Adv. Appl. Prob. **24**, 761–777 (1992)
Printed in N. Ireland
© *Applied Probability Trust* 1992

INTERNATIONAL CONFERENCE ON RANDOM MAPPINGS, PARTITIONS, AND PERMUTATIONS

A Conference Sponsored by the US Army Research Office, the National Science Foundation, the National Security Agency, the Office of Naval Research, and the University of Southern California at the University of Southern California, Los Angeles, 3–6 January 1992

CONTENTS

Introduction

Since the pioneering work of Erdős, Goncharov, Rényi and Turán, the application of probabilistic methods to combinatorial enumeration has grown into an independent branch of mathematics. There has been a remarkable resurgence of interest in classical combinatorial objects, among them random mappings, partitions and permutations. These structures are proving to be an inexhaustible source of challenging problems, of both theoretical and practical importance. The conference was intended to bring together leading researchers to discuss and synthesize a variety of disparate approaches and techniques.

The organizing committee was comprised of R. Arratia, S. W. Golomb, and S. Tavaré from USC, G.-C. Rota from MIT, and B. Harris from Madison. There were 18 invited lectures, each of 40 minutes duration. In addition, there were 15 contributed lectures, each lasting 20 minutes. Professor V. F. Kolchin of the Steklov Mathematical Institute gave the keynote address. We were also delighted to have a special lecture from Professor P. Erdős. The speakers' names and abstracts are listed alphabetically.

There were over 80 participants at the conference, including faculty and students from the University of Southern California. Five mathematicians from Russia spoke at the meeting, as well as two from England, and one each from Australia, Germany, Hungary, Poland, Sweden, and Switzerland. Up-to-date accounts were given of many major trends in research into the asymptotics of probabilistic enumeration.

The conference was supported by Army Research Office grant DAAL 03-91-G-0322, National Security Agency grant MSP-003-91, National Science Foundation grant DMS 91-16083, Office of Naval Research Grant N00014-92-J-1473, and the University of Southern California.

University of Southern California
Los Angeles

S. TAVARÉ

INVITED PAPERS

D. J. ALDOUS, *University of California, Berkeley*

Brownian bridge asymptotics for random mappings

Write $\boldsymbol{n} = \{1, 2, \cdots, n\}$. A function $f : \boldsymbol{n} \to \boldsymbol{n}$ may be regarded as a directed graph with edges $i \to f(i)$; note this allows an edge $i \to i$. By a *random mapping* F_n we mean a uniform random choice of f from all n^n functions $\boldsymbol{n} \to \boldsymbol{n}$. There is a large literature on combinatorial analysis of random mappings, much due to the Soviet school. Their results up to the early 1980s can be found summarized in Kolchin (1986). Three distinct methods have classically been used for proving $n \to \infty$ asymptotics for random mappings:

- Take limits in exact formulas.
- Generating function methods: see for example Flajolet and Odlyzko (1990).
- Representing certain quantities as i.i.d. random variables conditioned on their sum: see Kolchin (1986).

More recently Stein's method has been used to bound the errors in certain asymptotic approximations.

The purpose of this talk is to present a new method. We show how a mapping can be coded as a walk (with steps ± 1) of length $2n$. Our main result is that the random walk coded from the random mapping can be rescaled to converge as $n \to \infty$ to reflecting Brownian bridge. This one result encompasses many asymptotic results for particular statistics which have previously been treated separately—loosely, it gives limit distributions for all 'global' functionals of random mappings. Of course, the limit distribution is given in terms of a corresponding functional of reflecting Brownian bridge, which requires some calculation to evaluate explicitly. Fortunately most distributions of interest have already been discussed in the theoretical stochastic processes literature, or can be derived by known methods. This program parallels that of Aldous (1991) in which distributions associated with random trees are derived by coding trees as walks converging to Brownian excursion.

The exact way in which reflecting Brownian bridge approximates a random mapping is best described in pictures, but here is one aspect. Let $[G_1, D_1]$ be the excursion of Brownian bridge containing a uniform random time. Then we can decompose the Brownian bridge into three processes defined on $[0, G_1]$, $[G_1, D_1]$ and $[D_1, 1]$, and these three processes are rescaled versions of Brownian bridge, Brownian excursion and Brownian bridge respectively. If we take a random mapping and see where vertex 1 is, we can split the graph into three parts: (a) the tree-component containing 1, (b) the rest of the graph-component containing 1, and (c) the rest of the graph.

In our coded walk these parts appear in order (b, a, c), and approximate the tripartite decomposition of reflecting Brownian bridge described above. This is joint work with Jim Pitman.

ALDOUS, D. J. (1991) The continuum random tree II: an overview. In *Proc. Durham Symp. Stochastic Analysis* 1990, ed. M. T. Barlow and N. H. Bingham, Pp. 23–70, Cambridge University Press.

FLAJOLET, P. AND ODLYZKO, A. (1990) Random mapping statistics. In *Proc. Eurocrypt '89*, ed. J.-J. Quisquater, pp. 329–354, Lecture Notes in Computing Science **434**, Springer-Verlag, Berlin.

KOLCHIN, V. F. (1986) *Random Mappings*. Optimization Software, New York. (Translation of Russian original.)

