RECTANGULAR ARRAYS WITH FIXED MARGINS

PERSI DIACONIS* AND ANIL GANGOLLI†

Abstract. In a variety of combinatorial and statistical applications, one needs to know the number of rectangular arrays of nonnegative integers with given row and column sums. The combinatorial problems include counting magic squares, enumerating permutations by descent patterns and a variety of problems in representation theory. The statistical problems involve goodness of fit tests for contingency tables. We review these problems along with the available techniques for exact and approximate solution.

1. Introduction. Let $\mathbf{r} = (r_1, \cdots, r_m)$ and $\mathbf{c} = (c_1, \cdots, c_n)$ denote positive integer partitions of N. Let $\Sigma_{\mathbf{rc}}$ denote the set of all $m \times n$ nonnegative integer matrices in which row i has sum r_i and column j has sum c_j. Thus $\Sigma_{ij} T_{ij} = N$ for every $T \in \Sigma_{\mathbf{rc}}$. Throughout, we assume $m, n > 1$; otherwise $\Sigma_{\mathbf{rc}}$ has only one element. As will emerge, $\Sigma_{\mathbf{rc}}$ is always nonempty.

When $m = n$ and $r_i \equiv c_j \equiv r$, $\Sigma_{\mathbf{rc}}$ becomes the set of magical squares. The classical literature on these squares is reviewed in Section 2. For general $\mathbf{r}, \mathbf{c}$, the set $\Sigma_{\mathbf{rc}}$ arises in permutation enumeration problems. These include enumerating permutations by descents, enumerating double cosets, and describing tensor product decompositions. Section 3 describes these problems. Sections 4, 5 describe closely related problems in symmetric function theory.

In statistical applications, $\Sigma_{\mathbf{rc}}$ is called the set of contingency tables with given margins $\mathbf{r}$ and $\mathbf{c}$. Tests for independence and more general statistical models are classically quantified by the chi-square distribution. More accurate approximations require knowledge of $\Sigma_{\mathbf{rc}}$. These topics are covered in Section 6.

Remaining sections describe algorithms and theory for enumerating and approximating $|\Sigma_{\mathbf{rc}}|$. Section 7 describes asymptotic approximations. Section 8 describes algorithms for exact enumeration. Section 9 gives complexity results. Section 10 describes Monte Carlo Markov chain techniques. The different sections are independent and may seem to be quite disparate; the link throughout is the set of tables.

Each of these topics has seen active development in recent years. A useful review of the earlier work is in Good and Crook (1977).

2. Magical squares. Let $H_n(r)$ denote the number of $n \times n$ matrices of nonnegative integers each of whose row and column sums is r. Such matrices are called magical squares and were first studied by MacMahon (1916) who showed $H_3(r) = \binom{r+2}{2} + 3\binom{r+3}{4}$. Stanley (1973, 1986) gives a careful review of the history. He has proved that for $r \geq 0$ $H_n(r)$ is a

* Dept. of Mathematics, Harvard University, Cambridge, MA 02138.
† Silicon Graphics, 2011 N. Shoreline Blvd., Mount View, CA 94043-1389.

polynomial in r of degree exactly $(n-1)^2$. This polynomial is known for $n \le 6$: see Jackson and Van Rees (1975).

A polynomial of degree $(n-1)^2$ is determined by its values on $(n-1)^2+1$ points. Stanley shows that the polynomial H_n satisfies

$$H_n(-1) = H_n(-2) = \cdots = H_n(-n+1) = 0;$$
$$H_n(-n-r) = (-1)^{n-1} H_n(r), \quad \text{for all} \quad r.$$

Thus, the values $H_n(i)$ for $1 \le i \le \binom{n-1}{2}$ determine $H_n(r)$ for all r. The techniques developed below should enable computation of H_7, and perhaps H_8.

The leading coefficient of $H_n(r)$ is the volume of the polytope of $n \times n$ doubly stochastic matrices (Stanley (1986)). This does not have a closed form expression at this writing. Perhaps inspection of the data will permit a reasonable guess. Bona (1994) gives bounds on the volume of the doubly stochastic matrices.

The problem of enumerating magical squares with both diagonals summing to r is discussed in Section 8.2 below. The theory of the present section should apply here but little previous work has been done. We further mention recent work of Jia (1994) which uses multivariate spline techniques to give new proofs of Stanley's results as well as resolve some open problems. Finally, we mention work by Gessel (1990) and Goulden, Jackson, Reilly (1983). These authors show that the problems of this section are P-recursive.

3. Examples in the permutation group. Let S_n be the group of permutations of n objects. This section shows how $\Sigma_{\mathbf{rc}}$ arises in enumerating permutations by descents, in describing double cosets, and in decomposing induced representations and tensor products.

3.1. Descents. A permutation has a descent at i if $\pi(i) > \pi(i+1)$. The descent set of π is $D(\pi) = \{i : \pi(i) > \pi(i+1)\}$. Thus 2431 has descents at positions 2 and 3 so $D(2431) = \{2,3\}$. By definition, $D(\pi) \subseteq \{1, 2, \cdots, n-1\}$. There is a useful recoding of descent sets as compositions of $n : D \longleftrightarrow C(D)$. If $D = \{d_1, d_2, \cdots d_r\}$ with $d_1 < d_2 < \cdots < d_r$, set $c_1 = d_1$, $c_2 = d_2 - d_1 \cdots c_{r+1} = n - d_r$. Thus $D = \{2,3\}$ has $C(D) = (2,1,1)$. The map back has $D(C) = \{c_1, c_1+c_2, \cdots, c_1 + \cdots + c_{r-1}\}$. Descents are actively studied in several areas of mathematics. See, e.g., Stanley (1986), Gessel and Reutenauer (1994), Diaconis, McGrath, Pitman (1993) and the literature cited there. The following result is attributed to Foulkes in the folklore of combinatorics. The elegant bijective proof given below is due to Nantel Bergeron (personal communication).

THEOREM 3.1 (FOULKES). *Let* $\mathbf{r}$ *and* $\mathbf{c}$ *be compositions of* N. *The number of permutations* π *in* S_N *with* $D(\pi) \subseteq D(\mathbf{r})$ *and* $D(\pi^{-1}) \subseteq D(\mathbf{c})$ *is* $|\Sigma_{\mathbf{rc}}|$.

Example 3.2. The following display lists $\pi/D(\pi); \pi^{-1}/D(\pi^{-1})$ for π in S_4

1234/ϕ ; 1234/ϕ	2134/1 ; 2134/1	3124/1 ; 2314/2	4123/1 ; 2341/3
1243/3 ; 1243/3	2143/13 ; 2143/13	3142/13 ; 2413/2	4132/13 ; 2431/23
1324/2 ; 1324/2	2314/2 ; 3124/1	3214/12 ; 3214/12	4213/12 ; 3241/13
1342/3 ; 1423/2	2341/3 ; 4123/1	3241/13 ; 4213/12	4231/13 ; 4231/13
1423/2 ; 1342/3	2413/2 ; 3142/13	3412/2 ; 3412/2	4312/12 ; 3421/23
1432/23 ; 1432/23	2431/23 ; 4123/13	3421/23 ; 4312/12	4321/123 ; 4321/123

There are 7 tables with row and column sums 112:

100	010	100	010	001	001	001
010	100	001	001	100	010	001
002	002	011	101	011	101	110

Now $D(1,1,2) = \{1,2\}$; from the display, pairs π, π^{-1} with $D(\pi) \subseteq \{1,2\}$, $D(\pi^{-1}) \subseteq \{1,2\}$ are

1234,1234 1324,1324 2134,2134 2314,3124 3124,2314 3214,3214 3412,3412.

Proof. of the theorem. Consider two compositions, $\mathbf{r}$ and $\mathbf{c}$ of N. A permutation π can be represented by a permutation matrix

$$\rho(\pi)_{ij} = \begin{cases} 1 & \text{if } \pi(i) = j \\ 0 & \text{otherwise.} \end{cases}$$

Thus, the 1 in the i^{th} row is in position $\pi(i)$ and $\pi(i+1) < \pi(i)$ says the 1 in the $i+1$st row occurs to the left of the one in the i^{th} row. Divide ρ into blocks specified by $\mathbf{r}$ and $\mathbf{c}$. Then, π has $D(\pi) \subseteq \{r_1, r_1+r_2, \cdots, r_1 + \cdots + r_{m-1}\}$ if and only if the pattern of ones in each horizontal strip decreases from upper left to lower right. Since $\rho(\pi^{-1}) = \rho(\pi)^t$, $D(\pi^{-1}) \subset \{c_1, c_1+c_2, \cdots, c_1 + \cdots + c_{n-1}\}$ if and only if the pattern of ones in each vertical strip decreases from upper left to lower right.

With this representation, there is a one-to-one correspondence between tables $T \in \Sigma_{\mathbf{rc}}$ and permutations satisfying the constraints: form a permutation matrix with T_{ij} ones in the (i,j) block which also satisfies the monotonicity constraints. There is a unique way to do this: the T_{11} ones in the $(1,1)$ block must be contiguous along the diagonal, starting at $(1,1)$. The T_{12} ones in the $(1,2)$ block must be contiguous on a diagonal starting at $(T_{11}+1, c_1+1)$; that is as far to the left and high up as possible consistent with the monotonicity constraints. The first horizontal block of $\rho(\pi)$ is similarly specified. Now the entries in the (2,1) block and then the second horizontal strip are forced and so on. Continuing, we see that $\rho(\pi)$ is uniquely determined by T. □

Foulkes (1976) and Garsia and Remmel (1985, pp. 233–234) give related results in terms of Schur functions. Briefly, let $c^{\nu}_{\lambda,\mu}$ be the Littlewood-Richardson numbers. These may be defined by $s_\lambda s_\mu = \Sigma_\nu c^{\nu}_{\lambda,\mu} s_\nu$ where s_λ, s_μ, s_ν are Schur functions. To connect these to descents, let D be a subset of $\{1, 2, \cdots, n-1\}$. Construct a skew diagram $\nu(D)/\lambda(D)$ consisting of a connected string of n boxes starting in the first column and ending in the first row, by moving right at step i if $i \in D$ and up if $i \in D^c$. For example, $D = \{1, 2, 4, 6, 7\}$ corresponds to the skew diagram :

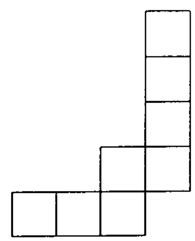

FIG. 3.1.

