Systems of distinct representatives

1 SDRs and Hall's Theorem

Let $(X_1,\ldots,X_n)$ be a family of subsets of a set A, indexed by the first n natural numbers. (We allow some of the sets to be equal.) A *system of distinct representatives*, or SDR, for the family, is a family $(x_1,\ldots,x_n)$ of elements of A satisfying the conditions

- $x_i \in X_i$ for $i = 1,\ldots,n$;
- $x_i \neq x_j$ for $i \neq j$.

(The first condition asserts that the elements are representatives of the sets, and the second that they are distinct.)

Clearly the existence of an SDR imposes conditions on the family of sets: any k sets must between them contain at least k elements (since they must have k distinct representatives). In particular, all the sets must be non-empty. Philip Hall [5] proved that this condition is also sufficient for the existence of an SDR. The result is often called *Hall's Marriage Theorem*, since it is stated in the following form: given n boys, if any k of the boys between them know at least k girls (for $1 \leq k \leq n$), then it is possible to marry each boy to a girl that he knows.

We introduce some notation to state the theorem formally. Given a family $(X_1,\ldots,X_n)$ of sets, for each set $I \subseteq \{1,\ldots,n\}$ we define

$$X(I) = \bigcup_{i \in I} X_i.$$

We say that the family satisfies *Hall's condition* if $|X(I)| \geq |I|$ for any $I \subseteq \{1,\ldots,n\}$.

Theorem 1 *A family of sets has a SDR if and only if it satisfies Hall's condition.*

There are many different proofs of this theorem, so we do not give one here. Note that there is a polynomial-time algorithm which either finds an SDR or shows that one cannot exist by finding a violation of Hall's condition.

2 Permanents

There is a well-known formula for the determinant of the square matrix $A = (a_{ij})$:

$$\det(A) = \sum_{\pi \in S_n} \mathrm{sgn}(\pi) a_{1\pi(1)} a_{2\pi(2)} \cdots a_{n\pi(n)},$$

where S_n is the symmetric group on $\{1, \ldots, n\}$, and sgn is the sign function (taking the value $+1$ on even permutations and -1 on odd permutations). The *permanent* is the function defined by the same formula without the sign factor:

$$\mathrm{per}(A) = \sum_{\pi \in S_n} a_{1\pi(1)} a_{2\pi(2)} \cdots a_{n\pi(n)}.$$

Somewhat surprisingly, while the determinant can be computed in polynomial time, the easier-looking permanent cannot (so far as we know): its calculation is #P-complete.

The connection with SDRs lies in the observation that, if A is the zero-one incidence matrix of a family of n subsets of an n-set, then $\mathrm{per}(A)$ is the number of SDRs of A: each non-zero term in the permanent arises from a permutation whose list of values $(\pi(1), \ldots, \pi(n))$ is a SDR, and every SDR contributes one to the sum.

A matrix A is said to be *doubly stochastic* if its entries are non-negative and all row and column sums are equal to 1. The name comes from the fact that the transition matrix of a Markov chain with finitely many states (the matrix whose (i, j) entry is the probability of a transition from state i to state j) is non-negative and has all row sums equal to 1 – such a matrix is called *stochastic*. A permutation matrix (a zero-one matrix with a single non-zero entry in any row or column) is doubly stochastic, though the corresponding Markov chain is "deterministic".

Theorem 2 *Let A be a doubly stochastic matrix. Then*

(a) $\mathrm{per}(A) \neq 0$*;*

(b) A is a convex combination of permutation matrices (that is, $A = x_1P_1 + \cdots + x_rP_r$*, where* P_i *are permutation matrices and* x_i *positive real numbers with* $x_1 + \cdots + x_r = 1$*).*

Part (a) is a consequence of Hall's Theorem. If X_i is the set of indices of columns having non-zero entries in the ith row, then the entries in the columns indexed by $X(I)$ have sum at least $|I|$, since they contain all the non-zero entries in the rows indexed by I; so there are at least $|I|$ such columns. Hence this family of sets has an SDR, of the form $(\pi(1),\ldots,\pi(n))$ for some permutation π. Then the corresponding term in the permanent is non-zero.

Part (b) is then proved by induction on the number of non-zero elements in the matrix. Given a SDR, we subtract a multiple of the corresponding permutation matrix and rescale to get a doubly stochastic matrix with fewer non-zero entries.

Corollary 3 *Let $(X_1,\ldots,X_n)$ be a family of k-element subsets of $\{1,\ldots,n\}$, and suppose that each element of $\{1,\ldots,n\}$ lies in exactly k of these sets. Then the family has an SDR, and indeed has k disjoint SDRs.*

The first part follows from the fact that, if A is the incidence matrix of the family, then $(1/k)A$ is doubly stochastic. The second part is proved by induction as above. (Here two SDRs are said to be disjoint if the representatives of any set in the two systems are different.)

