Review Macula's Theorem

$$\begin{array}{cc} & \binom{n}{k} \\ \binom{n}{d} & \left[\begin{array}{c} \\ \\ M \text{ is } d\text{-disjunct} \\ \\ \\ \end{array}\right] \end{array} \qquad k \geq d$$

Fact 1 If $d=1$, then M is 1-disjunct.

Fact 2 Fix $d=1$ and $k>1$. How can we change the set of columns to increase the disjunctness of M. Clearly we need to chop off some columns. (Combinatorial Designs)

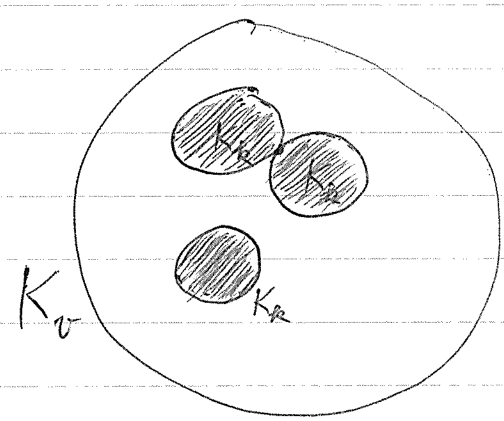
Fact 3 If we let $x = \max\{|C_i \cap C_j| \mid C_i, C_j \text{ are two columns of } M\}$, and $|C_i| = k$ for each column C_i, then the disjunctness is (going to be) at least $\lceil \frac{k}{x} \rceil - 1$.

Proof. One column can cover the other column at most x elements. So, you need $\lceil \frac{k}{x} \rceil$ columns to cover another column. This implies that such at least a matrix M is $(\lceil \frac{k}{x} \rceil - 1)$-disjunct. ∎

Corollary If we can find a collection $\mathbb{B}$ of k-subsets in $[1, v]$ such that two subsets have at most one element in common, then we have an $M_{v \times |\mathbb{B}|}$ which is $(k-1)$-disjunct.

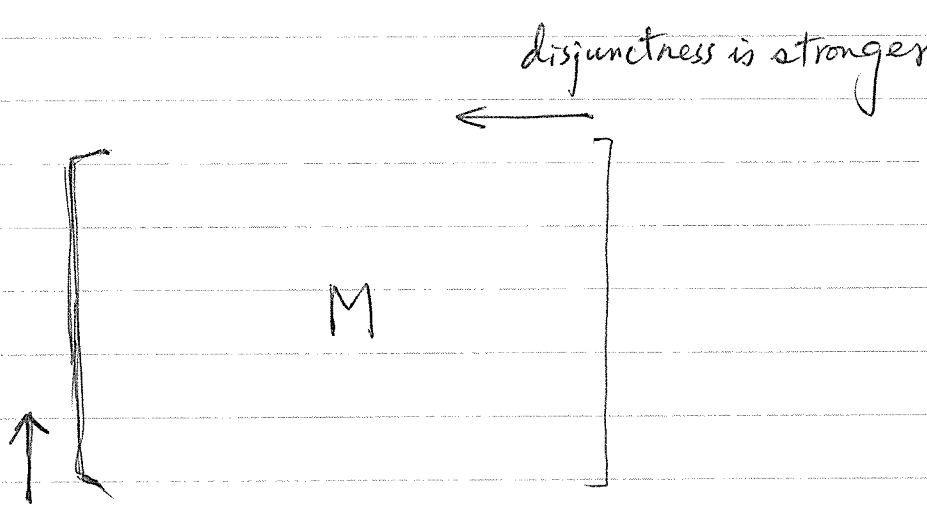
Fact 5 We can find at most $\left\lfloor \binom{v}{k} \Big/ \binom{k}{2} \right\rfloor$ such k-subsets.

Proof.



A k-subset of $[1, v]$ induces a k-clique and any two cliques contain at most one vertex in common.

