補充机率方法的资料 GT 16

1. Introduction and preliminaries

⋮

Definition 1.0. Let $A_1, A_2, \cdots, A_m$ be events in an arbitrary probability space. A graph $G=(V,E)$ on the set of vertices $V=\{1,2,\cdots,m\}$ is said to be a dependency graph for the events $A_1, A_2, \cdots, A_m$ if for each i, $1 \le i \le m$, the event A_i is mutually independent of a set of all the other events except for those A_j with $\{i,j\} \in E$.

In the following, e denotes the base of natural logrithms, i.e., $e \approx 2.71828$.

Theorem 1.1. (The Lovász Local Lemma; Symmetric case)

Let $A_1, A_2, \cdots, A_m$ be events in an arbitrary probability space. Suppose that each event A_i is mutually independent of a set of all the other events A_j but at most μ, and that $Pr(A_i) \le p$ for all $1 \le i \le m$. If $e \cdot p \cdot (\mu+1) \le 1$, then $Pr(\bigcap_{i=1}^{m} \overline{A_i}) > 0$.

⋮

2. The main results

For positive integers n and d, let $[n]$ denote the set $\{1, 2, \cdots, n\}$ and $\binom{[n]}{d}$ denote the collection of all d-subsets of $[n]$. Let $t(d,n)$ denote the minimum number of rows for a d-disjunct matrix with n columns. In [17], Yeh proved the following theorem by using Theorem 1.1.

Theorem 2.1. [17] $t(d,n) \le d^{-d} \cdot (d+1)^{d+1} \cdot \left(1 + \ln(d+1) + \ln\left[\binom{n}{d+1} - \binom{n-d-1}{d+1}\right]\right)$.

To generalize Theorem 2.1, we start by giving a more general definition.

Definition 2.2. A $t \times n$ binary matrix M is called $(d,r]$-disjunct if the union of any d columns does not contain the intersection of any other r columns in M.

It is worth of noting that a $(d,r]$-disjunct matrix can identify the up-to-d positives on the complex model with complexes of size at most r, see [2] for a reference.

Let $t(n,d,r]$ denote the minimum number of rows for a $(d,r]$-disjunct matrix with n columns. Then we have the following generalization of Theorem 2.1.

<u>Theorem 2.3.</u> $t(n,d,r] \le d^{-d} \cdot r^{-r} \cdot (d+r)^{d+r} \cdot \left(1+\ln\left[\binom{n}{d}\binom{n-d}{r} - \binom{n-d-r}{d}\binom{n-2d-r}{r}\right]\right)$.

<u>Proof.</u> Let $M=(m_{i,j})$ be a $t\times n$ random binary matrix with $Pr(m_{i,j}=1)=p$ and $Pr(m_{i,j}=0)=1-p$. Note that all $m_{i,j}$'s are chosen independently. Let $A_{D,\bar{R}}$ be the even that $\cap \bar{R} \subseteq \cup D$ where $D\in\binom{[n]}{d}$ and $\bar{R}\in\binom{[n]}{r}$ with $D\cap\bar{R}=\phi$. Then,

$$Pr(A_{D,\bar{R}}) = (1-p^r\cdot(1-p)^d)^t.$$

By Lovász's Local Lemma, a $t\times n$ $(d,r]$-disjunct matrix exists whenever $e\cdot(\mu+1)\cdot Pr(A_{D,\bar{R}}) \le 1$ where $\mu+1 = \binom{n}{d}\cdot\binom{n-d}{r} - \binom{n-(d+r)}{d}\cdot\binom{n-2d-r}{r}$. Here, μ is the maximum degree of dependency graph.