R. A. ARRATIA, *University of Southern California*

Independent process approximations for combinatorial structures

Many random combinatorial objects have a structure whose joint distribution is exactly equal to that of a process of mutually independent random variables, conditioned on the value of a weighted sum of these variables. It is interesting to compare the combinatorial structure directly to the independent discrete process, *without* renormalizing. The quality of approximation can often be conveniently quantified in terms of total variation distance.

In detail, consider the component structure $C(n) = (C_1(n), C_2(n), \cdots, C_n(n))$, where C_i represents the number of parts of size i. The random variables $C_1, C_2, \cdots, C_n$ are mutually dependent, since the weighted sum $C_1 + 2C_2 + \cdots + nC_n$ has the constant value n. For a given family of combinatorial objects, indexed by $n = 1, 2, \cdots$, and for each value of a real parameter $x > 0$, there are mutually independent random variables $Z_1, Z_2, \cdots$, with the following property. Let T_n be the weighted sum $T_n = Z_1 + 2Z_2 + \cdots + nZ_n$. For each n, the joint distribution of the combinatorial process $C(n)$ is equal to the joint distribution of $(Z_1, \cdots, Z_n)$, conditioned on the event $\{T_n = n\}$. The simple independent process $(Z_1, \cdots, Z_n)$, without conditioning on the value of T_n, may directly provide useful approximations to the distribution of $C(n)$.

To describe the independent random variables Z_i, let m_i be the number of possible structures available for each part of size i. For the class of combinatorial assemblies, which includes permutations, with $m_i = (i-1)!$, mapping functions, with $m_i = (i-1)!(1 + i + i^2/2 + \cdots + i^{i-1}/(i-1)!)$, and partitions of a set, with $m_i = 1$, we have that Z_i is Poisson with parameter $\lambda_i = m_i x^i/i!$, $x > 0$. For the class of multisets, which includes partitions of an integer, with $m_i = 1$, polynomials of degree n, with $m_i =$ the number of monic irreducible polynomials of degree i, and random mapping patterns, Z_i is negative binomial, corresponding to the sum of m_i independent geometric random variables Y with $P(Y = k) = (1 - x^i)(x^i)^k$ for $k \geqq 0$, $0 < x < 1$. For the class of selections, which includes partitions of an integer into distinct parts, and square free polynomials, Z_i is binomial with parameters m_i and $x^i/(1 + x^i)$, $x > 0$.

An appropriate choice of the parameter x corresponds roughly to maximizing $P(T_n = n)$. In some cases, a constant gives an appropriate choice of x; examples include $x = 1$ for random permutations, $x = e^{-1}$ for random mapping functions, $x = \rho \equiv 0{\cdot}3383 \cdots$ for random mapping patterns, and $x = q^{-1}$ for random polynomials over a field with q elements. In such cases, $C(n)$, viewed as an element of $\mathbb{R}^\infty$, converges in distribution to $(Z_1, Z_2, \cdots)$. In other examples an appropriate choice of x must vary with n; examples include $x =$ the solution of $xe^x = n$ for random partitions of the set $\{1, 2, \cdots, n\}$, $x = \exp(-\pi/\sqrt{6n})$ for random partitions of the integer n and $x = \exp(-\pi/\sqrt{12n})$ for partitions with no repeated parts.

For most examples, with an appropriate choice of x, for large n and individually for each $i = 1$ to n, Z_i is a good approximation for $C_i(n)$. More generally, for $B \subset \{1, \cdots, n\}$, the joint distribution of the independent process $(Z_i)_{i \in B}$ is a good approximation for the joint distribution of the process $(C_i)_{i \in B}$, provided that B is small in the sense that the contributions to the mean and variance of T_n from terms indexed by B are small compared to the mean and variance of T_n. This approximation can be quantified conveniently by the total variation metric, and allows effective approximation of the distribution of some functionals of the entire process $C(n)$ by the distribution of the same functional applied to the independent process $(Z_1, \cdots, Z_n)$. Clearly, not all functionals are approximated well in distribution, the extreme example being the indicator functional $h(a_1, \cdots, a_n) = 1\{a_1 + 2a_2 + \cdots + na_n = n\}$, since in all non-trivial examples $Eh(Z_1, \cdots, Z_n) = P(T_n = n) \to 0$.

We consider issues common to all the above examples, including equalities and upper bounds for total variation distances, heuristics for good approximations, the relation to standard generating functions, formulas for moments, refinement to the process which counts the number of parts of each possible type, the effect of conditioning on further restrictions, large deviation theory and non-uniform measures on combinatorial objects, and the possibility of getting useful upper bounds for the probability of unlikely events by simply giving a lower bound on $P(T_n = n)$. Detailed examples, which show the utility and tractability of these approximations of combinatorial processes by independent processes, will be published elsewhere. This is joint work with Simon Tavaré.