Foulkes showed that the number of permutations with descent set D whose inverse has descent set E is

$$\sum_{\mu} c^{\nu(D)}_{\lambda(D)\mu} c^{\nu(E)}_{\lambda(E)\mu}.$$

Because of the theorem above, summing this expression in D and E gives an expression for $|\Sigma_{\mathbf{rc}}|$.

3.2. Double cosets. Given a partition $\mathbf{r}$ of N, let $S_{\mathbf{r}}$ be the subgroup of the symmetric group S_N that permutes the first r_1 elements among themselves, the next r_2 elements among themselves, and so on. This $S_{\mathbf{r}}$ is called a Young subgroup and is a basic tool in developing the representation theory of S_N. It is isomorphic to the direct product of the S_{r_i}. Two Young subgroups $S_{\mathbf{r}}$ and $S_{\mathbf{c}}$ can be used to define double cosets. These are equivalence classes for the following relation:

$$\pi \sim \sigma \quad \text{if} \quad \rho\pi\kappa = \sigma \quad \text{for some } \rho \in S_{\mathbf{r}},\ \kappa \in S_{\mathbf{c}}.$$

The following lemma is a classical combinatorial fact:

LEMMA 3.3. *In the symmetric group S_N, the number of double cosets for $S_{\mathbf{r}}, S_{\mathbf{c}}$ equals $|\Sigma_{\mathbf{rc}}|$.*

Proof. The correspondence between tables and cosets has the following combinatorial interpretation: consider N balls labeled 1 up to N. Color the first r_1 balls with color 1, the next r_2 balls with color 2, and so on. Let $\pi \in S_N$ permute the labels. Construct a table $T(\pi)$ as follows: look at the first c_1 places in π and for each color i count how many balls of color i occur in these places. Call these numbers $T(\pi)_{i,1}$. Then consider the next c_2 places in π and count how many balls of each color i occur. Call these numbers $T(\pi)_{i,2}$. Continuing gives a table $T(\pi) \in \Sigma_{\mathbf{rc}}$. It is not hard to check that every table in $\Sigma_{\mathbf{rc}}$ is $T(\pi)$ for some π and that $T(\pi) = T(\sigma)$ if and only if π and σ are in the same double coset. Thus the number of double cosets equals $|\Sigma_{\mathbf{rc}}|$. □

A group-theoretic proof of the lemma is given by James and Kerber (1981, Cor. 1.3.11). The above proof clearly shows that $\Sigma_{\mathbf{rc}}$ is nonempty.

3.3. Induced representations and tensor products. Let G be a finite group; $H \subseteq G$ a subgroup, $X = G/H$ the associated coset space, and $L(X)$ the vector space of all real valued functions on X. The group G acts on $L(X)$ by left translation: $sf(x) = f(s^{-1}x)$. The resulting representation is denoted $Ind_H^G(triv)$: the representation of G induced from the trivial representation of H. For $G = S_N$, with Young subgroup $S_{\mathbf{r}}$, these representations arise in the statistical analysis of "partially ranked data of shape $\mathbf{r}$". See Diaconis (1988, 1989). They are also the building blocks of most constructions of the irreducible representations of S_N.

For $\mathbf{r}$ and $\mathbf{c}$ partitions of N a classical theorem of Mackey (see, e.g., James and Kerber (1981, p. 17)) studies the intertwining number $I(\mathbf{r}, \mathbf{c})$: the dimension of the space of linear maps from $Ind_{S_{\mathbf{r}}}^{S_N}(triv)$ to $Ind_{S_{\mathbf{c}}}^{S_N}(triv)$ that commute with the action of S_N.

THEOREM 3.4 (MACKEY). *For $\mathbf{r}$ and $\mathbf{c}$ partitions of N,*

$$I(\mathbf{r}, \mathbf{c}) = |\Sigma_{\mathbf{rc}}|.$$

There is a related appearance. Let (ρ_1, V_1) and (ρ_2, V_2) be linear representations of a finite group G. The tensor product is the set of matrices $\rho_1(s) \otimes \rho_2(s)$ for $s \in G$. This gives a representation of G on $V_1 \otimes V_2$. One of the basic problems of representation theory is the study of how tensor products decompose.

If $M^{\mathbf{r}} = Ind_{S_{\mathbf{r}}}^{S_N}(triv)$, one can state

THEOREM 3.5 (MACKEY). *For $\mathbf{r}$ and $\mathbf{c}$ partitions of N*

$$M^{\mathbf{r}} \otimes M^{\mathbf{c}} = \bigoplus_T M^S$$

where the sum runs over all tables T in $\Sigma_{\mathbf{rc}}$ and the partition $S = S(T)$ is derived from T by taking all the entries of T in order.

For example, take $\mathbf{r} = \mathbf{c} = (N-1, 1)$. There are two tables in $\Sigma_{\mathbf{rc}}$:

$$\begin{matrix} N-1 & 0 \\ 0 & 1 \end{matrix} \qquad \begin{matrix} N-2 & 1 \\ 1 & 0 \end{matrix}.$$

Thus

$$M^{N-1,1} \otimes M^{N-1,1} = M^{N-1,1} \oplus M^{N-2,1,1}.$$

This theorem is proved by James and Kerber (1981, pp. 95–98). It gives a convenient way of decomposing tensor products of irreducible representations as well. There is a general interrelation between double cosets, induced representations and tensor products which includes these results as a special case. Curtis and Reiner (1962) develop this in some detail. The special case has fascinating ramifications not developed here. This concerns Solomon's descent algebras which connect to Lie theory, card shuffling and much else. See Solomon (1976), Garsia (1990), Garsia and Reutenauer (1989) and Diaconis, McGrath, Pitman (1993).

4. Symmetric functions. A polynomial in n variables is called symmetric if it is invariant under every permutation of its variables. Let Λ_n be the ring of symmetric polynomials in variables $x_1, x_2, \cdots, x_n$, with integer coefficients. This ring decomposes into a direct sum of subrings

$$\Lambda_n = \bigoplus_{k \geq 0} \Lambda_n^k$$

where Λ_n^k consists of the homogeneous symmetric polynomials of degree k, together with the zero polynomial. The best reference for these matters is Macdonald (1979). We use his notation in this section.

There are a number of well known bases for Λ_n^k. All are indexed by partitions λ of k with n or fewer parts: $\lambda = (\lambda_1, \cdots, \lambda_n)$, $\lambda_1 + \cdots + \lambda_n = k$, $\lambda_i \geq 0$. Let $x^\lambda = x_1^{\lambda_1} \cdots x_n^{\lambda_n}$.

- Monomial symmetric functions m_λ are defined by

$$m_\lambda(x_1, \cdots, x_n) = \Sigma x^\alpha$$

summed over all distinct permutations of $(\lambda_1, \cdots, \lambda_n)$. Thus, if $n = 3, k = 4, \lambda = (2,1,1)$, $m_{211}(x_1, x_2, x_3) = x_1^2x_2x_3 + x_1x_2^2x_3 + x_1x_2x_3^2$. These m_λ are symmetric and form a basis for Λ_n^k.

- Elementary symmetric functions e_λ are defined by

$$e_j = \sum_{1 \leq i_1 < \cdots < i_j \leq n} x_{i_1} x_{i_2} \cdots x_{i_j} \quad \text{and} \quad e_\lambda = e_{\lambda_1} e_{\lambda_2} \cdots e_{\lambda_n}.$$

Thus $e_{211}(x_1, x_2, x_3) = (x_1x_2 + x_2x_3 + x_1x_3)(x_1 + x_2 + x_3)^2$. These also form a basis for Λ_n^k.

- The complete symmetric functions h_λ are defined by

$$h_j = \sum_{1 \leq i_1 \leq \cdots \leq i_j \leq n} x_{i_1} x_{i_2} \cdots x_{i_j} \quad \text{and} \quad h_\lambda = h_{\lambda_1} h_{\lambda_2} \cdots h_{\lambda_n}.$$

Thus $h_{211}(x_1, x_2, x_3) = (x_1^2 + x_2^2 + x_3^2 + x_1x_2 + x_1x_3 + x_2x_3)(x_1 + x_2 + x_3)^2$. These again form a basis for Λ_n^k.

- The power sum symmetric functions are defined by

$$p_j = \sum_i x_i^j, \qquad p_\lambda = p_{\lambda_1} \cdots p_{\lambda_n}.$$

Thus $p_{211}(x_1, x_2, x_3) = (x_1^2 + x_2^2 + x_3^2)(x_1 + x_2 + x_3)^2$. These form a basis for Λ_n^k over $\mathbb{Q}$ but not over $\mathbb{Z}: h_2 = \frac{1}{2}(p_1^2 + p_2)$.

A scalar product can be defined on Λ_n^k by requiring that h and m are dual:

$$\langle h_\lambda | m_\mu \rangle = \delta_{\lambda\mu}$$

for δ the Kronecker delta. With this choice, the p_λ bases are orthogonal:

$$\langle p_\lambda | p_\mu \rangle = \delta_{\lambda\mu} z_\lambda,$$

where z_λ is defined in terms of the partition $\lambda = 1^{a_1}2^{a_2} \cdots k^{a_k}$, by

$$z_\lambda = \prod_{i=1}^k i^{a_i} a_i!\,.$$

The following well known result connects all of this to arrays.

Theorem 4.1. *If* $\mathbf{r}$ *and* $\mathbf{c}$ *are partitions of* N,

$$\langle h_{\mathbf{r}} | h_{\mathbf{c}} \rangle = |\Sigma_{\mathbf{rc}}|.$$

Proof. The transition matrix between the dual bases h and m is given by Macdonald as

$$h_{\mathbf{r}} = \sum_{\mathbf{c}} |\Sigma_{\mathbf{rc}}| m_{\mathbf{c}}.$$

Taking the inner product on both sides with $h_{\mathbf{c}}$ completes the argument. □

We can draw two consequences from this result. The first is an algorithm for computing $|\Sigma_{\mathbf{rc}}|$. The second is a formula for this number.