Hence $e\cdot(1-p^r\cdot(1-p)^d)^t\cdot\left[\binom{n}{d}\cdot\binom{n-d}{r} - \binom{n-d-r}{d}\binom{n-2d-r}{r}\right] \le 1$ is required. This implies that

$$t \ge \frac{1+\ln\left[\binom{n}{d}\binom{n-d}{r} - \binom{n-d-r}{d}\binom{n-2d-r}{r}\right]}{-\ln(1-p^r\cdot(1-p)^d)}. \cdots (*)$$

Since $-\ln(1-x) \geq x$ for $0 \leq x < 1$, we conclude that if

$$t \geq \frac{1 + \ln\left[\binom{n}{d}\binom{n-d}{r} - \binom{n-d-r}{d}\cdot\binom{n-2d-r}{r}\right]}{p^r \cdot (1-p)^d} \quad (2)$$

holds, then the inequality (*) holds. By the fact that $p = \frac{r}{d+r}$ gives the minimum value of R.H.S of (2) we conclude the proof by pluging in p. ∎

By a similar technique, we also have the following result. First, we need a definition.

<u>Definition 2.4.</u> A $t \times n$ binary matrix M is called (d,r)-disjunct if the union of any d columns does not contain the union of any other r columns in M.

Note that an (h,d)-disjunct matrix can be applied to identify the positives on the (d,h)-inhibitor model.[6] Here, h is the number of inhibitors.

Let $t(n,d,r)$ denote the minimum number of rows for a (d,r)-disjunct matrix with n columns. The following is also a generalization of Theorem 2.1.

Theorem 2.5. $t(n,d,r) \le \left(1+\frac{d}{r}\right)\left(1+\frac{r}{d}\right)^{\frac{d}{r}} \cdot \left(1+\ln\left[\binom{n}{d}\binom{n-d}{r}-\binom{n-d-r}{d}\binom{n-2d-r}{r}\right]\right)$

Proof. By a similar set up as in Theorem 2.3. Let $A_{D,R}$ be the event that $\cup R \subseteq \cup D$ where $D \in \binom{[n]}{d}$, $R \in \binom{[n]}{r}$ and $D \cap R = \phi$. Then, $Pr(A_{D,R}) = \left(1-(1-p)^{d} \cdot [1-(1-p)^{r}]\right)^{t}$. By using the Lovász Local lemma, we conclude that a $t \times n$ (d,r)-disjunct matrix exists whenever

$e \cdot Pr(A_{D,R}) \cdot \left[\binom{n}{d}\binom{n-d}{r}-\binom{n-d-r}{d}\cdot\binom{n-2d-r}{r}\right] \le 1$. Hence, we conclude the proof by letting $p = 1-\left(\frac{d}{d+r}\right)^{\frac{1}{r}}$. ▨

We can use the following notion to extend the above two types of disjunct matrices.

Definition 2.6. For $1 \le s \le r$, a $t \times n$ binary matrix M is called $(d, s$ out of $r]$-disjunct if for any d columns and any other r columns of M, there exists a row index in which none of the d columns appear and at least s of the r columns do.

Clearly, $(d, 1$ out of $r]$-disjunctness is precisely the (d,r)-disjunctness and $(d, r$ out of $r]$-disjunctness is equivalent to $(d,\bar{r}]$-disjunctness.

Now, with the same technique we can also find a good upper bound for a the number of rows "t" in a $t\times n$ $(d, s$ out of $r]$-disjunct matrix.

<u>Theorem 2.7.</u> Let $t(n,d,r,s]$ denote the minimum number of rows we need in a $(d, s$ out of $r]$-disjunct matrix with n columns. Then, $t(n,d,r,s] \le \left(1+\ln\left[\binom{n}{d}\binom{n-d}{r}-\binom{n-d-r}{d}\binom{n-2d-r}{r}\right]\right)\Big/ f_{d,r,s}(p)$ for all $0<p<1$, where $f_{d,r,s}(p) = (1-p)^d\cdot\left[1-\sum_{i=0}^{s-1}\binom{r}{i}p^i(1-p)^{r-i}\right]$.