A. D. BARBOUR, *University of Zürich*

Refined approximations for the Ewens sampling formula

The Ewens sampling formula is a family of probability distributions over the space of cycle types of permutations of n objects, indexed by a real parameter θ. In the case $\theta = 1$, where the distribution reduces to that induced by the uniform distribution on all permutations, the joint distribution of the number of cycles of lengths less than $b = o(n)$ is extremely well approximated by a product of Poisson distributions, having mean $1/j$ for cycle length j: the error is super-exponentially small with nb^{-1}. For $\theta \neq 1$, the analogous approximation, with means adjusted to θ/j, is good, but with error only linear in $n^{-1}b$. In this paper, it is shown that, by choosing the means of the Poisson distributions more carefully, an error quadratic in $n^{-1}b$ can be achieved, and that essentially nothing better is possible.

B. BOLLOBÁS, *University of Cambridge*

Random partial orders: concentration of the height

The systematic study of random d-dimensional partial orders was started by Winkler in 1985, although random permutations corresponding to random two-dimensional partial orders were first investigated by Ulam in 1961. To get a random d-dimensional partial order $<$ on $[n] = \{1, 2, \cdots, n\}$, take d random linear orders (permutations) on $[n]$, say $<_1, <_2, \cdots, <_d$ and set $x < y$ iff $x <_i y$ for every i. Equivalently, pick n points at random from the d-dimensional cube $[0, 1]^d$, and take the coordinatewise partial order on these points.

Denote by $L_{n,d}$ the height of a random d-dimensional partial order. In 1988 it was proved by Winkler and the author that for every fixed $d \geqq 2$, there is a constant $c_d > 0$ such that $n^{-1/d}L_{n,d}$ tends to c_d in probability. In fact, for $d = 2$ much stronger results had been known; in 1977 Logan and Shepp proved that $c_2 \geqq 2$ and, independently, Vershik and Kerov showed that $c_2 = 2$.

Recently Frieze investigated the concentration of $L_{n,2}$ about its mean, showing that, for $\varepsilon > 0$ there is some constant $\beta > 0$ such that $P(|L_{n,2} - EL_{n,2}| \geqq n^{\frac{1}{3}+\varepsilon}) \leqq \exp(-n^{\beta})$. We present some similar concentration results for $L_{n,d}$, for all $d \geqq 2$, proved jointly with Graham Brightwell. In the case $d = 2$, our theorem shows that $n^{\frac{1}{3}+\varepsilon}$ above can be replaced by $n^{\frac{1}{4}+\varepsilon}$, which we believe to be essentially best possible.

The proof is based on the Hoeffding–Azuma martingale inequality, applied to a variant of the height of a random partial order obtained from a Poisson process in the cube $[0, 1]^n$.

Bollobás, B. and Winkler, P. M. (1988). The longest chain among random points in Euclidean space. *Proc. Amer. Math. Soc.* **103,** 347–353.

Frieze, A. M. (1991) On the length of the longest monotone subsequence in random permutation. *Ann. Appl. Prob.* **1,** 301–305.

Logan, B. F. and Shepp, L. A. (1977) A variational problem for Young tableaux. *Adv. Math.* **26,** 206–222.

Vershik, A. M. and Kerov, S. V. (1977) Asymptotics of the Plancherel measure of the symmetric group and the limiting form of Young tableaux. *Dokl. Akad Nauk. SSSR* **233,** 1024–1028.

Winkler, P. M. (1985) Random orders. *Order* **1,** 317–325.

P. DIACONIS, *Harvard University*

Comparison techniques for card shuffling

In joint work with Laurent Saloff-Coste a new set of techniques has emerged for bounding the rate of convergence of a reversible Markov chain to its stationary distribution. The techniques are here illustrated for a variety of 'shuffles': random walks on the permutation group. Given a symmetric set of permutations, create a random walk by repeatedly picking from this set (with replacement) and multiplying. Under mild restrictions, this process converges to the uniform distribution. The present techniques give sharp bounds on the rate of convergence in total variation in terms of 'geometric' properties of the set. For example, if the set is $\{\mathrm{id}, (1,2), (n, n-1, \cdots, 1), (1, 2, \cdots, n)\}$ and $k = 36n^3(\log n + c)$ then there are universal constants $\alpha, \beta > 0$ such that $\|P^{*k} - U\| \leqq \alpha e^{-\beta c}$, for $c > 0$.

As a second example, let a connected graph on n vertices have edge set E. Form a random walk on permutations by replacing a card at each vertex of the graph and at successive times choosing an edge at random, and switching the cards on that edge. The methods show that the bound displayed above holds, provided $k = n^{-1}|E|\,\gamma_* b_*(\log n + c)$. Here γ_* is the length of the longest path in the graph, $b_* = \max_e \sum_{\gamma_{xy} \ni e} 1$ where paths γ_{xy} have been chosen as in Diaconis and Stroock (1991). Examples show these bounds are sharp.

The techniques involve a new set of comparison inequalities which bound the eigenvalues of the chain of interest via comparison with eigenvalues of a known chain.

DIACONIS, P. AND STROOCK, D. (1991) Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Prob.* **1,** 36–61.

DIACONIS, P. AND SALOFF-COSTE, L. (1991) Comparison theorems for random walks on groups. Technical report, Department of Mathematics, Harvard University.

