J Comb Optim (2008) 15: 7–16
DOI 10.1007/s10878-007-9081-5

Hypothesis group testing for disjoint pairs

Morgan A. Bishop · Anthony J. Macula · Thomas E. Renz · Vladimir V. Ufimtsev

Published online: 12 July 2007

Abstract Classical group testing (CGT) is a widely applicable biotechnical technique used to identify a small number of distinguished objects from a population when the presence of any one of these distinguished objects among a *group* of others produces an observable result. This paper discusses a variant of CGT called *group testing for disjoint pairs* (GTAP). The difference between the two is that in GTDP, the distinguished items are pairs from, not individual objects in, the population. There are several biological examples of when this abstract model applies. One biological example is DNA hybridization. The presence of pairs of hybridized DNA strands can be detected in a pool of DNA strands. Another situation is the detection of binding interactions between prey and bait proteins. This paper gives a random pooling method, similar in spirit to hypothesis testing, which identifies pairs of objects from a population that collectively have an observable function. This method is simply to apply, achieves good results, is amenable to automation and can be easily modified to compensate for testing errors.

M.A. Bishop is supported by AFOSR FA8750-06-C-0007.

A.J. Macula is supported by NSF-0436298, AFOSR FA8750-06-C-0007.

M.A. Bishop
JEANSEE, 36 Westview Cr., Geneseo, NY, USA
e-mail: morgan.bishop@rl.af.mil

A.J. Macula (✉)
Department of Mathematics, SUNY Geneseo, Geneseo, NY 14454, USA
e-mail: macula@geneseo.edu

T.E. Renz · V.V. Ufimtsev
Air Force Research Lab, IFTC, Rome Research Site, Rome, NY 13441, USA

T.E. Renz
e-mail: thomas.renz@rl.af.mil

V.V. Ufimtsev
e-mail: ufimtsev.v@neu.edu

Keywords Group testing · Complexes · Nonadaptive · Pooling · Random algorithms

1 Introduction

Throughout this paper, all simple lower case roman variables are non-negative integers. Given set S, $|S|$ denotes its cardinality. Let $[n] = \{1, 2, \ldots, n\}$ represent a finite set with n elements. A k-set in $[n]$ is a k element subset and let $\binom{[n]}{k}$ denote the collection of k-sets in $[n]$. Let $\Gamma = \{C_1, \ldots, C_d\}$ be an unknown collection of d *disjoint* 2-sets of $[n]$. We refer to Γ as the set of *positive pairs* and will always consider these 2-sets to be disjoint. In the group testing for disjoint pairs (GTDP) problem, the goal is to identify as many positive pairs as possible by performing certain 0,1 tests on subsets from $[n]$. These tests are called complex group tests. A pool, $P \subseteq [n]$, is said to be positive if and only it completely contains a positive pair. In short, a pool is positive if and only if there is a C_i in Γ with $C_i \subseteq P$. GTDP is a simplified version of group testing for complexes. See (Dyachkov et al. 2002; Macula et al. 2004; Macula and Popyack 2004; Torney 1999). In (Bishop et al. 2007), the closely related problem of *group testing to annihilate pairs* is applied to DNA strands for the identification *at least one strand* in a pair of cross-hybridized oligonucleotides. If the method described herein were to be applied in the aforementioned situation, then *both strands* in a pair of cross-hybridized oligonucleotides would be identified. The decoding methods in (Bishop et al. 2007) and herein were inspired by (Knill et al. 1996) and different from those given in (Dyachkov et al. 2002; Macula et al. 2004; Macula and Popyack 2004; Torney 1999). See (Du and Hwang 2000) for a comprehensive overview of group testing.

A GTDP design is simply a collections of pools and we use the standard incidence matrix representation for N pools on a population $[n]$. See Table 4. Given an $N \times n$ binary matrix $X \equiv (X_{i,j})$, the ith pool, P_i, in this design is given by the ith row of X when P_i is taken to be the set of all columns of X that have a 1 in the ith row. Let $0 \leq p \leq 1$. In this paper X is an instance of a random binary matrix whose entries are i.i.d. Bernoulli trials where each $X_{i,j}$ is 1 with probability p. Thus each pool is a random pool and can be described as an n-sequence of i.i.d. Bernoulli trials $X_i = X_{i,1}, \ldots, X_{i,n}$. Given a random pooling design X, we define the binary random variable Y_i to be 1 if and only if the pool X_i is positive. The vector $Y = (Y_i)$ is called the output vector.