The theorem offers a variety of schemes for computing $|\Sigma_{\mathbf{rc}}|$ by using currently available algorithms for computing with symmetric functions. Perhaps the best available tool is John Stembridge's package which runs in connection with Maple. An algorithm can be based on the identities above together with the following result (Macdonald (1979, p. 17)) which expresses the relation between the h and m bases:

$$h_j = \sum_{\lambda \vdash j} p_\lambda / z_\lambda.$$

Algorithm 4.2. (*to compute* $|\Sigma_{\mathbf{rc}}|$). Let r, c be given. Express $h_{\mathbf{r}} = \prod_{i=1}^{m} h_{r_i}$ and $h_{\mathbf{c}}$ in terms of power sums. Then compute the inner product $\langle h_{\mathbf{r}} | h_{\mathbf{c}} \rangle$. Note that all terms involved in this algorithm involve positive quantities. Thus, truncation at any stage gives a lower bound.

As an example: there are two 2×2 tables with row and column sums $(2,1)$: $\begin{matrix} 2 & 0 \\ 0 & 1 \end{matrix}$ and $\begin{matrix} 1 & 1 \\ 1 & 0 \end{matrix}$. To compute $\langle h_{21} | h_{21} \rangle$, express $h_2 = \frac{1}{2}(p_2 + p_1^2)$, $h_1 = p_1$ and then $\langle h_{21} | h_{21} \rangle = \frac{1}{4}\langle (p_2 + p_1^2)p_1 | (p_2 + p_1^2)p_1 \rangle = \frac{1}{4}\{\langle p_2p_1 | p_2p_1 \rangle + \langle p_1^3 | p_3p_1 \rangle\} = \frac{1}{4}\{2 + 6\} = 2$.

Alas, in practice, computation does not seem so feasible for N above 100 or so. As will emerge, other algorithms work well with much larger N. As such N arise in practice, improvements in symmetric function technology are needed before algorithm (4.2) becomes feasible.

Just carrying through the algorithm gives the following formula:

COROLLARY 4.3. *Let* **r** *and* **c** *be partitions of* N. *Then*

$$|\Sigma_{\mathbf{rc}}| = \sum_{\substack{\mu^i \vdash r_i \\ \lambda^j \vdash c_j}} \delta_{\lambda\mu} z_\lambda \prod \frac{1}{z_{\mu_i} z_{\lambda_j}}.$$

The sum is over all partitions μ^i *of* r_i, λ^j *of* c_j *for* $1 \le i \le m$, $1 \le j \le n$ *and* $\lambda = (\lambda^1, \cdots, \lambda^m)$, $\mu = (\mu^1, \cdots, \mu^n)$ *are considered partitions of* N *by concatenating parts.*

Remark 4.4. In Section 7.2 we give the generating function for the number of tables as $\prod_{i,j}(1 - x_i y_j)^{-1} = e^{\Sigma p_i(x) p_i(y)/i}$. On the right, $p_i(x)$ is the power sum symmetric function. The corollary can also be read of this expansion by expanding the exponential in the usual way. Richard Stanley used this technique to get numerical answers to problem 27974 in the American Mathematical Monthly (1980).

5. Young tableaux and Kostka numbers. For λ and μ partitions of n, a *semi-standard Young tableau of shape* λ *and content* μ is a diagram of shape λ containing μ_1 ones, μ_2 twos, etc., arranged to be weakly increasing in rows and strictly increasing down columns. For example,

$$\begin{matrix} 1 & 1 & 2 \\ 2 & 3 & \end{matrix}$$

is semi-standard with shape 3,2 and content 2,2,1. Such tableaux are a basic ingredient in the description of the irreducible representations of the classical groups.

Define the Kostka number $K_{\lambda\mu}$ as the number of semi-standard tableaux of shape λ and content μ. The following classical result relates these to arrays.

THEOREM 5.1. *Let* **r** *and* **c** *be partitions of* N. *Then*

$$|\Sigma_{\mathbf{rc}}| = \sum_{\mu} K_{\mathbf{r}\mu} K_{\mathbf{c}\mu}.$$

Proof. The Schur functions s_λ are yet another basis for the ring Λ_n^k of Section 4. These can be expressed in terms of the h_λ as

$$h_\lambda = \sum_{\mu} K_{\lambda\mu} s_\mu.$$

Macdonald (1979, p. 57) proves this as well as showing that $\langle s_\lambda | s_\mu \rangle = \delta_{\lambda\mu}$. Now $|\Sigma_{\mathbf{rc}}| = \langle h_{\mathbf{r}} | h_{\mathbf{c}} \rangle$, so the result follows. □

Remark 5.2. 1. A direct combinatorial proof of the theorem is given by Knuth (1970). Extending ideas of Robinson, Schensted, and Schützenberger, he gives a bijection between tables with row sum **r** and column sum **c** and pairs of semi-standard tableaux with shape **r** and content **c**.
2. There has been some recent work on formulae for the Kostka numbers $K_{\lambda\mu}$. At the moment, these do not seem so useful, but here is a brief description. Kirillov and Reshitikhin (1986) have shown

$$K_{\lambda\mu} = \sum_{\alpha} \prod_{k,n \ge 1} [p_n^k(\alpha), \alpha_n^k - \alpha_{n+1}]$$

where the sum is over all sequences of partitions $\alpha = (\alpha^0, \alpha^1, \alpha^2 \cdots)$ such that $\alpha^0 = \mu$, $|\alpha^k| = \lambda_{k+1} + \lambda_{k+2} + \cdots$, $k \ge 1$, further $p_n^k(\alpha) = \sum_{i=1}^{n}(\alpha_i^{k-1} - 2\alpha_i^k + \alpha_i^{k+1})$ and $[a,b] = \binom{a+b}{b}$.

6. Tables and statistics. This section motivates statistical uses of arrays and explains how enumeration of $\Sigma_{\mathbf{rc}}$ arises in application. Data is often categorized into 2-way contingency tables. The following example is typical: 592 subjects were classified by hair and eye color (Snee (1974)). Such data are often analyzed under the assumption that the cell counts T_{ij} follow a multinomial distribution with probability p_{ij}. The independence hypothesis can be specified as

$$p_{ij} = p_{i\cdot} p_{\cdot j} \quad \text{for all } i \text{ and } j, \quad \text{where} \quad p_{i\cdot} = \sum_j p_{ij}, \quad p_{\cdot j} = \sum_i p_{ij}.$$

A standard test of independence uses the chi-squared statistic

$$\chi^2 = \sum_{i,j} \frac{(T_{ij} - \frac{r_i c_j}{n})^2}{\frac{r_i c_j}{n}}, \qquad \text{with} \quad r_i = \sum_j T_{ij}, \quad c_j = \sum_i T_{ij}.$$

In Table 6.1, $\chi^2 = 138.29$.

TABLE 6.1

	Black	Brunette	Red	Blonde	
Brown	68	119	26	7	220
Blue	20	84	17	94	215
Hazel	15	54	14	10	93
Green	5	29	14	16	64
	108	286	71	127	592

The chi-squared value is usually compared with an approximation from the chi-squared distribution. In this example, the approximation has mean 9 and standard deviation $\sqrt{18}$, so 138.29 is a huge value and independence is rejected.

Huge values of chi-squared are sufficiently common that statisticians have developed other ways of calibrating the chi-square statistic. For example, Diaconis and Efron (1985) assumed that the underlying probabilities p_{ij} were unknown and put a uniform prior on them. This is just Lebesgue measure on the simplex $\{p_{ij} : p_{ij} \geq 0, \Sigma p_{ij} = 1\}$. Under this assumption, it turns out that the table T_{ij} is also uniform: all tables have an equal chance of occurring. This suggests calibrating the distribution of χ^2 under the uniform distribution as an antagonistic alternative to the model of independence.

In statistics, it is customary to fix (or condition on) the row and column sums of the observed table and ask for calibration of test statistics in the set of possible tables. This leads to the following combinatorial problem: for r and c partitions of N, find the proportion of tables in $\Sigma_{\mathbf{rc}}$ with chi-squared values smaller than t, as t varies.

Diaconis and Efron (1985) develop a variety of techniques to approximate this distribution and describe related work by Good and other statisticians. A Monte Carlo algorithm discussed in Section 10 below suggests that in fact, about 15.4% of all tables have $\chi^2 \leq 138.29$. Thus here the data is compatible with the antagonistic alternative.

Bad!

Enumeration of $\Sigma_{\mathbf{rc}}$ enters the picture at several points. At present, the only way we have of exactly solving the calibration problem under the uniform distribution on $\Sigma_{\mathbf{rc}}$ is to systematically run through all tables and actually calculate the statistic. It is of obvious interest to have an estimate of the number of tables to have some idea of the running time.

The size of $|\Sigma_{\mathbf{rc}}|$ is similarly used to estimate the running time of algorithms for exact enumeration of the chi-squared statistic under other distributions on $\Sigma_{\mathbf{rc}}$. The most important of these is the Fisher-Yates (or multiple hypergeometric) distribution. A variety of algorithms for doing these computations are reviewed in Sections 8 and 9.

Finally, the size of $|\Sigma_{\mathbf{rc}}|$ was required as input to an approximation procedure proposed by Diaconis and Efron (1985).

In Section 8.1 we give algorithms for calculating $|\Sigma_{\mathbf{rc}}|$ which work for tables like Table 6.1: there are approximately 10^{15} tables with the same row and column sums as Table 6.1; in fact, there are $1,225,914,276,768,514$ such tables.

We conclude this section by mentioning some literature related to privacy issues. Suppose a contingency table is summarized by giving its row and column sums and perhaps a few other functions. How much about the individual cell entries can be deduced? Gusfield (1988) gives best possible bounds on the entries and a review of this literature.

7. Asymptotic approximations. A variety of asymptotic approximations to $|\Sigma_{\mathbf{r},\mathbf{c}}|$ have been suggested (and proved!). O'Neil (1969), Békéssy, Békéssy and Komlós (1972) followed by Good and Crook (1977) give

$$(7.1) \qquad |\Sigma_{\mathbf{rc}}| \sim \frac{N!}{\Pi r_i! \Pi c_j!} \exp\left\{\frac{2}{N^2}\sum_{i,j}\binom{r_i}{2}\binom{c_j}{2}\right\}.$$

O'Neil proved this as m, n tend to infinity with the row sums of less than $\{\log \max(m,n)\}^{1/4-\epsilon}$; that is for large, sparse tables. Bender (1974) developed variations for prescribed zeros and bounded entries. Békéssy, Békéssy and Komlós proved it was valid for m fixed and n large. Alas, this approximation is useless for tables like Table 6.1, 8.1 above where both m and n are small and N is large. They are off by many orders of magnitude for these examples.