P. J. DONNELLY, *Queen Mary and Westfield College, London*

Labellings, size-biased permutations and the GEM distribution

In proving limit theorems for the 'sizes' of 'components' of combinatorial objects there are usually several ways of labelling the components. One labelling is by decreasing size order, another is a particular random labelling called a size-biased permutation. Continuity results usually guarantee that convergence with one labelling is equivalent to convergence with the other. In many cases (random permutations, random mappings, population genetics, prime divisors) normalised sizes converge to the Poisson–Dirichlet distribution with the ordered labelling and to the GEM distribution with the size-biased labelling. Apart from its inherent interest and natural interpretation in some settings, use of the size-biased permutation often greatly facilitates the proof of convergence results. Furthermore, in stark contrast to the Poisson–Dirichlet, the GEM distribution which arises in the limit is extremely tractable. Amongst residual allocation models the GEM distribution is characterized by invariance under size-biased permutation. This characterization is equivalent to a characterization of the Ewens sampling formula (which describes a sample from GEM or Poisson–Dirichlet populations) as the only partition structure enjoying a certain 'non-interference' property. More generally, invariance under size-biased permutation is equivalent to invariance under the so-called heaps mechanism. We conjecture various optimality properties of size-biased permutations.

P. ERDŐS, *Hungarian Academy of Sciences*

Recent problems in probabilistic number theory and combinatorics

The following problems and results were discussed.

Tetali and I (1990) proved that there exists a sequence $a_1 < a_2 < \cdots$ such that if $f(n)$ denotes the number of solutions of

$$n = \sum_i \varepsilon_i a_i, \qquad \varepsilon_i = 0 \text{ or } 1, \qquad \sum \varepsilon_i \leqq k$$

then

$$\limsup \frac{f(n)}{\log n} = c_1 < \infty, \qquad \liminf \frac{f(n)}{\log n} = c_2 > 0.$$

I proved this for $k = 2$ long ago, and Wirsing also proved it for $k = 3$. The general case presents difficulties.

Let us now restrict ourselves to $k = 2$. Turán and I conjectured 50 years ago that if $f(n) > 0$ for every $n > n_0$ then $\limsup f(n) = \infty$, and perhaps even $\limsup f(n)/\log n > 0$. I offer \$500 for a proof or disproof. I also conjectured that

$$\lim \frac{f(n)}{\log n} = c, \; 0 < c < \infty$$

is impossible. I offer \$500 for a proof or disproof of this also.

Nathanson constructed for every k a minimal asymptotic basis of order k. The proof in Erdős and Tetali (1990) gives a thin minimal asymptotic basis. Several other problems were discussed.

Erdős, P. and Tetali, P. (1990) Representations of integers as the sum of k terms. *Random Struct. Alg.* **1,** 245–262.

L. H. HARPER, *University of California, Riverside*

In search of maximum antichains of partitions

An antichain in a poset P is a subset of P having no comparable members. In 1928 Sperner showed that the largest antichain in the set of all subsets of an n-set ordered by containment, is its largest rank.

Around 1967 Rota asked if the partitions of an n-set, ordered by refinement, have the Sperner property, i.e. is the largest antichain the largest rank? Erdős has asked the same question about the partitions of n.

In 1978 Canfield showed that the answer to Rota's question is no; that for n sufficiently large ($n > 6 \times 10^{24}$) antichains larger than any rank exist. Subsequent papers by Shearer and Kleitman lowered the upper bound on the Canfield number to 6×10^6, but did not give any lower bounds on it nor ascertain whether there were antichains significantly larger than the largest rank.

In 1985 the present author:

(i) Showed how to approximate the poset of partitions of an n-set by a Gaussian process ordered by a cone.
(ii) Solved the finite-dimensional analog of the Sperner problem in (i);
(iii) Carried out calculations, based on the assumption that the solution of the Sperner problem is preserved by the limiting process of (i), which show that the Canfield number is about 6×10^6 and that the ratio of the largest antichain to the largest rank converges to 1·69.

Recently J. Chavez and I have been investigating whether the same technique can answer Erdős's question. Our latest results were presented, indicating that the poset of partitions of n is asymptotically Sperner.

Canfield, E. R. (1978) On a problem of Rota. *Adv. Math.* **29,** 1–10.
Harper, L. H. (1985) On a continuous analog of Sperner's problem. *Pacific J. Math.* **118,** 411–425.
Rota, G.-C. (1967) Research problem: A generalization of Sperner's theorem. *J. Combinatorial Theory* **2,** 104.
Shearer, J. B. (1979) A simple counterexample to a conjecture of Rota. *Discrete Math.* **28,** 327–330.
Sperner, E. (1928) Ein Satz uber Untermengen einer endlichen Menge. *Math Z.* **27,** 544–548.

B. HARRIS, *University of Wisconsin-Madison*

The early history of the theory of random mappings

Much of the early work in the theory of random mappings was reviewed. This includes the contributions of Ulam, Katz, Riordan and the author. In particular, work by Golomb, Rubin and Sitgreaves, Folkert and Lenard was discussed. These works are of particular interest, because they have not been published.

RUBIN, H. AND SITGREAVES, R. (1954) Probability distributions related to random transformation of a finite set. Stanford University Technical Report, 19 A.

HARRIS, B. (1960) Probability distributions related to random mappings. *Ann. Math. Statist.* **31,** 1045–1062.