<u>Proof.</u> By using the same set up as above, let $\overline{A_{D,R}}$ be the event that there exists a row index in which none of the column $C_j, j\in D$ appear and at least s of the columns $C_k, k\in R$, do. Then,

$$Pr(A_{D,R}) = \left\{1-(1-p)^d\cdot\left[1-\sum_{i=0}^{s-1}\binom{r}{i}p^i\cdot(1-p)^{r-i}\right]\right\}^t.$$

For convenience, let $f_{d,r,s}(p) = (1-p)^d\cdot\left[1-\sum_{i=0}^{s-1}\binom{r}{i}p^i\cdot(1-p)^{r-i}\right]^{\cancel{t}}$.

Again, by Theorem 1.1, a $t\times n$ $(d, s$ out of $r]$-disjunct matrix exists whenever

$$e\cdot[1-f_{d,r,s}(p)]^t\cdot\left[\binom{n}{d}\binom{n-d}{r}-\binom{n-d-r}{d}\binom{n-2d-r}{r}\right]\le 1.$$

Hence, $t \ge 1+\ln\left[\binom{n}{d}\binom{n-d}{r}-\binom{n-d-r}{d}\binom{n-2d-r}{r}\right]\Big/ f_{d,r,s}(p)$ by using the fact $-\ln(1-x)\ge x$ for $0\le x<1$. $\blacksquare$

We can use similar approaches to obtain an upper bound for the minimum size (of rows) of a (k,m,n)-selector. Review that a (k,m,n)-selector with t rows is a $t \times n$ binary matrix M such that any submatrix of M obtained by choosing arbitrary k columns of M contains at least m rows of the identity matrix I_k. The integer t is commonly known as the size of the (k,m,n)-selector. Du and Hwang [9] proved that a (k,m,n)-selector is $(m-1, k-m+1)$-disjunct, but in general the reverse statement is not true. The (k,m,n)-selector does have quite a few applications in group testing, see [5] for example.

Let $t_s(k,m,n)$ denote the minimum size of a (k,m,n)-selector. De Bonis et. al. [5] obtained upper bounds for $t_s(k,m,n)$ by translating the problem into the hypergraph language.

<u>Theorem 2.8.</u> [5] $t_s(k,m,n) < \frac{ek^2}{k-m+1}\left(\ln\frac{n}{k}\right) + \frac{ek(2k-1)}{k-m+1}$.

The following upper bound is obtained by applying Theorem 1.1.

<u>Theorem 2.9.</u> $t_s(k,m,n) \le \frac{m}{\binom{k}{m}\cdot m!}\cdot\left[k\left(1+\frac{1}{k-1}\right)^{k-1}\right]^m\cdot\left[1+\ln\left(\binom{n}{k}-\binom{n-k}{k}\right)\right]$.

<u>Proof.</u> Let $M^* = (m^*_{i,j})$ be a $t\times n$ random binary matrix with $Pr(m^*_{i,j}=1) = p$ and $Pr(m^*_{i,j}=0) = 1-p$. For $K \in \binom{[n]}{k}$ and $M \in \binom{[t]}{m}$, define A_K be the event that the $t\times k$ submatrix of M^* corresponding to K contains at most $m-1$ rows of I_k, and $A_{K,M}$ be the event that the $m\times k$ submatrix of M^* corresponding to K and M doesn't consist m distinct rows of I_k. Observe that
$$A_K = \bigcap_{M\in\binom{[t]}{m}} A_{K,M}.$$

For convenience, assume that $m|t$. Let $M_i = \{m(i-1)+1, m(i-1)+2, \cdots, mi\}$ for $1\le i\le \frac{t}{m}$. Then

$$Pr(A_K) = Pr\Big(\bigcap_{M\in\binom{[t]}{m}} A_{K,M}\Big) \le Pr\Big(\bigcap_{i=1}^{t/m} A_{K,M_i}\Big)$$

$$= \left[1-\binom{k}{m} m!\cdot p^m\cdot (1-p)^{m(k-1)}\right]^{t/m}.$$

Note that A_K is mutually independent of all the other events $A_{K'}$ except for those $K'\cap K\ne\emptyset$. There are exactly $\binom{n}{k}-\binom{n-k}{k}-1$ such events. Now, by Theorem 1.1, a $t\times n$ (k,m,n)-selector exists whenever

$$e \cdot \left[1-\binom{k}{m} m!\, p^{m} \cdot (1-p)^{m(k-1)}\right]^{\frac{t}{m}} \cdot \left[\binom{n}{k}-\binom{n-k}{k}\right] \le 1 .$$