An application of this result gives the existence of Youden "squares" [6, 4]. A block design is a set Ω of plots, with two partitions B and T of Ω (the block and treatment partitions). It is binary if any treatment and any block intersect in at most one plot. A square BIBD is a binary design satisfying the three conditions

- $|B| = |T| = v$;
- each part of T and each part of B has size k, where $k < v$;
- any two treatments occur together in exactly λ blocks, where $\lambda > 0$.

A Youden "square" supported by such a design is a further partition of Ω into k parts of size v orthogonal to the block and treatment partitions (that is, each part of the Youden partition meets each part of the other two partitions in a single plot).

Corollary 4 *Any square BIBD supports a Youden "square".*

For, if we translate the block design into an incidence structure (with the treatments as points and the blocks regarded as sets of points), then a part of the Youden partition is an SDR for the resulting family of sets, and the whole partition is a set of k pairwise disjoint SDRs, whose existence is guaranteed by Corollary 3. Note that the third condition of the definition of a square BIBD is not used in this argument.

3 The van der Waerden conjecture

One of the most important results about permanents was conjectured by van der Waerden and proved by Egorychev [2] and Falikman [3].

Theorem 5 *Let A be an $n\times n$ doubly stochastic matrix. Then* $\mathrm{per}(A)\geq n!/n^n$, *with equality if and only if $A=(1/n)J$, where J is the all-*1 *matrix.*

This shows, for example, that a family of sets satisfying the hypotheses of Corollary 3 has at least $n!(k/n)^n$ SDRs. For the number of SDRs is the permanent of A, and so is k^n times the permanent of the doubly stochastic matrix $(1/k)A$.

In the case of a symmetric BIBD, we can do better. The incidence matrix satisfies $AA^\top=(k-\lambda)I+\lambda J$, from which we find that

$$\det(A)^2=\det(AA^\top)=k^2(k-\lambda)^{v-1}.$$

Thus

$$\mathrm{per}(A)\geq|\det A|=k(k-\lambda)^{(v-1)/2},$$

the first inequality holding since the determinant contains terms of both signs which are all positive in the permanent.

4 Latin squares

The number $L(n)$ of Latin squares of order n has been studied for a long time. One of the motivations was the suggestion of Yates that, to randomize an experimental design based on a Latin square, one should choose at random from the set of all Latin squares of the appropriate size, and for this one should know how many there are!

Clearly $L(n)\leq n^{n^2}$, since there are at most n choices for the entries in each of the n^2 cells of the square. This upper bound can be further improved to $(n!)^n$, since each row is a permutation of $(1,\ldots,n)$. Further improvements are possible. What about lower bounds?

Let $X_i=\{1,\ldots,n\}$ for $i=1,\ldots,n$. Then each row of a Latin square of order n is an SDR for the family $(X_!,\ldots,X_n)$; and distinct rows correspond to disjoint transversals. So the counting problem is a special case of those considered in the last section, and we can derive lower bounds as follows.

The ith row is an SDR for the family of sets obtained by omitting from X_j the $i-1$ entries already used in the jth column in preceding rows. (This guarantees the disjointness.) The resulting sets satisfy the hypotheses of Corollary 3 with $k = n - i + 1$. So the number of choices for the ith row is at least $n!((n-i+1)/n)^n$. Multiplying these numbers together for $i = 1, \ldots, n$, we obtain

$$L(n) \geq \frac{(n!)^{2n}}{n^{n^2}}.$$

Since $n! \geq (n/\mathrm{e})^n$, this gives $L(n) \geq (n/\mathrm{e}^2)^{n^2}$, which is not too far from the trivial upper bound (their logarithms are asymptotically equal).

Babai [1] exploited this lower bound to show that, asymptotically, almost all Latin squares have trivial automorphism group. The existence of a non-trivial automorphism reduces the upper bound so drastically that, summed over all possible permutations, the total is much smaller than the lower bound for $L(n)$.

References

[1] L. Babai, Almost all Steiner triple systems are asymmetric, in *Topics in Steiner systems* (ed. C. C. Lindner and A. Rosa), *Ann. Discrete Math.* **7**, Elsevier, Amsterdam, 1979, pp. 37-39.

[2] G. P. Egorychev, Solution of the van der Waerden problem for permanents, *Soviet Math., Dokl.* **23** (1981), 619–622; translation from *Dokl. Akad. Nauk SSSR* **258** (1981), 1041–1044.

[3] D. I. Falikman, Proof of the van der Waerden conjecture regarding the permanent of a doubly stochastic matrix, *Math. Notes* **29** (1981), 475–479; translation from *Mat. Zametki* **29** (1981), 931–938.

[4] R. A. Fisher, The mathematics of experimentation, *Nature* **142** (1938), 442–443, reprinted in *Collected Papers of R. A. Fisher* (ed. J. H. Bennett), vol. 4, University of Adelaide, Adelaide, 1974, pp. 155–158.

[5] P. Hall, On representatives of subsets, *J. London Math. Soc.* **10** (1935), 26–30.

[6] W. J. Youden, Use of incomplete block replications in estimating tobacco-mosaic virus, *Contrib. Boyce Thompson Inst.* **9** (1937), 41–48.

Peter J. Cameron

May 30, 2003