V. F. KOLCHIN, *Steklov Mathematical Institute, Moscow*

Cycles in random graphs and hypergraphs

For a $T \times n$ matrix $A = \|a_{ij}\|$ in $GF(2)$ we define a hypergraph G_A with n vertices and T hyperedges $e_t = \{j : a_{tj} = 1\}$, $t = 1, \ldots, T$. Denote $a_t = (a_{t1}, \cdots, a_{tn})$, $t = 1, \cdots, T$. A set of row numbers $\{t_1, \cdots, t_n\}$ is called a critical set if the sum of vectors $a_{t_1} + \cdots + a_{t_m}$ is the zero vector. We can naturally define the concept of independence for critical sets and determine the maximal number $s(A)$ of independent critical sets in A. The total number of critical sets $S(A)$ is equal to $2^{s(A)} - 1$. It is not difficult to see that $s(A)$ and the rank $r(A)$ of the matrix A are connected by the equality $r(A) + s(A) = T$. Therefore we can use the more tractable characteristic $s(A)$ for the investigations of rank of the matrix A.

We consider a matrix A of a special form which corresponds to the following system of T random equations in $GF(2)$:

$$x_{i_1(t)} + \cdots + x_{i_r(t)} = b_t, \qquad t = 1, \cdots, T,$$

where $i_1(t), \cdots, i_r(t)$, $t = 1, \cdots, T$, are independent identically distributed random variables which take values $1, \cdots, n$ with equal probabilities. We denote by $A_{r,n,T}$ the matrix of this system.

In the case $r = 2$ a critical set of the matrix $A_{2,n,T}$ corresponds to an ordinary cycle in the ordinary graph $G_{A_{2,n,T}}$. In the case $r > 2$ we introduce a concept of a hypercycle as a set of hyperedges which corresponds to a critical set of $A_{r,n,T}$. We prove a threshold property for the mean number of hypercycles in the hypergraph $G_{A_{r,n,T}}$.

Let $r \geqq 3$ be fixed, $T, n \to \infty$ in such a way that $T/n \to \alpha$. Then there exists a constant α_r such that the mathematical expectations $MS(A_{r,n,T}) \to 0$ if $\alpha < \alpha_r$, and $MS(A_{r,n,T}) \to \infty$ if $\alpha > \alpha_r$.

The constant α_r is the first component of the vector which is the only solution of the following system of equations in three unknowns a, x, λ:

$$e^{-x} \cosh \lambda \left(\frac{ar}{ar - x}\right)^a = 1, \qquad \frac{x}{\lambda}\left(\frac{ar - x}{x}\right)^{1/r} = 1, \qquad \lambda \tanh \lambda = x.$$

A. M. ODLYZKO, *AT&T Bell Laboratories*

Search for the maximum of a random walk

Let $X_1, X_2, \cdots$ be independent and identically distributed with $P(X_j = 1) = P(X_j = -1) = \frac{1}{2}$, and let $S_k = X_1 + X_2 + \cdots + X_k$. Thus S_k is the position of a symmetric random walk on the line after k steps. Any algorithm that determines $\max\{S_0, \cdots, S_n\}$ with certainty must examine at least $c_1 n^{\frac{1}{2}}$ of the S_k on average for a certain constant $c_1 > 0$, if all random walks with n steps are considered equally likely. There is also an algorithm that on average examines only $c_2 n^{\frac{1}{2}}$ of the S_k to determine their maximum for another constant c_2. These results can be used to model some search problems on functions that are difficult to compute.

J. W. PITMAN, *University of California, Berkeley*

Cycles and descents of random permutations

Formulae for the joint distribution of the cycle structure and number of descents of a random permutation can be derived from simpler formulae for the distribution of the cycle structure for a model of random riffle shiffles with at most $a-1$ descents. Following Bayer and Diaconis (1992), for positive integer a, define an *a-shuffle* to be the probability distribution on the permutation group S_n that assigns $\pi \in S_n$ the probability

$$P_a(\pi) = \binom{a+n-d-1}{n} a^{-n}$$

where $d = d(\pi)$ is the number of descents in π. For non-negative n_j with $\sum jn_j = n$, the P_a probability that a permutation has n_j cycles of length j, $1 \leqq j \leqq n$, is

$$a^{-n} \prod_{j=1}^{n} \binom{f_{ja} + n_j - 1}{n_j}$$

where f_{ja} is the number of aperiodic circular words of length j for an alphabet of a letters. This formula is a consequence of a bijection discovered by Gessel and Reutenauer (1991) between $\{1, \cdots, a\}^n$ and the collection of multisets of aperiodic necklaces with total length n. As an application of the formulae, the asymptotic behaviour of the distribution of the cycle structure induced by an a-shuffle can be described as $n \to \infty$, for fixed a. The behavior of the large cycles is governed by Poisson-Dirichlet asymptotics, exactly as in the uniform case. But the limiting joint distribution of the numbers of j-cycles for a random a-shuffle, as $n \to \infty$, is the distribution of independent negative binomial variables with parameters (f_{ja}, a^{-j}). Only as $a \to \infty$ does this approach the well-known limiting distribution for the uniform case defined by independent Poisson (j^{-1}) variables. This is joint work with Persi Diaconis and Michael McGrath.