Throughout this paper p is reserved and assumed to be $p = Prob(X_{i,j} = 1)$ *and is called the pooling probability*

Given any $N \times n$ binary matrix M, let M^2 be the $N \times \binom{n}{2}$ binary matrix whose columns are indexed by $\binom{[n]}{2}$ and $(M^2)_{i,\{j_1,j_2\}} = 1$ if and only if both M_{i,j_1} and M_{i,j_2} are both 1. Given a fixed $\Gamma = \{C_1, \ldots, C_d\}$, it is easy to see that:

$Y_i = 1 \Longleftrightarrow$ there is $\{j_1, j_2\} \in \Gamma$ such that $(X^2)_{i,\{j_1,j_2\}} = 1$.

Note that X^2 is an instance of a random matrix with the $Prob((X^2)_{i,\{j_1,j_2\}} = 1) = p^2$, but the entries are not independent Bernoulli trials.

Table 1 Joint distribution of $(X^2)_{i,\{j_1,j_2\}}$ and Y_i for $\{j_1, j_2\}$ of type 0

$\{j_1, j_2\}$ type 0	$Y_i = 0$	$Y_i = 1$
$(X^2)_{i,\{j_1,j_2\}} = 0$	$p^0_{0,0} \equiv (1-p^2)(1-p^2)^d$	$p^0_{0,0} \equiv (1-p^2) - p^0_{0,0}$
$(X^2)_{i,\{j_1,j_2\}} = 1$	$p^0_{1,0} \equiv p^2(1-p^2)^d$	$p^0_{1,1} \equiv p^2 - p^0_{1,0}$

Table 2 Joint distribution of $(X^2)_{i,\{j_1,j_2\}}$ and Y_i for $\{j_1, j_2\}$ of type 1

$\{j_1, j_2\}$ type 1	$Y_i = 0$	$Y_i = 1$
$(X^2)_{i,\{j_1,j_2\}} = 0$	$p^1_{0,0} \equiv (1-p)(p(1-p)+1)(1-p^2)^{d-1}$	$p^1_{0,1} \equiv (1-p^2) - p^1_{0,0}$
$(X^2)_{i,\{j_1,j_2\}} = 1$	$p^1_{1,0} \equiv p^2(1-p)(1-p^2)^{d-1}$	$p^1_{1,1} \equiv p^2 - p^1_{1,0}$

Table 3 Joint distribution of $(X^2)_{i,\{j_1,j_2\}}$ and Y_i for $\{j_1, j_2\}$ of type 2

$\{j_1, j_2\}$ type 2	$Y_i = 0$	$Y_i = 1$
$(X^2)_{i,\{j_1,j_2\}} = 0$	$p^2_{0,0} \equiv (2p+1)(1-p)^2(1-p^2)^{d-2}$	$p^2_{0,1} \equiv (1-p^2) - p^2_{0,0}$
$(X^2)_{i,\{j_1,j_2\}} = 1$	$p^2_{1,0} \equiv p^2(1-p)^2(1-p^2)^{d-2}$	$p^2_{1,1} \equiv p^2 - p^2_{1,0}$

Table 4 Joint distribution of $(X^2)_{i,\{j_1,j_2\}}$ and Y_i for $\{j_1, j_2\}$ of type 3

$\{j_1, j_2\}$ type 3	$Y_i = 0$	$Y_i = 1$
$(X^2)_{i,\{j_1,j_2\}} = 0$	$p^3_{0,0} \equiv (1-p^2)^d$	$p^3_{0,1} \equiv (1-p^2) - p^3_{0,0}$
$(X^2)_{i,\{j_1,j_2\}} = 1$	$p^3_{1,0} \equiv 0$	$p^3_{1,1} \equiv p^2$

Definition 1 Given $\Gamma = \{C_1, \ldots, C_d\}$, let $H = \bigcup_{i=1}^{d} C_i$. Let $\{j_1, j_2\}$ be a 2-set in $[n]$ that is not in Γ. Then we say that $\{j_1, j_2\}$ is of type t, where $t = 0, 1, 2$, if $|\{j_1, j_2\} \cap H| = t$. If $\{j_1, j_2\}$ is in Γ, then we say it is of type 3. In other words, a positive pair is of type 3.