Diaconis and Efron (1986) gave approximations which seem to work for m, n small, N large. To state the result, let

$$w = \frac{1}{1 + mn/2N}, \quad k = \frac{n+1}{n\Sigma \bar{r}_i^2} - \frac{1}{n}, \quad \bar{r}_i = \frac{1-w}{m} + \frac{wr_i}{N}, \quad \bar{c}_j = \frac{1-w}{n} + \frac{wc_j}{N}.$$

Then Diaconis and Efron suggest (without proof)

$$|\Sigma_{\mathbf{rc}}| \sim \left(\frac{2N+mn}{2}\right)^{(m-1)(n-1)} \left(\prod_{i=1}^{m} \bar{r}_i\right)^{n-1} \left(\prod_{j=1}^{m} \bar{c}_i\right)^{k-1} \frac{\Gamma(nk)}{\Gamma(n)^m (k)^n}.$$

(7.2)

For example, this gives 1.235×10^{15} as an approximation of the number of tables with the same margins as Table 6.1. The right answer is 1.226×10^{15}. For Table 8.1 it gives 2.33×10^8, the right answer being 2.39×10^8.

Good and Crook (1977) suggest several further approximations the simplest of these, (6.5) of their paper, gives 1.1432×10^15. For Table 6.1 and the remarkable 2.3939×10^8 for table 8.1.

Finally, Gail and Mantel (1977) have suggested a normal approximation which we have not found terribly reliable.

8. Algorithms for exact enumeration. This section describes many algorithms which have been developed for counting or actually running through all $m \times n$ tables with row sum $\mathbf{r}$ and column sums $\mathbf{c}$. For the $2 \times n$ and $3 \times n$ case, formulas of Mann are described in Section 8.4 below. Algorithms using symmetric functions were described in Sections 4 and 5. Of course, the space of tables $\Sigma_{\mathbf{rc}}$ is large and the complexity considerations of Section 9 may rule out success. Nonetheless, we have been pleasantly surprised at how often exact answers are available for problems of practical interest.

8.1. Exact enumeration. A number of authors, March (1972), Boulton and Wallace (1973), Hancock (1974), Balmer (1988) have suggested algorithms for exhaustively stepping through the set $\Sigma_{\mathbf{rc}}$ one table at a time. These begin at some canonically constructed initial table and proceed by making small changes to the cell entries so that the tables increase monotonically in some linear order. This affords a means of calculating $|\Sigma_{\mathbf{rc}}|$ as well as various other functions on the set such as the proportion of tables with a chi-square (or other statistic) smaller than a given value.

As an example, Table 8.1 below was chosen as a table with irregular margins. There are 239,382,173 tables in $\Sigma_{\mathbf{rc}}$. The exhaustive enumeration

TABLE 8.1

5	2	3	10
50	7	5	62
3	6	4	13
5	3	3	11
2	7	30	39
65	25	45	135

using Pagano and Taylor-Halvorsen's algorithm on a DEC station 3100 (a vintage 1990 desktop workstation) took 54665 CPU seconds and 15 hours, 21 minutes of real time to calculate the exact chi-square distribution. John Mount reported that counting alone took 17 seconds using the divide and conquer ideas explained below.

Table 8.1 has $\chi^2 = 72.18$. A histogram of the χ^2 values for all tables in $\Sigma_{\mathbf{rc}}$ is shown in Figure 8.1. The exact proportion of tables with chi-square value smaller than 72.18 is .76086 to 5 significant figures. This will be used to illustrate some other approaches (Figure 8.1 is explained in Section 10.2.)

If one is interested in simply determining the proportion of tables with chi-square values smaller than a given value, one may follow R.A. Fisher (1935) and "work in from the end". This avoids generating every table.

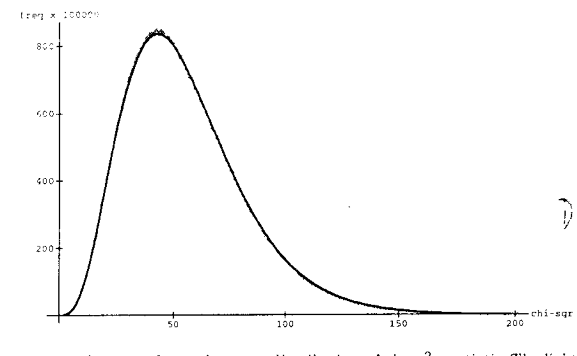

FIG. 8.1. *The dark curve shows the exact distribution of the χ^2 statistic. The light curve shows a Monte Carlo approximation using the random walk of section 10.2*

Pagano and Taylor-Halvorsen (1981) have developed such a short cut for the hypergeometric distribution on $\Sigma_{\mathbf{rc}}$.

One straightforward approach to table enumeration uses a recurrence. Gail and Mantell (1977) carry this out. For example, in the $m \times 3$ case, let $T(r_1, \cdots, r_m; c_1, c_2) = |\Sigma_{\mathbf{rc}}|$ then

$$T(\mathbf{r}, \mathbf{c}) = \sum_{k_1, k_2} T(r_1, r_2, \cdots, r_{m-1}; c_1 - k_1, c_2 - k_2).$$

The sum on the right runs over values k_1, k_2 with $0 \leq k_i \leq \min(r_m, c_i)$ and $k_1 + k_2 \leq r_m$. Of course, the generalization of this recurrence to arbitrary dimensions takes exponential time to compute.

We mention briefly two other algorithms which have seen extensive empirical application. Stein and Stein (1970) have proposed a branching algorithm. This is extended by Good and Crook (1977) who give a clear description. Finally, the network algorithm of Mehta and Patel (1983) is a mainstay of the commercial program Statexact. It is geared for exact evaluation of hypergeometric distributions for statistics on $\Sigma_{\mathbf{rc}}$.

Finally, we mention that David des Jardin (personal communication) and John Mount (personal communication) have employed a divide and conquer algorithm. Mount's algorithm works by running through all possible $2 \times n$ tables T with column sums c and row sums $\mathbf{r}'$, where $r'_1 = r_1 + \cdots + r_{\lfloor \frac{m}{2} \rfloor}$, $r'_2 = r_{\lfloor \frac{m}{2} \rfloor + 1} + \cdots + r_m$. Let $\mathbf{r}^\ell$ be the $\lfloor \frac{m}{2} \rfloor$ vector consisting of the first $\lfloor \frac{m}{2} \rfloor$ entries of $\mathbf{r}$ and $\mathbf{r}^R$ be the $m - \lfloor \frac{m}{2} \rfloor$ vector consisting of the last entries of $\mathbf{r}$. Let T_i denote the i^{th} row of T. Then

$$m(\mathbf{r}, \mathbf{c}) = \Sigma_T m(\mathbf{r}^\ell, T_1) \times m(\mathbf{r}^\ell, T_2),$$

where T runs through all legal $2 \times n$ tables having row sums $\mathbf{r}'$ and column sums c. The smaller counting problems are solved by the same technique. The recursion ends when a $1 \times n$ or $2 \times n$ problem is reached. This algorithm has solved the hardest problems to date, the 4×4 example in Table 6.1, which has 1,225,914,276,768,514 tables with the same row and column sums.

8.2. Generating functions and Fourier transforms. Let $x_1, x_2, \cdots, x_m$ and $y_1, y_2, \cdots, y_n$ be variables. Form the generating function

$$\prod_{i,j}(1 - x_i y_j)^{-1} = (1 + x_1 y_1 + (x_1 y_1)^2 + \cdots)(1 + x_1 y_2 + (x_1 y_2)^2 + \cdots) \cdots (1 + x_m y_m + (x_m y_m)^2 + \cdots). \tag{8.1}$$

By inspection, the coefficient of $x_1^{r_1} x_2^{r_2} \cdots x_m^{r_m} y_1^{c_1} y_2^{c_2} \cdots y_n^{c_n}$ is $|\Sigma_{\mathbf{rc}}|$. For example, the coefficient of $x_1^2 x_2 y_1 y_2 y_3$ is 3. The 3 tables with $\mathbf{r} = (2, 1)$, $\mathbf{c} = (1, 1, 1)$ are

$$\begin{matrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{matrix} \qquad \begin{matrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{matrix} \qquad \begin{matrix} 0 & 1 & 1 \\ 1 & 0 & 0 \end{matrix} .$$

The coefficients in the generating function can be expressed as contour integrals and one can attempt asymptotic approximations. Good (1976) pursues this line; see Section 9.

Along more algorithmic lines, we can truncate the expansions above and compute the initial portions of the product by multiplying polynomials. This can in turn be done using fast Fourier transform (F.F.T.)

We outline the method and study its running time. Let $\mathbf{i} = (i_1, i_2, \cdots, i_k)$ and let $z^{\mathbf{i}} = z_1^{i_1} \cdots z_k^{i_k}$. A polynomial is $f(z_1, \cdots, z_k) = f(z) = \Sigma_{\mathbf{i}} a_{\mathbf{i}} z^{\mathbf{i}}$. Two k variable polynomials of degree at most d in each variable can be multiplied by using the F.F.T. on the group $\mathbf{Z}_{2d}^k$ in $O(k(2d)^k \log d)$ operations. To see this, suppose $f(z) = \Sigma_{\mathbf{i}} a_{\mathbf{i}} z^{\mathbf{i}}$ and $g(z) = \Sigma_{\mathbf{i}} b_{\mathbf{i}} z^{\mathbf{i}}$. Their product is $h(z) = \Sigma_k c_{\mathbf{k}} z^{\mathbf{k}}$ where $c_{\mathbf{k}} = \Sigma_{\mathbf{i}+\mathbf{j}=k} a_{\mathbf{i}} b_{\mathbf{j}}$. Such convolutions can be computed by the F.F.T. on $\mathbf{Z}_{2d}^k$ in $O(k(2d)^k \log d)$ operations. One useful reference is Cormen, Leiserson, Rivest (1990, Chapter 32).