BAYER, D. AND DIACONIS, P. (1992) Trailing the dovetail shuffle to its lair. *Ann. Appl. Prob.* To appear.

GESSEL, I. M. AND REUTENAUER, C. (1991) Counting permutations with given cycle structure and descent set. Preprint.

B. G. PITTEL, *Ohio State University*

Random permutations and stable matchings

A matching on a set of an even number of members is stable—with regard to a given system of members' preferences for a partner—if no two unmatched members prefer each other to their partners under the matching. We study the set of stable matchings for a random instance of the ranking system, under an assumption that each member rank orders potential partners uniformly at random, independently of other members. For the bipartite version ('stable marriages') with n men and n women, we prove that almost surely the total number of stable matchings (marriages) is at least $(n/\log n)^{\frac{1}{2}}$. We show an almost sure (a.s.) existence of an 'egalitarian' marriage, for which the total rank of all spouses is about $2n^{\frac{3}{2}}$, as opposed to $n^2/\log n$ for the extreme—female optimal and male optimal—marriages. Almost surely this particular matching is also (asymptotically) a 'minimum regret' stable marriage, with the largest rank of a spouse in it being close to $n^{\frac{1}{2}} \log n$. Quite unexpectedly, the stable matchings obey—statistically—a law of hyperbola. It states that almost surely the product of the sum of husbands' ranks and the sum of wives' ranks in a stable marriage is asymptotic to n^3, *uniformly* over all stable marriages.

We also study a non-bipartite version of the stable matching problem, which is colloquially known as a 'stable roommates' problem. Here, in a set of even cardinality n, each member ranks all others in order of preference. It is well known that unlike the bipartite version (marriages), a stable matching may not exist. We prove that, for the random instance of the ranking system, the mean and variance of the number of stable matchings are asymptotic to $e^{\frac{1}{2}}$ and $(\pi n/4e)^{\frac{1}{2}}$, respectively. (For the marriages, the mean is about $n \log n/e$.) Thus $P(n)$, the probability that a solution exists, is at least $const/n^{\frac{1}{4}}$. What is $\lim P(n)$? Rob Irving (1985) has performed extensive computer runs using his two-stage proposal algorithm. (The algorithm delivers a stable matching whenever there is one.) The empirical data gathered so far indicate that the possibility of a positive limit cannot be ruled out (Irving, personal communication 1991). We present some preliminary results concerning the likely behavior of Irving's algorithm, and a hyperbola law which holds for stable tables, Irving's extension—to the non-bipartite case—of the notion of stable marriages. We also discuss likely structure of closely related stable (cyclic) partitions, which have been discovered recently by Tan (1991).

Finally, we look at the probabilistic aspects of a conceptually (and mathematically) related trade model introduced and investigated by Shapley and Scarf (1974).

Detailed exposition of the above results is presented in the author's papers Pittel (1989), (1992a, b, c).

IRVING, R. W. (1985) An efficient algorithm for the stable roommates problem. *J. Algorithms* **6**, 577–595.

PITTEL, B. (1989) The average number of stable matchings. *SIAM J. Discrete Math.* **2**, 530–549.

PITTEL, B. (1992a) On likely solutions of a stable marriage problem. *Ann. Appl. Prob.* **2**, 358–406.

PITTEL, B. (1992b) The 'stable roommates' problem with random preferences. *Ann. Prob.* To appear.

PITTEL, B. (1992c) On a random instance of a 'stable roommates' problem. Likely behaviour of the proposal algorithm. Submitted.

SHAPLEY, L. S. AND SCARF, H. (1974) On cores and indivisibility. *J. Math. Econ.* **1**, 23–28.

TAN, J. J. M. (1991) A necessary and sufficient condition for the existence of a complete stable matching. *J. Algorithms* **12**, 154–178.

L. A. SHEPP, *AT&T Bell Laboratories*

Linear and non-linear codes for a special channel

Gargano et al. (1992) want to construct a subset, S, of the group $G_n = \{0, 1, -1\}^n$ with componentwise addition modulo 3, with as many elements as possible and with the property that any two codewords (elements of S) x, y are far apart in the sense that some component of the difference is 1, i.e. $x(i) - y(i) = 1$ for some i. (It follows that some other component of $x - y$ is -1, by interchanging the roles of x and y.) Let $A(n)$ be the cardinality of any set S attaining maximum cardinality with the property. It is easy to see that $A(m+n) \geqq A(m)A(n)$ and so $A(n) \approx a^n$ for $n \to \infty$ for some a, where $2 \leqq a \leqq 3$. The lower bound comes from the example, due to the proposers, of the set

$$S = \{x : x \text{ has } n/2 \text{ 1s' and the rest 0's}\}.$$

If S is also required to be a subgroup of G_n, then the maximum cardinality will be a number, $B(n)$, and again $B(n) \approx b^n$ as $n \to \infty$ for some b, where $\sqrt{3} \leqq b \leqq a$. The lower bound comes from the example

$$S = \{c_1(1, -1, 0, 0, \cdots) + c_2(0, 0, 1, -1, 0, 0, \cdots) + c_3(0, 0, 0, 0, 1, -1, 0, 0, \cdots) + \cdots\}$$

of $n/2$ generators of the subgroup S with $n/2$ independent coefficients $c_1, c_2, \cdots$ in $\{0, 1, -1\}$.