Suppose that Γ is fixed. Let the experiment be the construction of a random pooling design followed by performance of the group tests with the results recorded. For this experiment, we give the following joint distributions for $(X^2)_{i,\{j_1,j_2\}}$ and Y_i. These joint distributions are different depending upon the type of $\{j_1, j_2\}$ and it is this difference that is used to identify the positive pairs. Note that for $a, b = 0, 1$ and $t = 0, 1, 2$, $p^t_{a,b} = \text{Prob}((X^2)_{i,\{j_1,j_2\}} = a \text{ and } Y_i = b)$ when $\{j_1, j_2\}$ is type t.

2 Interpreting output

We use a combination of the random matrix X and its associated output vector Y to define a new $N \times \binom{n}{2}$ matrix $(Z_{i,\{j_1,j_2\}})$. For each $(i, \{j_1, j_2\}) \in N \times \binom{[n]}{2}$, define the random variable $Z_{i,\{j_1,j_2\}} = X_{i,j_1} \cdot X_{i,j_2} + Y_i + 1 \pmod 2$. Thus for each

$(j_1, j_2) \in \binom{[n]}{2}$, $Z_{\{j_1,j_2\}} = Z_{1,\{j_1,j_2\}}, \ldots, Z_{N,\{j_1,j_2\}}$ is a sequence of i.i.d. Bernoulli variables which are 1 exactly when $X^2_{i,\{j_1,j_2\}} = Y_i$. For $t = 0, 1, 2, 3$, let Z_t denote the Bernoulli variable $Z_{i,\{j_1,j_2\}}$ for $\{j_1, j_2\}$ of type t. From the joint distribution tables above we have that $\text{Prob}(Z_t = 1) = p^t_{0,0} + p^t_{1,1} \equiv \rho_t$. Note that ρ_t is a function of the pooling probability p. The following Proposition is straightforward to verify.

Proposition 1 *Suppose $d \geq 2$ is fixed. For $p \in (0, 1)$, $\rho_0 < \rho_1 < \rho_2 < \rho_3$.*

For each $\{j_1, j_2\} \in \binom{[n]}{2}$, we define $F_{N,\{j_1,j_2\}} = \sum_{i=1}^{N} Z_{i,\{j_1,j_2\}}$. Since $Z_{i,\{j_1,j_2\}}$ is Bernoulli, $F_{N,\{j_1,j_2\}}$ has the binomial distribution, and, if $\{j_1, j_2\}$ is of type t, then $F_{N,\{j_1,j_2\}} = b(N, \rho_t)$. Since $\rho_0 < \rho_1 < \rho_2 < \rho_3$, then, for a given degree of certainty, if N is large enough, then the unknown type of a $\{j_1, j_2\}$ can be determined from observations. Let $B(N, \rho, k) \equiv \sum_{i=k}^{N} \binom{N}{i} \rho^i (1-\rho)^{N-i}$.

Given an instance of a random pooling design and the associated output vector, we have an instance of the random matrix Z. Let the weight column $\{j_1, j_2\}$ be denoted by $\omega_{\{j_1,j_2\}}$. Let $\rho_2 < \tau < \rho_3$ be called the *acceptance threshold*. τ is in this range because we want to identify at least 50% of the positive pairs and we don't want to misidentify more than 50% of type 0, 1, or 2 pairs. Let $\{j_1, j_2\}$ be of type t. Then the probability that $\omega_{\{j_1,j_2\}} \geq \lceil N\tau \rceil$ is $B(N, \rho_t, \lceil N\tau \rceil)$. Thus a hypothesis testing approach can be applied to the identification of Γ.