Moving back to tables, let $d_* = \max\{r_i, c_j, 1 \le i \le m, 1 \le j \le n\}$. The coefficient of the term $x^{\mathbf{r}} y^{\mathbf{c}}$ in the generating function (1.1) cannot depend on any term in the product with degree exceeding d_*. Successively multiplying each of the mn polynomials and discarding terms of excess degree gives the result we want. We summarize:

LEMMA 8.1. *Given vectors* $\mathbf{r}$ *and* $\mathbf{c}$, *of length* m *and* n, *let* $d_* = \max\{r_i, c_j, 1 \le i \le m, 1 \le j \le n\}$. *There is an algorithm for computing* $|\Sigma_{\mathbf{rc}}|$ *using* $O(mn(m+n)(2d_*)^{m+n} \log d_*)$ *operations on* $0((2d_*)^k)$ *numbers.*

Remark 8.2. 1. For fixed m and n this gives an algorithm for computing $|\Sigma_{\mathbf{rc}}|$ whose running time is a polynomial in N. Similar ideas are suggested by Good and Crook (1977).

2. The polynomial technique can be adapted to count elements T in $\Sigma_{\mathbf{rc}}$ which satisfy additional linear constraints of the form

$$\sum_{i,j} a_{ij} T_{ij} = a.$$

This can be done by using the generating function

$$\prod_{i,j}(1 - s^{a_{ij}} x_i y_j)^{-1}.$$

The coefficient of $s^a x^{\mathbf{r}} y^{\mathbf{c}}$ gives the number of tables in $\Sigma_{\mathbf{rc}}$ satisfying the constraint. Any number of constraints can be handled this way.

For example, the generating function for $n \times n$ "magic squares" with diagonal sums equal to row and column sums can be expressed as

$$\prod_{i,j}(1 - s^{a_{ij}} t^{b_{ij}} x_i y_j)^{-1},$$

where $a_{ij} = 1$ if $i = j$ and zero otherwise, and $b_{ij} = 1$ if $i + j = n$ and zero otherwise.

Such additional constraints arise naturally in contingency table analysis. For example, Agresti, Mehta and Patel (1990) needed the number of tables with prescribed row and column sums and an additional constraint as above with $a_{ij} = u_i v_j$ for specific u_i, v_j.

The computational approach centered around polynomials has been actively developed by workers in computational statistics. Baglivio (1994) gives a book length development centered around this theme. She gives an extensive review of the statistical literature.

8.3. Ehrhart polynomials and toric ideals. There has been an active recent development in combinatorial mathematics, commutative algebra, and algebraic geometry which leads to useful algorithms for table enumeration. Briefly, the number of tables with given row and column sums can be shown to be a piecewise polynomial in $\mathbf{r}$ and $\mathbf{c}$ which shifts its coefficients at boundaries specified by well-specified hyperplanes. Further, the polynomials can be identified by using geometric properties of an associated "toric ideal". The following example was kindly communicated by Bernd Sturmfels.

Let $T(r_2, r_3, r_4, c_1, c_2, c_3, c_4)$ be the number of 4×4 tables with row sums r_1, r_2, r_3, r_4 and column and sums c_1, c_2, c_3, c_4, where $r_1 = c_1 + c_2 + c_3 + c_4 - r_2 - r_3 - r_4$. Consider the following closed convex polyhedral cone C which is defined by the following 11 linear inequalities in the 7-dimensional space of marginal totals:

$$c_1 \ge r_3, c_4 \ge r_3, c_3 \ge r_4, c_1 + c_4 \ge r_2, r_3 + r_4 \ge c_4, r_3 + r_4 \ge c_1, c_2 \ge r_2 + r_4$$
$$r_2 + r_4 \ge c_1 + c_4, c_2 + c_3 \ge r_2 + r_3, r_2 + r_3 \ge c_1 + c_3 + c_4, r_2 + r_3 + r_4 \ge c_2 - c_3.$$

This cone has 18 extreme rays and it contains the marginal totals of the data in Table 6.1 in its interior: $(215, 93, 64, 108, 286, 71, 127) \in Int(\mathcal{C})$. Sturmfels has proved that the restriction of T to $\mathcal{C}$ is a polynomial with rational coefficients of degree 9. As an example of what can be done, suppose that c_4 is varied by δ. How does that effect the number of tables? The polynomial specializes to the following formula which is valid for $-20 \leq \delta \leq 2$:

$$T(215, 93, 64, 108, 286, 71, 127 + \delta) =$$
$$-\tfrac{1}{40320}\delta^9 - \tfrac{37}{5040}\delta^8 + \tfrac{1871}{2240}\delta^7 + \tfrac{5269}{36}\delta^6 - \tfrac{60566509}{1920}\delta^5 - \tfrac{5875690069}{720}\delta^4$$
$$-\tfrac{19659536000567}{10080}\delta^3 - \tfrac{485892234635}{84}\delta^2 + \tfrac{105240591524160}{7}\delta + 1225914276768514.$$

Of course, when $\delta = 0$ we recover the number of tables from the final coefficient.

That T is a piecewise polynomial follows from the theory developed by E. Ehrhart. The clearest elementary treatment is in Stanley (1986, Sections 4.4, 4.6). This contains background and references. The quasi polynomial nature of the answer suggests an exciting possibility. One can determine the polynomial by computing with tiny marginal totals in the same part of the cone. This is how the magic square polynomials of Section 2 were determined.

The methods used to find the polynomial above lie somewhat deeper. Briefly, consider a set of integer vectors in $\mathbb{R}^d$. Form their convex hull $\mathcal{H}$ and consider the problem of finding N the number of lattice points in $\mathcal{H}$. Clearly the problem of table enumeration fits into this mold. Associated to $\mathcal{H}$ is a complex variety called a toric variety. There is a formula for N in terms of the geometry of this toric variety. It uses fairly abstract constructions such as the cohomology and Todd classes associated to the variety. Further, one can actually calculate the numerical ingredients of these geometric quantities. It would take us too far afield to develop these topics here. Danilov (1978) or Fulton (1993) give fine introductions which get to the relevant parts of the subject. Fulton gives references to recent work by Barvinock, Morelli, and Pomerschein on this subject.

8.4. Tables with fixed m and n. In applications, one is often in the situation in which many subjects are classified into a small number of categories. Table 6.1 of Section 6 is a typical example. We give a polynomial time algorithm (in N) for solving the problems in Section 8.2. Mann (1994) has given the following formulae for the case of $2 \times n$ or $3 \times n$ tables.

$$m_2(\mathbf{r};\mathbf{c}) = \sum_{\sigma \subset [1,n]} \left[\begin{matrix} r_1 - c_\sigma - |\sigma| \\ n-1 \end{matrix} \right] (-1)^{|\sigma|}$$
$$m_3(\mathbf{r};\mathbf{c}) = \left[\begin{matrix} r_1 \\ n-1 \end{matrix} \right] m_2(r_3, c_{[1,a]} - r_3; \mathbf{c})$$
$$\sum_{\substack{\sigma \subset [1,n] \\ |\sigma| \neq 0, n}} \quad \sum_{\substack{\tau : \tau \cap \sigma = \emptyset \\ \gamma : \gamma \subset \sigma}} (-1)^{|\sigma|+|\tau|+|\gamma|} f$$

where

$$f = \sum_{k=0}^{|\sigma|-1} (-1)^k \left[\begin{matrix} A+B-k \\ 2n - |\sigma| + k - 1 \end{matrix} \right] \left[\begin{matrix} A + c_\gamma + |\gamma| - |\sigma| \\ |\sigma| - |\gamma \end{matrix} \right] \left[\begin{matrix} n-1 \\ k \end{matrix} \right]$$
$$+(-1)^\sigma \sum_{j=1}^{n} \left[\begin{matrix} n-j \\ |\sigma|-1 \end{matrix} \right] \left[\begin{matrix} A + c_\gamma + |\gamma| - |\sigma| \\ j-1 \end{matrix} \right] \left[\begin{matrix} B - c_\gamma - |\gamma \\ 2n-j-1 \end{matrix} \right].$$

In these sums, $\left[\begin{matrix} n \\ k \end{matrix} \right] = {}^{n+k}_{\ \ k}$, $[1,n]$ is the interval from 1 to n; Greek letters denote subsets, and $c_\sigma = \sum_{j \in \sigma} c_j$. Finally, $A = r_1 - c_\sigma - |\sigma|$ and $B = r_3 - c_\tau - |\tau|$.

Similar formulae for the volume of the associated convex polyhedra appear in Diaconis and Efron (1985). For fixed n, these give a polynomial computation for $|\Sigma_{\mathbf{rc}}|$ in $\mathbf{r}$ and $\mathbf{c}$. The computation is also clearly exponential as n grows. The methods used to derive these formulae can be extended to ever more unsightly formulae for $m \times n$ tables. Dahmen and Micchelli (1988) relate such formulae to multivariate splines.

This puts the tension back in the problem; in a specific example, what is large? It seems feasible to us that there might also be tractable methods in sparse cases where m and n are large, but N is small or moderate.

9. Complexity results. This section records what is currently known about the complexity of computing $|\Sigma_{\mathbf{rc}}|$. There are 3 natural parameters that give a crude measure of problem size: m, n, and N. Briefly, the problem is hard (#-P complete) if m or n is large (theorem of Dyer, Kannan and Mount). If m and n are fixed and N is large, the problems are theoretically tractable. We begin with a brief self contained description of #-P completeness. The initiated may skip to Sections 9.2 and 9.3 which describe our results.

9.1. #-P completeness. The standard reference on this subject is Gary and Johnson (1979). One clear way to represent the relevant complexity classes begins with a finite alphabet Σ. Let Σ^* denote all finite sequences (words) of elements in Σ. We use $|x|$ to denote the length of x.

Let $R \subseteq \Sigma^* \times \Sigma^*$ be a relation on words. Write $R(x, y)$ to denote $(x, y) \in R$ and $R(x) = \{y : R(x, y)\}$: this is called the solution set of x. R is a *p-relation* if there are polynomials p and q such that:

- The predicate $R(x, y)$ can be checked in $p(|x|)$ operations.
- If $y \in R(x)$ then $|y| \leq q(|x|)$.