We prove that these lower bounds are asymptotically best possible (for $A(n)$, only in the weak sense of $\approx$), so that $b = \sqrt{3}$, and $a = 2$. Thus there are considerably fewer codewords possible if S is required to be a linear, or group code, which seems to contradict, to some extent, the naive belief that the more efficient (in the sense of decodability) group codes give up little in the sense of capacity. This is joint work with Rob Calderbank, Peter Frankl, Ron Graham and Wen-Ch'ing Winnie Li.

GARGANO, L., KORNER, J. AND VACCARO, U. (1992) Sperner theorems on directed graphs and qualitative independence. *J. Combinatorial Theory* **A**. To appear.

J. SPENCER, *Courant Institute, New York*

The Poisson paradigm and random graphs

When a random variable X is the sum of many indicator random variables, each rare and mostly independent, the Poisson paradigm is that X has close to the Poisson distribution. In particular, if $E[X]=\mu$ then $P[X=0]$ should be close to $e^{-\mu}$. This is a natural situation in random graphs. For example, in the original papers of Erdős and Rényi on random graphs it was shown that in $G(n,p)$ if $p=p(n)$ is such that the expected number of triangles is a constant μ then indeed the probability that there is no triangle approaches $e^{-\mu}$.

A few years ago Svante Janson, employing a variant of the Stein–Chen method, found a pair of inequalities now known as the Janson inequalities. With these, results such as the above come out with a fairly elementary calculation on random graphs, involving no more, basically, than evaluation of the second moment. Applications include the following:

- Bounds on the probability that $G\sim G(n,p)$ contains no subgraph H, for a fixed H and various $p=p(n)$.
- Fine threshold behavior for every vertex to lie in a triangle, and similar extension statements.
- The existence of sets S of positive integers so that the number of representations $n=x+y+z$ with $x, y, z\in S$ is $\Theta(\ln n)$.

ALON, N., ERDÖS, P., AND SPENCER, J. (1991) *The Probabilistic Method.* Wiley, New York.

L. TAKÁCS, *Case Western Reserve University*

On the heights and widths of random rooted trees

Denote by S_n the set of all distinct rooted trees with vertices $1,2,\ldots,n$ in which the root is labeled 1. Each tree is represented by an ordered sequence of n non-negative integers $(i_1,i_2,\cdots,i_n)$ satisfying the conditions $i_1+i_2+\cdots+i_n=n-1$, and $i_1+i_2+\cdots+i_r\geqq r$ for $1\leqq r<n$. In a tree, represented by $(i_1,i_2,\cdots,i_n)$, two vertices r and s $(1\leqq r<s\leqq n)$ are joined by an edge if and only if $i_0+i_1+\cdots+i_{r-1}<s\leqq i_0+i_1+\cdots+i_r$ where $i_0=1$. The number of trees in S_n is $|S_n|=C_{n-1}=\frac{1}{n}\binom{2n-2}{n-1}$, where $C_0=C_1=1$, $C_2=2$, $C_3=5$, $C_4=14,\cdots$ are the Catalan numbers.

Let $\{p_j\}$ be a probability distribution on the set of non-negative integers with expectation 1 and standard deviation σ $(0<\sigma<\infty)$. Let us choose a tree at random in S_n, assuming that the probability of a tree represented by $(i_1,i_2,\cdots,i_n)$ is $p(i_1,i_2,\cdots,i_n)=A_np_{i_1}p_{i_2}\cdots p_{i_n}$ where A_n is determined by the requirement $\sum_{S_n}p(i_1,i_2,\cdots,i_n)=1$.

For a tree chosen at random in S_n define the random variable $\tau_n(m)$ as the number of vertices at distance m from the root. Furthermore, define $\mu_n=\max\{m:\tau_n(m)>0\}$ as the height of the tree, $\delta_n=\max\{\tau_n(m):m\geqq 0\}$ as the width of the tree and $\tau_n=\sum_{m\geqq 0}m\tau_n(m)$ as the total height of the tree.

Explicit formulas are given for the asymptotic distributions of τ_n, μ_n, δ_n and $\tau_n(m)$ if $m=[2\alpha\sqrt{n}/\sigma]$ where $0<\alpha<\infty$ and $n\to\infty$.

V. A. VATUTIN, *Steklov Mathematical Institute, Moscow*

Limit theorems for the height of a random planted plane tree

Let $\tau_N=\tau_N(T)$ be the height of a planted tree T chosen at random from the set $\mathcal{F}_N$ of all planted plane trees having N vertices. Using a natural correspondence between the set $\mathcal{F}_N$ and the set of all realizations of a simple random walk S_n: $S_0=0$, $S_n>0$, $1\leqq n<2N$, $S_{2N}=0$, we find the asymptotics of $P\{\tau_N>n\}$ as $N\to\infty$, $n^2N^{-1}\to\infty$, $nN^{-1}\to 0$, and of $P\{\tau_N=n\}$ as $n^2N^{-1}\to x\in(0,\infty)$.