Suppose $|\Gamma| = d$ and that a proportion $0.5 < \lambda_3 < 1$ of Γ is to be identified on average. Further suppose that the accepted proportion of non-positive pairs that can be misidentified is β. This performance can be achieved if an N, p and τ are found for which the following equations (1–4) simultaneously hold:

$$B(N, \rho_3, \lceil N\tau \rceil) \geq \lambda_3, \tag{1}$$

$$B(N, \rho_0, \lceil N\tau \rceil) \leq \lambda_0, \tag{2}$$

$$B(N, \rho_1, \lceil N\tau \rceil) \leq \lambda_1, \tag{3}$$

$$B(N, \rho_2, \lceil N\tau \rceil) \leq \lambda_2 \tag{4}$$

where

$$\beta \cdot \left(\binom{n}{2} - d \right) = \lambda_0 \cdot T_0 + \lambda_2 \cdot T_1 + \lambda_2 \cdot T_2 \tag{5}$$

and T_t is the number of pairs of type t in $[n]$. Note that $T_0 = \binom{n-2d}{2}$, $T_2 = \binom{2d}{2} - d$, and $T_1 = \binom{n}{2} - \binom{n-2d}{2} - \binom{2d}{2}$.

Using the normal approximation to the binomial, we can think of the bounds on the probabilities λ_t in terms of z-scores. Assume $b(N, \rho_t)$, the distribution of the weight of a column of type t in Z, is normal with $\mu_t = N\rho_t$ and $\sigma_t = \sqrt{N\rho_t(1-\rho_t)}$. Let z_{λ_t} be the z-score so that $Prob(N(0, 1) \geq z_{\lambda_t}) = \lambda_t$. Since we assume that $0 < \lambda_t < 0.5 < \lambda_3 < 1$, we have $z_{\lambda_3} < 0 < z_{\lambda_t}$ for $t = 0, 1, 2$. Let $z_{\{j_1,j_2\}}$ be the z-score of $\omega_{\{j_1,j_2\}}$ where $\{j_1, j_2\}$ is of type t. If $t = 0, 1, 2$, then $B(N, \rho_t, \lceil N\tau \rceil) \leq \lambda_t$ when

$$z_{\{j_1,j_2\}} \equiv \frac{N\tau - N\rho_t}{\sqrt{N\rho_t(1-\rho_t)}} \geq z_{\lambda_t} \tag{6}$$

and if $t = 3$, then $B(N, \rho_3, \lceil N\tau \rceil) \geq \lambda_3$ when

$$z_{\{j_1,j_2\}} \equiv \frac{N\tau - N\rho_t}{\sqrt{N\rho_t(1-\rho_t)}} \leq z_{\lambda_3}. \tag{7}$$

Solving for N and noting that both z_{λ_3} and $\tau - \rho_3$ are less than zero while z_{λ_t} and $\tau - \rho_t$ are greater than zero for $t = 0, 1, 2$ we have for $t = 0, 1, 2$

$$\sqrt{N} \geq \frac{z_{\lambda_t}\sqrt{\rho_t(1-\rho_t)}}{(\tau - \rho_t)}, \tag{8}$$

while for $t = 3$, we have

$$\sqrt{N} \leq \frac{z_{\lambda_3}\sqrt{\rho_3(1-\rho_3)}}{(\tau - \rho_3)}. \tag{9}$$

Setting (8) equal to (9), we can solve for the acceptance threshold τ as a function of pooling probability p and find that

$$\tau_t(p) = \frac{\rho_3 z_{\lambda_3}\sqrt{\rho_3(1-\rho_3)} - \rho_t z_{\lambda_t}\sqrt{\rho_t(1-\rho_t)}}{z_{\lambda_3}\sqrt{\rho_3(1-\rho_3)} - z_{\lambda_t}\sqrt{\rho_t(1-\rho_t)}}. \tag{10}$$

Substituting (10) back into the right side of (9), we have three equations ($t = 0, 1, 2$) that express the number of pools N as a function of the pooling probability

$$\sqrt{N_t(p)} \equiv \frac{z_{\lambda_t}\sqrt{\rho_t(1-\rho_t)}}{(\tau_t(p) - \rho_t)}. \tag{11}$$