Concluding remark



disjunctness is weaker

Row: the number of tests
Column: the number of items

In a general setting, we can consider the following modification. Let $\underset{t\times n}{M}$ be a pooling design with fixed number of tests. Foreach column C_i, define $\varphi(C_i)$ be the number of minimum columns which can be used to cover C_i. So, if M is d-disjunct, then $\varphi(C_i) \geq d+1$. Now, let

and $\varphi(C_i) = n$ if the union of $n-1$ other columns can not cover C_i.

$$\varphi(\underset{t\times n}{M}) = \Big(\sum_{i=1}^{n} \varphi(C_i)\Big)\Big/ n.$$

Again, if M is d-disjunct, then $\varphi(M) \geq d+1$. But, the converse statement may not be true. But, the following conclusion is indeed quite convicing: If $\varphi(M) \geq d+1$, then M is almost very sure d-disjunct.

(•) Question: If $\varphi(M) = 3.5$, then what can we say about the matrix M. Is M 2-disjunct? A quick conclusion is of course M is not 3-disjunct.

(••) If we use M is to find three defective items, then what is the probability that the non-adaptive algorithm using M

$$M: \begin{array}{cccccccccccc} C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 & C_8 & C_9 & C_{10} & C_{11} & C_{12} \end{array}$$

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

(•) Each column can be covered by three other columns but not two columns. $\sum_{i=1}^{12} d(C_i) / 12 = 3$.

(•) M is 2-disjunct.

(•) Assume we have three defective items 3, 5, 7. Then, apply M, we obtain $\vec{y}^t = (1,1,0,0,1,0,1,1,1)$.

(*) By using the 0-pool tests, we conclude that $\{C_1, C_6, C_9, C_{12}\} \cup \{C_2, C_4, C_9, C_{11}\} \cup \{C_2, C_6, C_8, C_{10}\}$ is the set of negatives (good items) which is $\{C_1, C_2, C_4, C_6, C_8, C_9, C_{10}, C_{11}, C_{12}\}$. This is exactly the set of negatives and the others are positives.

(**) But, if the set of defective items is $\{1,2,3\}$, then the

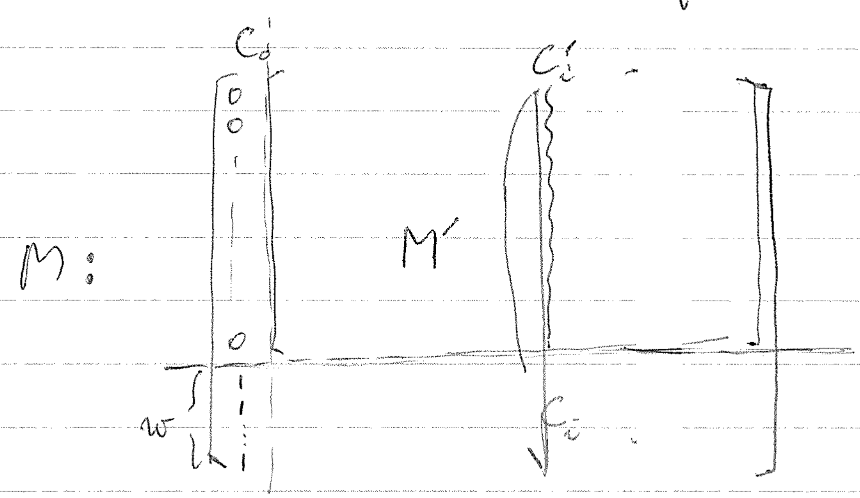
<u>Problem</u> What is the successful probability to find $d+1$ positive items from a set of n items by using a d-disjunct matrix? Here, we consider $\frac{n}{3} > d > 1$.

<u>Definition</u>

Let $t(d,n)$ denote the minimum number of rows (tests) for a d-disjunct matrix with n columns (items).

<u>Fact 6.</u> $t(d,n) \geq w + t(d-1, n-1)$ where $w = \max_{i=1}^{n} |C_i|$, C_i's are the columns.

<u>Proof.</u> Assume that the matrix with $t(d,n)$ rows is as follows:



Here, let $|C_0| = w$. Then M' is $(d-1)$-disjunct. Suppose not. In M', there exists one column $C_d' \subseteq C_1' \cup C_2' \cup \cdots \cup C_{d-1}'$. This implies

Hence, M' is a $(d-1)$-disjunct matrix with $n-1$ columns.