The well known class NP can be identified with the decision problems "Given x, is $R(x)$ non-empty?", where R is a P-relation. Similarly, the class $\#P$ can be identified with the counting problems "Given x what is the cardinality of $R(x)$?" where R is a P-relation.

Both in the case NP and $\#P$, there exist problems in the class that are complete under *polynomial time reductions*; these are problems ψ such that the existence of polynomial-time algorithms for ψ would imply the existence of polynomial-time algorithms for each problem in the class. Such problems are called NP complete or $\#P$ complete, respectively.

9.2. Some $\#$-P completeness results for table enumeration. Dyer, Kannan and Mount (1994) have announced the result that for $m = 2$ and large n, N, the problem of determining $|\Sigma_{\mathbf{rc}}|$ is $\#P$-complete. We present here an earlier result of Gangolli (1991) which proves $\#$-P completeness for a practical class of problems.

Let Z be an $m \times n$ binary matrix. The set $\Sigma^Z_{\mathbf{rc}}$ of contingency tables with structural zeros at Z is the subset of $\Sigma_{\mathbf{rc}}$ in which every table has only zeros in positions where $Z(i, j) = 1_0$. This set arises naturally in the analysis of contingency tables where the row/column classification gives rise to forbidden combinations. For example, a table that classifies a population of subjects by sex (rows) and cause of death (columns) might have a forbidden entry representing males with uterine cancer. See Bishop, Fienberg and Holland (1975) for background.

THEOREM 9.1. *With* $\mathbf{r}, \mathbf{c}$ *and* Z *as parameters, the counting problem for* $\Sigma^Z_{\mathbf{rc}}$ *is* $\#$ P *complete.*

Proof. First note that the counting function for the set $\Sigma^Z_{\mathbf{rc}}$ is in $\#$-P. To prove completeness, we give a reduction from the problem of computing the permanent. Here, if A is the $n \times n$ adjacency matrix of a bipartite graph on two sets of n vertices, $per(A)$ is the number of perfect matchings; this is $\sum_\pi \prod_{i=1}^n A_{i\pi(i)}$ summed over the permutations of n. The problem of computing the permanent was shown to be $\#$-P complete by Valiant (1979).

Computing the permanent is a special case of computing $|\Sigma^Z_{\mathbf{rc}}|$. Given the $n \times n$ adjacency matrix A for a bipartite graph G let $\mathbf{r} = \mathbf{c} = (1, 1, \cdots, 1)$ (length n). Let $Z_{ij} = 1 - A_{ij}$. Now it is clear that a table T is in $\Sigma^Z_{\mathbf{rc}}$ if and only if T is the adjacency matrix of a perfect matching in G. Thus, $|\Sigma^Z_{\mathbf{rc}}|$ equals the number of perfect matchings. The reduction can be done in linear time and logarithmic space in the size of A. $\square$

Remark 9.2. The theorem holds even if the inputs are expressed in unary. Roughly speaking, this means that the difficulty of this problem is really due to the structure of the problem and not just the size of the marginal totals. See Garey and Johnson (1979) for further discussion.

Throughout, it is natural to inquire about the natural generalizations to three-dimensional arrays. The problems are surprisingly more difficult; even determining if the set of tables with given line sums is non-empty is NP complete. See Irving and Jerrum (1990).

10. Approximate counting using sampling. In this section we show that there are randomized algorithms that give accurate approximations to $|\Sigma_{\mathbf{rc}}|$ in a polynomial number of steps in m, n, N. The algorithms work by choosing tables at random and using these to approximate the number of tables. The conversion is explained first, followed by two methods of random choice—a combinatorial random walk and a convex set approach due to Dyer, Kannan and Mount.

10.1. Random walks on $\Sigma_{\mathbf{rc}}$. Fix row and column sums $\mathbf{r}, \mathbf{c}$. A variety of random walks on $\Sigma_{\mathbf{rc}}$ have been in active use by statisticians. The idea is simple: pick a pair of rows i, i' and a pair of columns j, j' uniformly at random. These rows and columns intersect in four entries. The walk proceeds by changing the current table into a new one by adding and subtracting one in these entries according to the following pattern:

$$\begin{matrix} + & - \\ - & + \end{matrix} \quad \text{or} \quad \begin{matrix} - & + \\ + & - \end{matrix} .$$

The final choice is made with probability 1/2. If a step forces a negative table entry, the random walk stays at the original table. It is easy to see that this is a connected, symmetric, aperiodic Markov chain on $\Sigma_{\mathbf{rc}}$. It converges to the uniform distribution. A formal proof and discussion of rates of convergence appears in Section 10.2 below.

As an example, consider Table 8.1 of Section 8. This had a chi-square statistic of 72.18. A random walk was run on the tables with the same row and column sums in each run, the walk was run 49,936 as an initial randomization. Then, the walk was run 2,297,000 steps. For each step, the chi-square statistic was computed. A 1 is recorded if this is smaller than 72.18. The number of 1's/2,297,000 gives an estimate of the proportion of tables with chi-square smaller than 72.18. The whole procedure was repeated 5 times. The median of these five values is .7638.

In this example, an exhaustive enumeration gave the exact proportion as .76086 (to 5 significant figures) of the 239,382,173 tables. Thus, the Monte Carlo gives accurate answers for this example. The entire procedure took 23 minutes and 31 seconds. This is about 1/30 the time required for the exhaustive method. Figure 8.1 in Section 8 shows that the random walk gives a remarkably accurate approximation to the true distribution.

A history, many variations, and extensions of this algorithm are described by Diaconis and Sturmfels (1993). They also describe other problems (e.g., three-dimensional arrays) where similar algorithms are available.

10.2. From sampling to counting. This section gives a brief synopsis of work of Jerrum, Valiant, and Vazirani (1986). Sinclair (1993) gives a more complete short expository account.

If V is a finite set, H a subset and we can efficiently choose randomly from the uniform distribution over V, then we can estimate $|H|/|V|$ by seeing what proportion of samples fall into H. If H is small enough to be enumerated but large enough that $|H|/|V|$ is "not too small" (so the ratio can be reliably estimated) then the ingredients can be put together to give an estimate of $|V|$. In practice, a nested decreasing subsequence $V \supset H_1 \supset H_2 \cdots \supset H_r$ is used to meet the goals.

To make this rigorous, we will use the notation of Section 9. The definitions are a bit abstract; they are immediately followed by an example. Let Σ be a finite alphabet, Σ^* the finite sequences (or words) with each term in Σ. A relation $R \subset \Sigma^* \times \Sigma^*$ is *polynomially self-reducible* if

- There is a deterministic polynomial-time computable function $g : \Sigma^* \to \mathbb{N}$ such that if $R(x,y)$ then $|y| = g(x)$.
- There exist polynomial-time computable functions $\psi : \Sigma^* \times \Sigma^* \to \Sigma^*$ and $\sigma : \Sigma^* \to N$ such that for all $x, w \in \Sigma^*$

  $\sigma(x) \le c \log |x|$ for some constant c,
  $g(x) > 0$ implies $\sigma(x) > 0$,
  $|\psi(x,w)| \le |x|$,

  and such that for all $x \in \Sigma^*$, if $y = wz$ with $|y| = g(x)$ and $|w| = \sigma(x)$ then $R(x, wz)$ if and only if $R(\psi(x,w), z)$.

These conditions say that $R(x, wz)$ can be determined by first computing $\psi(x,w)$ and then determining $R(\psi(x,w), z)$. Since $|w| = \sigma(x) = O(\log|x|)$, the entire solution set $R(x)$ can be expressed as the disjoint union of a polynomial number of solution sets of the same relation on smaller instances.

Example 10.1. (*Spanning Trees*). We show here how the spanning trees in a graph can be coded in the language of this section. Let x represent a graph with n vertices. We suppose x is represented by its adjacency list which is indexed by $\ell = \lceil \log n \rceil$ bit integers. Spanning trees y can be represented as lists of $n-1$ edges, each edge being a pair of vertex indices. Thus $\Sigma = \{0,1\}$, and $R(x,y)$ if y is a spanning tree of x. Here, each y has $|y| = g(x) = 2\ell(n-1)$. Let $\sigma(x) = 2\ell$ with the first $\sigma(x)$ characters of y representing the first edge in the list. For y of length $g(x)$, write $y = wz$, where $|w| = \sigma(x)$ and w represents one edge. Let $\psi(z,w)$ be the result of contracting the edge w in x (merging the vertices at the ends of w and erasing any resulting multiple edges). This yields a smaller graph: $|\psi(z,w)| \le 2\ell(n-2) < |x|$. Note that $R(x, wz)$ if and only if $R(\psi(x,w), z)$.

That is, $y = wz$ is a spanning tree of x if and only if when we contract the edge w in x, z represents a spanning tree of the resulting $\psi(x,w)$.

Jerrum, Valiant, and Vazirani (1986) have given the following theorem which in essence says that for self reducible problems, efficient sampling yields efficient approximate counting. See Section 7 for the definition of polynomial relation.

THEOREM 10.2. (*Jerrum, Valiant, Vazirani*). *Let R be a self-reducible polynomial relation that is also in P. Suppose that we have an algorithm that*

- *takes input $x \in \Sigma^*$ and ϵ, $0 < \epsilon \le 1$*
- *runs in time polynomial in $|x|$ and $ln(1/\epsilon)$*
- *generates a random element $y \in R(x)$ whose distribution is within ratio $(1 \pm \epsilon)$ of the uniform distribution on $R(x)$.*

Then, given an x, ϵ, δ a random count c can be computed so that c approximates $|R(x)|$ within ratio $1 \pm \epsilon$, with probability at least $1 - \delta$. Moreover, this computation can be done in time polynomial in $|x|, 1/\epsilon$, and $\log(1/\delta)$.

Remark 10.3. Jerrum, Valiant and Vazirani (1986) show that the requirement that the relation R is in P can be dropped if a slightly different notion of near uniform sampling is used or if we work only with x for which $R(x)$ is nonempty.

