279	326	388
0.10	59	88	126	153	179	212
0.15	36	65	89	112	132	156
0.20	36	55	78	94	110	131
0.25	32	50	70	86	100	119
0.30	32	48	68	82	96	113
0.35	37	50	70	84	97	115
0.40	47	54	76	91	106	125
0.45	59	64	91	105	122	145
0.50	59	79	107	128	149	176

Table 2: Values of $N^*(n, \leq t, s)$, see Corollary 4.3, for $t = 3$, $s = \lceil \lambda n \rceil$, various n and λ.

n	$t=2$	$t=3$	$t=4$	$t=5$	$t=10$
500	44	111	207	326	1249
1 000	49	125	232	367	1425
2 000	53	137	256	407	1600
5 000	60	156	290	461	1832
10 000	65	169	315	502	2008
30 000	72	189	354	566	2286
50 000	76	199	373	596	2415

Table 3: Values of $N_*(n, \leq t)$, see (70), for various t and n.

n	$t=3$			$t=4$		
	N_*	$N_{0.05}^{(\mathrm{as})}$	$N_{0.0001}^{(\mathrm{as})}$	N_*	$N_{0.05}^{(\mathrm{as})}$	$N_{0.0001}^{(\mathrm{as})}$
500	100	33	42	100	42	51
1 000	111	36	45	111	46	55
2 000	112	39	48	112	50	59
5 000	138	43	52	138	55	64
10 000	150	46	55	150	59	68
20 000	162	49	58	162	63	72
30 000	169	50	59	169	65	74

Table 4: Values of $N_*(n, t)$, see (67), and $N_\gamma^{(\mathrm{as})}(n, t)$, see (72), for $\mathcal{T} = \mathcal{P}_n^t$, $t = 3, 4$.

Multiaccess channel

λ	$n=10$	$n=20$	$n=50$	$n=100$	$n=200$	$n=500$
0.05	32	98	116	169	198	236
0.10	32	49	71	86	101	120
0.15	17	34	46	59	69	82
0.20	17	26	38	46	54	64
0.25	12	22	31	39	46	54
0.30	12	20	29	35	41	49
0.35	12	19	27	33	39	46
0.40	12	19	27	33	38	46
0.45	13	19	28	34	40	47
0.50	13	21	30	36	42	50

Table 5: Values of $N^*(n, t, s)$, see (44), for $\mathcal{T} = \mathcal{P}_n^3$, $\mathcal{X} = \mathcal{P}_n^s$ with $s = \lceil \lambda n \rceil$.

λ	$n = 10$	$n = 20$	$n = 50$	$n = 100$	$n = 200$	$n = 500$
0.05	37	126	159	236	281	339
0.10	37	64	99	124	148	178
0.15	19	45	68	90	107	129
0.20	19	38	60	76	91	110
0.25	18	35	56	71	85	103
0.30	18	36	58	72	86	104
0.35	27	40	64	79	94	113
0.40	27	48	73	90	107	129
0.45	58	60	93	109	129	155
0.50	58	82	115	140	165	197

Table 6: Values of $N^*(n, t, s)$ for $t = 5$, $\mathcal{X} = \mathcal{P}_n^s$, $s = \lceil \lambda n \rceil$, various n and λ.

n	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 10$
100	19	33	52	74	224
500	26	46	72	103	331
1 000	29	51	81	116	377
5 000	36	64	101	146	484
10 000	39	69	109	159	530
100 000	49	87	138	201	682
1 000 000	59	105	167	243	835

Table 7: Values of $N^{(\mathrm{as})}(n, t, \lambda)$, see (77), for $\mathcal{X} = \mathcal{P}_n^s$, $\mathcal{T} = \mathcal{P}_n^t$, various t, n and optimum λ.

n	$t = 4$	$t = 5$	$t = 6$	$t = 10$
100	56	91	132	351
500	77	127	188	521
1 000	86	143	211	594
5 000	108	180	267	764
10 000	117	195	291	837
100 000	148	248	370	1080
1 000 000	178	300	450	1322

Table 8: Values of $N^{(\mathrm{as})}(n, \leq t, \lambda)$, see (80), for $\mathcal{X} = \mathcal{P}_n^s$, $\mathcal{T} = \mathcal{P}_n^{\leq t}$, various t, n and optimum λ.