$t(d,n) - w \geq \underline{t(d-1,n-1)}$, since $t(d-1,n-1)$ is the minimum number of tests we need to find $d-1$ positives in $n-1$ items. ∎

<u>Remark</u> A good d-disjunct $t \times n$ matrix M does not contain a column which is too heavy (with larger weight) since it is easier to contain the union of the other columns.

<u>Problem</u> In a d-disjunct matrix $M_{t \times n}$, what is the distribution of column weights if $t \approx t(d,n)$?

<u>Fact 7.</u> $t(d,n) \geq \min\left\{ \binom{d+2}{2}, n \right\}$.

在 adaptive algorithm，最多问 $n-1$ 次；但是在 non-adaptive 最多为 n 次。

<u>Proof.</u> By induction on n, and $n=1$ is true.

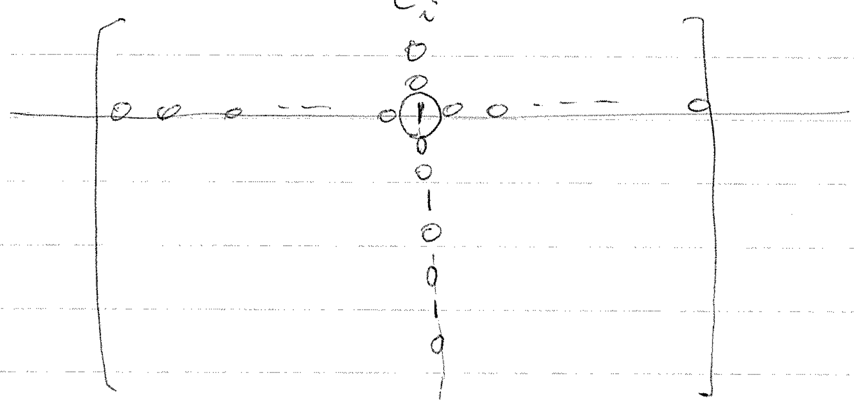
Let M be a d-disjunct $t(d,n) \times n$ matrix. First, we consider that M contains a column of weight $w \geq d+1$. By Fact 6,

$$t(d,n) \geq d+1 + \underline{t(d-1,n-1)} \geq d+1 + \underbrace{\min\left\{ \binom{d+1}{2}, n-1 \right\}}_{\text{(by induction)}}$$

$$= \min\left\{ d+1+\binom{d+1}{2}, d+1+n-1 \right\} \geq \min\left\{ \binom{d+2}{2}, n \right\}.$$ On the other

each C_i (column) (isolated) contains at least one element which is not an element of all the other columns. For otherwise, C_i will be covered by d other columns, a contradiction.

C_i



Now, since every column has an element which does not occur in other columns, we need at least n tests to find the positives, i.e., $t(d,n) \geq n \geq \min\left\{ \binom{d+2}{2}, n \right\}$. ■

<u>Theorem</u> (Erdős, Frankl and Füredi, Israel J. Math., 1985, 79–89)

Let $c(w)$ be the number of columns with weight w. If $M_{t \times n}$ is a d-disjunct matrix, then $c(w) \leq \binom{t}{w} \Big/ \binom{w-1}{v-1}$ where $v = \left\lceil \frac{w}{d} \right\rceil$.

<u>Proof.</u> (Exercise)

(*) Finding a suitable w is an important job in pooling e.g. $t=9$, $n=12$ and 2-disjunct : 9×12 matrix

$$c(3) = 12 \leq \binom{12}{3} \Big/ \binom{2}{1} \quad \text{where} \quad v = \left\lceil \frac{3}{d} \right\rceil = 2,\ w = 3.$$

(∘) This upper bound is in general "too big". Can you find a better one?

(∘∘) So far, the best : $t(d,n) > d(1+o(1))\log_2 n$. (?)

(*) $\bar{d}$-separable matrices do have "worse" decoding algorithm (in complexity), but comparing to d-disjunct matrices, we need less tests if the number of items is fixed to be n. That is, if $\bar{t}(d,n)$ denotes the number of tests in a $\bar{d}$-separable matrix with n items, then $\bar{t}(d,n) \leq t(d,n)$.

(***) It is easier to extend a $(0,1)$-matrix M (pooling design) to a $[0,q-1]$-matrix M_q if separable matrices are concerned. (See next lecture.)