The equations in (11) give the pairs $(p, \sqrt{N_t(p)})$ that, for each $t = 0, 1, 2$, make the following pair of equations (12) and (13) simultaneously true:

$$\frac{N_t(p)\tau_t(p) - N_t(p)\rho_t}{\sqrt{N_t(p)\rho_t(1-\rho_t)}} = z_{\lambda_t}, \tag{12}$$

$$\frac{N_t(p)\tau_t(p) - N_t(p)\rho_3}{\sqrt{N_t(p)\rho_3(1-\rho_3)}} = z_{\lambda_3}. \tag{13}$$

In other words, for $t = 0, 1, 2$, the values $N_t(p)$, $\tau_t(p)$, p are those for which

$$B(N_t(p), \rho_3(p), \lceil N_t(p)\tau_t(p) \rceil) = \lambda_3, \tag{14}$$

$$B(N_t(p), \rho_t(p), \lceil N_t(p)\tau_t(p) \rceil) = \lambda_t. \tag{15}$$

It is also clear that if $N \geq N_t(p)$, then

$$B(N, \rho_3(p), \lceil N\tau_t(p) \rceil) \geq \lambda_3, \tag{16}$$

$$B(N, \rho_t(p), \lceil N\tau_t(p) \rceil) \leq \lambda_t. \tag{17}$$

Hence if (11) is minimized over p, then $N_t(p)$ is the smallest number of tests that can achieve the probability requirements in (16) and (17).

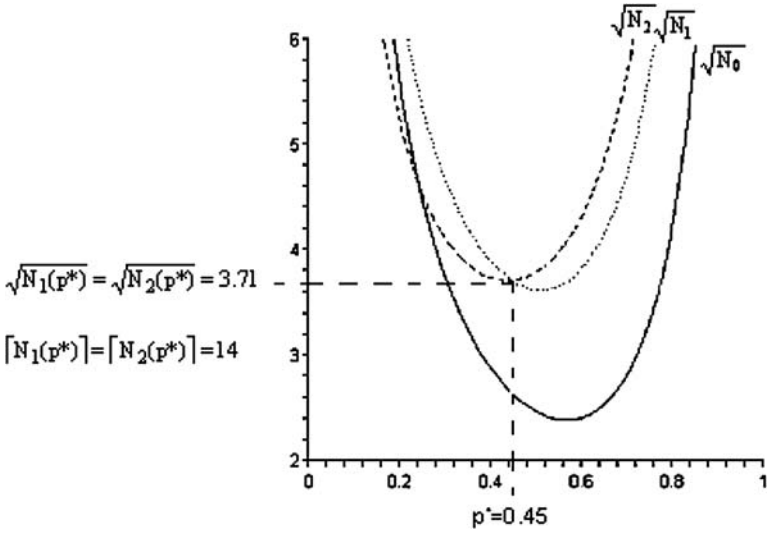

Fig. 1 The curves $\sqrt{N_t}$ for Example 1

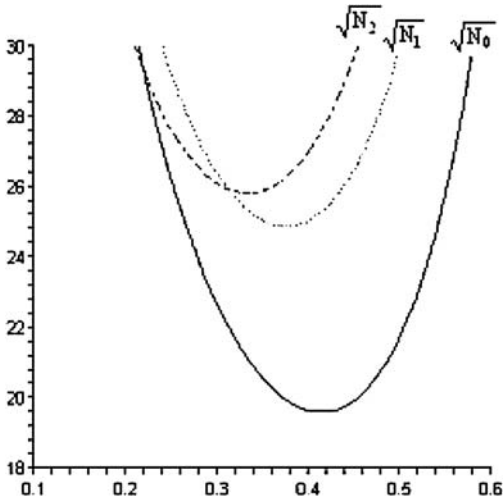

Fig. 2 The curves $\sqrt{N_t}$ for Example 2

Note that for $z_{\lambda_3} < 0 < z_{\lambda_t}$ for $t = 0, 1, 2$, the set of points $\{(N, p) : N \geq N_t(p), 0 < \tau_t(p) < \rho_t(p)\}$ is convex. See Figs. 1 and 2. Thus, by checking all points $(p, \sqrt{N_t(p)})$ that are either absolute minimums of $\sqrt{N_t(p)}$ or are a point of intersection of two $\sqrt{N_t(p)}s$, we can choose the smallest value $\lceil N_{t^*}(p^*) \rceil$ such that (18–21) below are true:

$$B(N_{t^*}(p^*), \rho_3(p^*), \lceil N_{t^*}(p^*)\tau_{t*}(p^*)\rceil) \geq \lambda_3, \tag{18}$$

$$B(N_{t^*}(p^*), \rho_0(p^*), \lceil N_{t^*}(p^*)\tau_{t*}(p^*)\rceil) \leq \lambda_0, \tag{19}$$