This implies that

$$t \ge m \cdot \frac{1+\ln\left[\binom{n}{k}-\binom{n-k}{k}\right]}{-\ln\left[1-\binom{k}{m} m!\, p^{m}(1-p)^{m(k-1)}\right]} \quad \text{and}$$

thus $$t \ge m \cdot \frac{1+\ln\left[\binom{n}{k}-\binom{n-k}{k}\right]}{\binom{k}{m} m! \cdot p^{m}(1-p)^{m(k-1)}} .$$

Now, by plugging in $p=\frac{1}{k}$, we obtain the desired upper bound for $t_s(k,m,n)$. ∎

References

~~Bibliography~~

[1] Noga Alon, Joel H. Spencer, The Probabilistic Method, 2nd ed., *John Wiley and Sons, Inc.*, (2000).

[2] H. B. Chen, D. Z. Du and F. K. Hwang, An unexpected meeting of four seemingly unrealated problems: graph testing, DNA complex screening, superimposed codes and secure key distribution, *J. Combin. Opt.*, to appear.

[3] H. B. Chen, H. L. Fu and F. K. Hwang, An upper bound of the number of tests in pooling designs for the error-tolerant complex model, Opt. Letters, to appear.

[4] P. Damaschke, Threshold Group Testing, *Electronic Notes in Discrete Mathematics 21*, (2005), 265-271.

[5] A. De Bonis, Leszek Gąsieniec, U. Vaccaro, Optimal Two-Stage Algorithms for Group Testing Problems, *SIAM J. Comput. Vol. 34, No. 5*, (2005), 1253-1270.

[6] A. De Bonis and U. Vaccaro, Improved algorithms for group testing with inhibitors, *Inform. Process Lett. 67*, (1998), 57-64.

[7] D. Deng, D. R. Stinson, R. Wei, The Lovász Local Lemma and Its Applications to some Combinatorial Arrays, *Designs, Codes and Cryptography, 32*, (2004), 121-134.

[8] Ding-Zhu Du, Frank K. Hwang, Combinatorial Group Testing and Its Applications, 2nd ed., *World Scientific*, (2000).

[9] Ding-Zhu Du, Frank K. Hwang, Pooling Designs and Nonadaptive Group Testing - Important Tools for DNA Sequencing, *World Scientific*, (2006).

[10] P. Erdős, L. Lovász, Problems and Results on 3-chromatic Hypergraphs and Some Related Questions, *in: Infinite and Finite Sets*(A. Hajnal et al., eds.), North-Holland, Amsterdam, (1975), 609-628.

[11] M. Farach, S. Kannan, E. Knill, S. Muthukrishnan, Group Testing Problem with Sequences in Experimental Molecular Biology, *Proc. Compression and Complexity of Sequences*, (1997), 357-367.

[12] C. H. Li, A sequential method for screening experimental variables, *J. Amer. Statist. Assoc. 57*, (1962), 455-477.

[13] L. Lovász, On the Ratio of Optimal Integral and Fractional Covers, *Discrete Math., 13*, (1975), 383-390.

[14] F. P. Ramsey, On A Problem of Formal Logic, *Proc. London Math. Soc. 30(2)*, (1929), 264-286.

[15] Joel Spencer, Probabilistic Methods, *Graphs and Combinatorics 1*, (1985), 357-382.

[16] D. R. Stinson and R. Wei, Generalized cover-free families, *Discrete Math. 279*, (2004), 463-477.

[17] Hong-Gwa Yeh, d-Disjunct matrices: bounds and Lovász Local Lemma, *Discrete Mathematics 253*, (2002), 97-107.