We will now show that we can cast $\Sigma_{\mathbf{rc}}$ in a self-reducible form. For any table $F \in \Sigma_{\mathbf{rc}}$ and ordered pair $(k,\ell) \in [m] \times [n]$, let $[\Sigma_{\mathbf{rc}}]$ denote the set of all tables T in $\Sigma_{\mathbf{rc}}$ with

$$T_{ij} = F_{ij} \quad \text{and whenever} \quad i = k \quad \text{and} \quad j < \ell.$$

In other words, $[\Sigma_{\mathbf{rc}}|F;(k,\ell)]$ is the subset of $\Sigma_{\mathbf{rc}}$ tables whose entries match the table F in all positions *strictly* preceding (k,ℓ) in the lexicographic order. (We will use the symbols $\succ$ (and $\succeq$) for this order relation.) Notice that we have the following properties for any $F \in \Sigma_{\mathbf{rc}}$: $\Sigma_{\mathbf{rc}}$: (a) $F \in [\Sigma_{\mathbf{rc}}|F;(k,\ell)]$, (b) $[\Sigma_{\mathbf{rc}}|F;(1,1)] = \Sigma_{\mathbf{rc}}$, and (c) for any $(k,\ell) \succ (m-1, n-1)$ we have $[\Sigma_{\mathbf{rc}}|F;(k,\ell)] = \{F\}$, since all remaining entries are then determined by the sum constraints.

If we use $F_{k\ell \leftarrow i}$ to denote the table obtained from F by setting $F_{k\ell} = i$, then we may write

$$[\Sigma_{\mathbf{rc}}|F;(k,\ell)] = \bigcup_{0 \le i \le N} [\Sigma_{\mathbf{rc}}|F_{k\ell \leftarrow i}; succ(k,\ell)),$$

where $succ(k,\ell)$ is the ordered pair that is the immediate successor of (k,ℓ) in the lexicographic order. Note that some of the sets on the right may be empty. This relation expresses the decomposition needed to show that

$[\Sigma_{rc}|F;(k,\ell)]$ is polynomially self-reducible in the parameters in m, n, and N.

Now we will show that we can use a random walk to draw samples from $[\Sigma_{rc}|F;(k,\ell)]$ as we used to draw sample from Σ_{rc}. This walk is explained in section 10.1 above.

Algorithm 10.4. (*Random walk on* $[\Sigma_{rc}|F;(k,\ell)]$). For a given k and ℓ, modify the basic walk of Section 10.1 so that it chooses only amongst values of i_1, i_2, j_1 and j_2 such that $(i_1, j_1) \succeq (k, \ell)$. That is, at each stage uniformly choose a pair of rows i_1 and i_2, $i_1 < i_2 \leq m$, and a pair of columns j_1 and j_2,

$$j_1 < j_2 \leq n, \qquad \text{such that} \quad (i_1, j_1) \succeq (k, \ell).$$

THEOREM 10.5. *The random walk generated by the Algorithm above and started on any $F \in \Sigma_{rc}$ is ergodic and has uniform stationary distribution on $[\Sigma_{rc}|F;(k,\ell)]$.*

Proof. We need to show that for each X and Y in Σ_{rc}, there is a path between X and Y, using only possible steps of the walk. Note that since each step is reversible, a path from X to Y implies a symmetric one from Y to X.

We prove that there is a path joining X and Y by induction on a distance measure between the two tables X and Y. Define the distance $d(X,Y)$ between X and Y as $d(X,Y) = \sum_{i,j} |X_{ij} - Y_{ij}|$. Observe that $d(X,Y) = 0$ if and only if $X = Y$. Further note that, since the grand sums in the table are the same, this distance is always a multiple of two.

Let k be a nonnegative integer, and assume the following induction hypothesis: if $0 \leq d(X,Y) \leq 2k$ and if (i,j) is the lexicographically-first coordinate in which X and Y differ, then there is a path joining X and Y using only steps of the walk that do not involve coordinates lexicographically preceding (i,j). This is vacuously true for $k = 0$.

For the induction step, let X and Y be two elements of Σ_{rc} and suppose $d(X,Y) = 2(k+1)$, where (i,j) is the first coordinate in which they differ. We will show that either there is a move from X to X' where $d(X',Y) \leq 2k$ or there is a move from Y to Y' where $d(X,Y') \leq 2k$, where no coordinates preceding (i,j) are involved. The induction hypothesis will then imply that there is an entire path between X and Y.

In step 2 of the algorithm choose $i_1 = i$ and $j_1 = j$. Then there are two cases to consider.

Case 10.6. ($X_{i_1,j_1} < Y_{i_1,j_1}$). Then since each row and column of X has the same sum as in Y, we have

$$\exists j_2 \quad \text{such that} \quad X_{i_1,j_2} > Y_{i_1,j_2}$$
$$\exists i_2 \quad \text{such that} \quad X_{i_2,j_1} > Y_{i_2,j_1}.$$

We must have $i_1 < i_2$ and $j_1 < j_2$, since (i_1, j_1) was chosen as the lexicographically first position in which X and Y differ. Moreover, the entries X_{i_1,j_2} and X_{i_2,j_1} are both positive, since they are greater than their nonnegative counterparts in Y. This means, that letting $d = +1$, the move

$$X'_{i_1,j_1} = X_{i_1,j_1} + 1 \qquad X'_{i_1,j_2} = X_{i_1,j_2} - 1$$
$$X'_{i_2,j_1} = X_{i_2,j_1} - 1 \qquad X'_{i_2,j_2} = X_{i_2,j_2} + 1$$

yields an X' having nonnegative entries as well as sharing the same row and column sums as X.

By moving from X to X', the difference with respect to Y on least the three coordinates (i_1,j_1), (i_1,j_2), and (i_2,j_1) decreased by 1. The difference at (i_2,j_2) may have increased by 1, but the net change in all four coordinates must in any case be a decrease of at least 2. That is, $d(X',Y) \leq d(X,Y) - 2$, so $d(X',Y') \leq 2k$. Now by the induction hypothesis there is a path from X' to Y. Adding the step from X to X' completes the path from X to Y, without altering any coordinates lexicographically preceding (i,j) (in which X and Y already agree).

Case 10.7. ($X'_{i_1,j_1} > Y_{i_1,j_1}$). This case is entirely symmetric. Swapping the roles of X and Y, the same argument as in Case 10.6 shows that there is a move from Y to Y' with $d(X,Y') \leq 2k$. Thus, by the induction hypothesis, there is a path between X and Y', and hence a path between X and Y via Y'. □

Remark 10.8. 1. To have a provably polynomial randomized algorithm it must be shown that the random walk above is rapidly mixing. This has been done for fixed dimensional problems in Diaconis and Saloff-Coste (1994). Here is a statement of their result. Let $P(x,y)$ be the transition matrix of the random walk described above. Here $x, y \in \Sigma_{rc}$. Let U be the uniform distribution on Σ_{rc}. Let γ be the diameter of Σ_{rc}. This is the smallest n so $P^n(x,y) > 0$ for all $x, y \in \Sigma_{rc}$.

THEOREM 10.9. *There are constants a, b, α, β such that*

$$\|P_x^k - U\| \leq \alpha e^{-c} \quad \text{for} \quad k \geq ac\gamma^2 \quad \text{and all } x$$
$$\|P_x^k - U\| \geq \beta > 0 \quad \text{for} \quad k \leq b\gamma^2 \quad \text{for some } x.$$

Here α, β, a, b depend on the dimensions m, n. But not otherwise on $\mathbf{r}, \mathbf{c}$.

This gives a randomized algorithm that runs in a polynomial number of steps in N for fixed m, n.

2. Chung, Graham and Yau (1994) have announced much more sweeping results which are currently under close scrutiny. It seems likely that a definitive solution will be available.

3. In practice, one uses much more vigorous random walks: one need not

move one each time and more complex patterns than the basic $\begin{matrix} + & - \\ - & + \end{matrix}$ can be used. Diaconis and Sturmfels contains further discussion and examples.

10.3. Convex sets for random generation. There is a somewhat different approach to random generation (and hence counting) which links into the healthy developments of computer science theory. Consider the set of all $m \times n$ arrays with *real* nonnegative entries, row sums $\mathbf{r}$ and column sums $\mathbf{c}$. This is a convex polyhedron containing Σ_{rc}. There has been a great deal of work in the computer science literature on approximations to the volume of such convex polyhedra. Briefly, such problems are #-P complete but have polynomial time approximations using randomness. The already large literature on this subset is surveyed in Dyer and Frieze (1991). See Lovasz and Shimonovitz (1990) for recent results. It seems natural to try and adapt the ideas developed for the volume problem to the problem of counting Σ_{rc}. Gangolli (1991) made an early application of these ideas to table enumeration. He needed to assume the row and column sums were fairly balanced.

Dyer, Kannan and Mount (1994) have recently made a breakthrough in this problem. Briefly, consider Σ_{rc} as a subset of lattice points in $(m-1) \times (n-1)$ dimensions. About each table, construct a parallelopiped with sides aligned to the lattice. Each such box is identified with a unique table. Let Σ be the convex hull of the union of these boxes. Points can be picked uniformly at random in Σ to good approximation using random walk. If the point is in one of the boxes, output the associated table as the choice. If the point is at the fringes, repeat.

A formal theorem requires a mild restriction on the row and column sums. Here is a simplified version of their results: suppose $r_i > n(n-1)(m-1)$ for each i and $c_j > m(n-1)(n-1)$ for each j. Then there is an algorithm for generating a random table with a distribution within ϵ of uniform invariation distance which runs in time bounded by a polynomial in $I, J, \max_{ij}(\log r_i, \log c_j)$ and $\log(1/\epsilon)$.

For m and n small, the restrictions on r_i and c_j allow tables of practical interest, for example, with $m = n = 4$. The restrictions are $r_i, c_j \geq 36$. The above is a simplified version of their work which actually is more general. The authors have done much more: they have implemented their algorithm and have it up and running. At present it produces about 1 table per second for tables like Table 6.1. The existence of an independent check of this sort has already been highly useful. It allowed us to pick up clear errors in some other algorithms.