Table 5 Random pooling design X

	1	2	3	4	5	6	7	Y
pool 1	0	1	0	0	1	0	0	0
pool 2	1	0	0	0	0	1	1	0
pool 3	1	0	0	1	1	1	1	1
pool 4	1	0	1	1	0	0	0	0
pool 5	1	1	0	1	0	1	0	1
pool 6	0	0	0	1	1	0	1	1
pool 7	0	1	1	0	1	0	0	0
pool 8	0	0	1	1	0	1	0	0
pool 9	0	0	1	1	0	1	1	0
pool 10	0	0	0	0	0	0	1	0
pool 11	0	1	0	0	1	1	0	0
pool 12	1	0	1	1	0	0	1	0
pool 13	1	1	0	1	1	1	0	1
pool 14	0	1	0	1	1	1	1	1

$$B(N_{t^*}(p^*), \rho_1(p^*), \lceil N_{t^*}(p^*)\tau_{t*}(p^*)\rceil) \le \lambda_1, \tag{20}$$

$$B(N_{t^*}(p^*), \rho_2(p^*), \lceil N_{t^*}(p^*)\tau_{t*}(p^*)\rceil) \le \lambda_2. \tag{21}$$

Example 1 In Tables 5–7, the random pooling and decoding method is exhibited for a small population $n = 7$ and the number of positive pairs $d = 2$ with $\Gamma = \{\{1, 2\}, \{4, 5\}\}$. Thus the number of pair types is $T_0 = 3$, $T_1 = 4$, and $T_2 = 12$. With $\lambda_0 = 0.16$, $\lambda_1 = 0.16$, $\lambda_2 = 0.27$ and $\lambda_3 = 0.69$ the optimization method described above and depicted in Fig. 1, the optimized pooling probability is $p^* = 0.44$, the number of required pools is $\lceil N_{t^*}(p^*)\rceil = 14$ and the acceptance threshold is $\tau_{t*}(p^*) = 0.80$. Theoretically, by using the normal approximation, the expected number of misidentified pairs given by (5) should be 4.36 and the expected of positive pairs identified is 1.38. From Table 7, the pairs $\{2, 4\}$, $\{3, 7\}$, and $\{4, 5\}$ attained the threshold. The first two pairs where incorrectly labeled and the last was correctly labeled a positive pair.

Example 2 Let the population be $n = 150$ and suppose there are $d = 5$ positive pairs. Then $T_0 = 9730$, $T_1 = 1405$ and $T_2 = 40$. Let $\lambda_0 = 0.000032$, $\lambda_1 = 0.000233$, $\lambda_2 = 0.008198$ and $\lambda_3 = 0.903200$. Then $z_0 = 4$, $z_1 = 3.5$, $z_2 = 2.4$ and $z_3 = -1.3$. Then from Fig. 2, the optimized pooling probability is $p^* = 0.38$, the number of required pools is $\lceil N_{t^*}(p^*)\rceil = 652$ and the acceptance threshold $\tau_{t*}(p^*) = 0.65$. Since $\lambda_3 = 0.90$ an expected number of 4.5 of the positive pairs will be above the threshold. The expected number of non-positives above the threshold is $9730\lambda_0 + 1405\lambda_1 + 40\lambda_2 = 0.97$.

3 Conclusion

The method described above is simple to apply, achieves good results, is amenable to automation and can be easily modified to compensate for testing errors. Furthermore

Table 6 The matrix X^2

	{(1, 2)}	{(1, 3)}	{(1, 4)}	{(1, 5)}	{(1, 6)}	{(1, 7)}	{(2, 3)}	{(2, 4)}	{(2, 5)}	{(2, 6)}	{(2, 7)}	{(3, 4)}	{(3, 5)}	{(3, 6)}	{(3, 7)}	{(4, 5)}	{(4, 6)}	{(4, 7)}	{(5, 6)}	{(5, 7)}	{(6, 7)}	Y
pairs in pool 1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
pairs in pool 2	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
pairs in pool 3	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
pairs in pool 4	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
pairs in pool 5	1	0	1	0	1	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0	1
pairs in pool 6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1
pairs in pool 7	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
pairs in pool 8	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0
pairs in pool 9	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	1	0	0	1	0
pairs in pool 10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
pairs in pool 11	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0
pairs in pool 12	0	1	1	0	0	1	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0
pairs in pool 13	1	0	1	1	1	0	0	1	1	1	0	0	0	0	0	1	1	0	1	0	0	1
pairs in pool 14	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1

This was constructed relative to the X given in Table 5. A 1 in row i and column $\{a, b\}$ indicates that $\{a, b\}$ is contained in pool i given in Table 5

Table 7 The matrix Z

	{(1, 2)}	{(1, 3)}	{(1, 4)}	{(1, 5)}	{(1, 6)}	{(1, 7)}	{(2, 3)}	{(2, 4)}	{(2, 5)}	{(2, 6)}	{(2, 7)}	{(3, 4)}	{(3, 5)}	{(3, 6)}	{(3, 7)}	{(4, 5)}	{(4, 6)}	{(4, 7)}	{(5, 6)}	{(5, 7)}	{(6, 7)}
	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
	0	0	1	1	1	1	0	0	0	0	0	1	0	0	0	1	1	1	1	1	1
	1	0	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1
	1	0	1	0	1	0	0	1	0	1	0	1	0	0	0	0	1	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	1	0
	1	1	1	1	1	1	0	1	0	1	1	1	0	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	0	1	1	1	1
	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	0	0	1	1	0
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	0	1	1
	1	0	0	1	1	0	1	1	1	1	1	0	1	1	0	1	1	0	1	1	1
	1	0	1	1	1	0	0	1	1	1	0	1	0	0	0	1	1	0	1	0	0
	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1
frequency	0.79	0.50	0.71	0.79	0.79	0.57	0.57	0.86	0.57	0.79	0.71	0.71	0.57	0.50	0.50	0.93	0.79	0.71	0.79	0.86	0.64
threshold attained								yes								yes				yes	
correct positive pair								no								yes				no	

This was constructed relative to the X^2 given in Table 6. The relative frequency of a 1 in column $\{a, b\}$ is recorded in the third row from the bottom. The threshold here is 0.80. Columns above the threshold are indicated in the second row from the bottom

is should be noted that the method is also *nonadaptive*. In nonadaptive group testing (NGT), the exact determination of all the pools that are to be tested must done before any testing occurs. In NGT, pools can often be run in parallel. Adaptive algorithms are multi-stage algorithms and group testing algorithms can be classified by the degree to which they are adaptive. In a two-stage algorithm, an initial battery of tests is carried out (in parallel if possible). Then using the information from the first stage, the second and final battery of tests are constituted and carried out. Because bioinformatic applications are often automated, nonadaptive or few stage methods are generally preferred over more adaptive methods. It seems likely that adaptive modifications of the NGT method described above will improve the pooling efficiency.

References

Bishop M et al (2007) Group testing to annihilate pairs applied to DNA cross-hybridization elimination using SYBR green. J Comput Biol 14:84–96

Du D-Z, Hwang F (2000) Combinatorial group testing and its applications, 2nd edn. World Scientific, Singapore

Dyachkov A et al (2002) On families of subsets where the intersection of l subsets are not covered by the union of s others. J Comb Theory Ser A 99(2):195–218

Knill E, Schliep A, Torney DC (1996) Interpretation of pooling experiments using the Markov chain Monte Carlo method. J Comput Biol 3(3):395–406

Macula A, Popyack L (2004) A group testing method for finding patterns in data. Discret Appl Math 144:149–157

Macula A, Rykov V, Yekhanin S (2004) Trivial two-stage group testing for complexes using almost disjunct matrices. Discret Appl Math 137:97–107

Torney D (1999) Sets pooling designs. Ann Comb 3:95–101