Note Added in Proof. John Mount (1994) has made spectacular progress in deriving closed form expression for the number of tables. R.B. Holmes and L.K. Jones (1994), " *On the Uniform Generation of Two-Way tables with fixed margins and conditional volume test of Diaconis and Efron*", Technical Report, Dept of Mathematics, U-Mass-Lowell, offer a clever rejection algorithm for Monte Carlo sampling. Rates of convergence for the random walk of Section 10.2 are in Diaconis, P. and Saloff-Coste L., (1994), Random Walk on Contingency Tables with Fixed Row and Column Sums, Technical Report, Dept. of Mathematics, Harvard University.

Acknowledgement. We thank Nantel Bergeron, Fan Chung, Ira Gessel, I.J. Good, Ron Graham, Susan Holmes, Ravi Kannan, Brad Mann, John Mount, Richard Stanley, John Stembridge, and Bernd Sturmfels for their help with this project.

REFERENCES

[1] Baglivo, J. (1994) Forthcoming Book.
[2] Balmer, D. (1988) Recursive enumeration of $r \times c$ tables for exact likelihood evaluation, *Applied Statistics* **37**, 290-301.
[3] Békéssy, A., Békéssy, P., Komlós, J. (1972) Asymptotic enumeration of regular matrices, *Studia Sci. Math. Hung.* **7**, 343-353.
[4] Bender, E. (1974) The asymptotic number of non-negative integer matrices, with given row and column sums, *Discrete Math.* **10**, 217-223.
[5] Bishop, Y., Fienber, S., and Holland, P. (1975) *Discrete Multivariate Analysis*, MIT Press, Cambridge, MA.
[6] Bona, M. (1994) Bounds on the volume of the set of doubly stochastic matrices, Technical Report, Dept. of Mathematics, M.I.T.
[7] Boulton, D., and Wallace, C. (1973) Occupancy of a regular array, *Computing* **16**, 57-63.
[8] Chung, F., Graham, R., and Yau, S. (1994) On sampling in combinatorial structures, unpublished manuscript.
[9] Cormen, T., Leiserson, C., and Rivest, R. (1990) *Introduction to algorithms*, McGraw-Hill, New York.
[10] Curtis, C., and Reiner, I. (1962) *Representation Theory of Finite Groups and Associative Algebras*, Wiley Interscience, New York.
[11] Dahmen, W., and Micchelli, C. (1988) The number of solutions to linear Diophantine equations and multivariate splines, *Trans. Amer. Math. Soc.* **308**, 509-532.
[12] Danilov, V. (1978) The geometry of toric varieties, Russian Math. Surveys **33**, 97-154.
[13] Diaconis, P., and Efron, B. (1985) Testing for independence in a two-way table: new interpretations of the chi-square statistic (with discussion), *Ann. Statist.* **13**, 845-913.
[14] Diaconis, P. (1988) *Group representations in probability and statistics*, IMS Lecture Notes – Monograph Series, II. Institute of Mathematical Statistics, Hayward, CA.
[15] Diaconis, P. (1989) Spectral analysis for ranked data, *Ann. Statist.* **17**, 781-809.
[16] Diaconis, P., McGrath, M., and Pitman, J. (1993) Descents and random permutations, to appear, *Combinatorica*.
[17] Diaconis, P., and Saloff-Coste, L. (1993) Nash inequalities for Markov chains, Technical Report, Dept. of Statistics, Harvard University.
[18] Diaconis, P., and Sturmfels, B. (1993) Algebraic algorithms for sampling from conditional distributions, Technical Report, Dept. of Statistics, Stanford University.
[19] Dyer, A., and Frieze, A. (1991) Computing the volume of convex bodies: a case where randomness provably helps, in B. Bollobás, *Probabilistic Combinatorics and Its Applications*, Amer. Math. Soc., Providence, RI.
[20] Dyer, M., Kannan, R., and Mount, J. (1994) Unpublished manuscript.

[21] Everett, C., and Stein, P. (1971) The asymptotic number of integer stochastic matrices, *Discrete Math.* **1**, 55-72.

[22] Fisher, R.A. (1935) *The Design of Experiments*, Oliver and Boyd, London.

[23] Foulkes, H. (1976) Enumeration of permutations with prescribed up-down and inversion sequences, *Discrete Math.* **15**, 235-252.

[24] Foulkes, H. (1980) Eulerian numbers, Newcomb's problem and representations of symmetric groups, *Discrete Math.* **30**, 3-99.

[25] Fulton, W. (1993) *Introduction to Toric Varieties*, Princeton University Press, Princeton, NJ.

[26] Gail, M., and Mantel, N. (1977) Counting the number of $r \times c$ contingency tables with fixed margine, *Jour. Amer. Statist. Assoc.* **72**, 859-862.

[27] Gangolli, A. (1991) Convergence bounds for Markov chains and applications to sampling, Ph.D. Thesis, Computer Science Dept., Stanford University.

[28] Garey, M., and Johnson, D. (1979) *Computers and Intractability*, Freeman, San Francisco.

[29] Garsia, A. (1990) Combinatorics of the free Lie algebra and the symmetric group, *Analysis and etc.* (P.H. Rabinowitz and E. Zehnder, eds.), pp. 309-382. Academic Press, Boston.

[30] Garsia, A., and Reutenauer, C. (1989) A decomposition of Soloman's descent algebra, *Adv. Math.* **77**, 189-262.

[31] Gessel, I. (1990) Symmetric functions and P-recursiveness, *Jour. Comb. Th.* **A 53**, 257-285.

[32] Gessel, I., and Reutenauer, C. (1993) Counting permutations with given cycle structure and descent set, *Jour. Comb. Th.* **A 56**, 189-215.

[33] Good, I.J. (1976) On the application of symmetric Dirichlet distributions and their mixtures to contingency tables, *Ann. Statist.* **4**, 1159-1189.

[34] Good, I.J., and Crook, J. (1977) The enumeration of arrays and a generalization related to contingency tables, *Discrete Math.* **19**, 23-65.

[35] Goulden, I., Jackson, D. and Reilly, J. (1983) The Hammond series of a symmetric function and its application to P-recursiveness, *SIAM Jour., Alg. Discrete Methods* **4**, 179-183.

[36] Gusfield, D. (1988) A graph-theoretic approach to statistical data security, *SIAM Jour. Comp.* **17**, 552-571.

[37] Hancock, J. (1974) Remark on Algorithm 434, *Communications of the ACM* **18**, 117-119.

[38] Humphreys, J. (1990) *Reflection Groups and Coxeter Groups*, Cambridge University Press, Cambridge.

[39] Irving, R., and Jerrum, M. (1990) 3-D statistical data security problems, Technical Report, Dept. of Computing Science, University of Glasgow.

[40] Jackson, D., and Van Rees (1975) The enumeration of generalized double stochastic non-negative integer square matrices, *SIAM Jour. Comp.* **4**, 475-477.

[41] James, G.D., and Kerber, A. (1981) *The representation theory of the symmetric group*, Addison-Wesley, Reading, MA.

[42] Jerrum, M., and Sinclair, A. (1989) Approximating the permanent, *SIAM Jour. Comp.* **18**, 1149-1178.

[43] Jerrum, M., Valiant, L., and Vazirani, V. (1986) Random generation of combinatorial structures from a uniform distribution, *Theoretical Computer Science* **43**, 169-180.

[44] Jia, R. (1994) Symmetric magic squares and multivariate splines, Technical Report, Dept. of Mathematics, University of Alberta.

[45] Kerov, S.V., Kirillov, A.N., and Reshetikin, N.Y. (1986) Combinatorics, Bethe Ansatz, and representations of the symmetric group, translated from: *Zapiski Mauchnykh Seminarov Leningradskogo Otdeleniya Matematicheskogo Instituta im. V.A. Steklova An SSSR* **155**, 50-64.

[46] Knuth, D. (1970) Permutations, matrices, and generalized Young tableaux, *Pacific Jour.* **34**, 709-727.

[47] Knuth, D. (1973) *The Art of Computer Programming*. Vol. 3. Addison-Wesley. Reading, MA.

[48] Lovasz, L., and Simonovits, M. (1990). The mixing rate of Markov chains, an isoperimetric inequality, and computing the volume. Preprint 27, Hung. Acad. Sci.

[49] MacMahon, P. (1916) *Combinatorial Analysis*, Cambridge University Press, Cambridge.

[50] Macdonald, J. (1979) *Symmetric Functions and Hall Polynomial*, Clarendon, Press, Oxford.

[51] Mann, B. (1994) Unpublished manuscript.

[52] March, D. (1972) Exact probabilities for $r \times c$ contingency tables, *Communications of the ACM* **15**, 991-992.

[53] Mehta, C., and Patel, N. (1983) A network algorithm for performing Fisher's exact test in $r \times c$ contingency tables, *Jour. Amer. Statist. Assoc.* **78**, 427-434.

[54] Mehta, C., and Patel, N. (1992) Statexact.

[55] Mount, J. (1994) Ph.D. Thesis, Dept. of Computer Science, Carnegie Mellon University.

[56] O'Neil, P. (1969) Asymptotics and random matrices with row-sum and column-sum restrictions, *Bull. Amer. Math. Soc.* **75**, 1276-1282.

[57] Pagano, M., and Taylor-Halvorsen, K. (1981) An algorithm for finding the exact significance levels of $r \times c$ contingency tables, *Jour. Amer. Statist. Assoc.* **76**, 931-934.

[58] Sinclair, A. (1993) *Algorithms for random generalization and counting: a Markov chain approach*, Birkhäuser, Boston.

[59] Sinclair, A., and Jerrum, M. (1989) Approximate counting, uniform generation and rapidly mixing Markov chains, *Information and Computation* **82**, 93-133.

[60] Snee, R. (1974) Graphical display of two-way contingency tables, *Amer. Statistician* **38**, 9-12.

[61] Soloman, L. (1976) A Mackey formula in the group ring of a Coxeter group, *J. Alg.* **41**, 255-264.

[62] Stanley, R. (1973) Linear homogeneous Diophantine equations and magic labelings of graphs, *Duke Math. Jour.* **40**, 607-632.

[63] Stanley, R. (1986) *Enumerative Combinatorics*, Wadsworth, Monterey, CA.

[64] Stein, M.L., and Stein, P. (1970) Enumeration of stochastic matrices with integer elements, Los Alamos Scientific Laboratory Report, LA-4434.

[65] Valiant, L. (1979) The complexity of computing the permanent, *Theoretical Computer Sci.* **8**, 189-201.