A. M. VERSHIK, *St. Petersburg University*

Random permutations, limit shapes and asymptotic problems of partition theory

Asymptotic properties of the limit measures which appear in additive problems in partition theory and number theory can be considered in a geometric manner. There are different types of asymptotic behaviour: 'ergodic', including problems like the LLN and CLT, and 'non-ergodic', in which one can calculate the limit distribution and the boundary. All these kinds of examples appear in the context of random permutations and statistics on the partitions of integers or reals. It turns out that completely different problems can give us the same limit measures.

O. V. VISKOV, *Steklov Mathematical Institute, Moscow*

The Rota umbral calculus and the Heisenberg–Weyl algebra

The talk was based on some of the author's papers (Viskov (1986) and references therein, and Viskov (1981), (1991)).

The Heisenberg-Weyl algebra is the algebra freely generated by three variables A, B and C subject to the identities

$$C=[A, B]=AB-BA, \qquad [A, C]=[B, C]=0. \tag{1}$$

The main purpose of the talk was to emphasize the role played by a suitable representation of this algebra in contemporary umbral or operator calculus originated in a seminal paper of Rota (1964).

Let $p=\{p_n(x), n=0, 1, 2, \cdots\}$ be an arbitrary basis in the commutative algebra $\mathscr{P}$ of all polynomials of a single variable x with coefficients in a field of characteristic zero and let $\mathscr{L}$ be the set of linear maps from $\mathscr{P}$ into $\mathscr{P}$. Since every operator in $\mathscr{L}$ is uniquely determined by its actions on an arbitrary basis p of $\mathscr{P}$, the relations

$$\begin{aligned} A[p_0(x)]=0, \qquad A[p_n(x)]=np_{n-1}(x), \qquad n=1, 2, \cdots; \\ B[p_n(x)]=p_{n+1}(x), \qquad n=0, 1, 2, \cdots, \end{aligned} \tag{2}$$

give us the desirable representation of the Heisenberg-Weyl algebra if we take the identity map as C. The simplest particular case of this situation is $p_n(x)=x^n$, $n=0, 1, 2, \cdots$, and then $A=d/dx$ and the operator B is multiplication by x.

Representation (2) and the systematic use of relations (1) allow us to obtain easily many important formulae in the finite operator calculus (Rota (1975), Roman (1984)) including the recurrence and transfer formulae, umbral operators and effective tool for the composition and inversion of power series. Moreover, the proofs become essentially simpler.

This approach admits a simple generalization to the multivariate case. It is also useful for analysis of many other situations, for instance, in the Stanley (1988) theory of differential posets.

Roman, S. M. (1984) *The Umbral Calculus.* Academic Press, Orlando, FL.

Rota, G.-C. (1964) The number of partitions of a set. *Amer. Math. Monthly* **71,** 498–504.

Rota, G.-C. (1975) *Finite Operator Calculus* Academic Press, New York.

Stanley, R. P. (1988) Differential posets. *J. Amer. Math. Soc.* **1,** 919–961.

Viskov, O. V. (1981) A class of linear operators. In *Generalized Functions and their Applications, Proc. Inter. Conference, Moscow,* pp. 110–120 (in Russian).

Viskov, O. V. (1986) A noncommutative approach to classical problems of analysis. *Trudy Mat. Inst. Steklov* **177,** 21–32. (in Russian). English translation in *Proc. Steklov Inst. Math.* 1988 (4), 21–32.

Viskov, O. V. (1991) On the ordered form of noncommutative binomial. *Uspechi Mat. Nauk* **46,** 209–210. (in Russian).

H. S. WILF, *University of Pennsylvania*

Ascending subsequences of permutations and Young tableaux

It is well known, from Shensted's algorithm, that there is a relationship between the longest increasing subsequence in a permutation and the length of the first row of a Young tableau. This talk presented a quantitative version, i.e. an explicit relationship between the numbers of permutations of n letters whose longest increasing subsequence is of length k and of Young tableaux of n cells whose first row has length k. The relationship is surprisingly simple. A number of unsolved problems were raised.

CONTRIBUTED PAPERS

W. J. EWENS, *University of Pennsylvania*

Sampling properties of random mappings

A random mapping of $\{1, 2, \cdots, m\}$ to $\{1, 2, \cdots, m\}$ leads to a (random) component structure. We consider the induced structure on a sample of n integers, which can be taken to be $\{1, 2, \cdots, n\}, n \leqq m$. Aldous (1985) has shown that, as $m \to \infty$, with n fixed, the probability structure of the induced partition of $\{1, 2, \cdots, n\}$ into components converges weakly to the Ewens sampling formula with parameter $\frac{1}{2}$. However, in this formula, as $n \to \infty$, various interesting quantities do not converge to the corresponding limiting $(m \to \infty)$ properties of the original component partition. For example, if $U_1(n, m)$ is the number of induced components of size 1 in the sample, then

$$\lim_{m \to \infty} EU_1(n, m) = e^{-1},$$

and

$$\lim_{n \to \infty} \lim_{m \to \infty} EU_1(n, m) = \tfrac{1}{2}.$$

The possibility for the inequality arises because $U_1(n, m)$ cannot be expressed as a bounded continuous function of the component sizes. Corresponding problems do not arise for random permutations.

ALDOUS, D. J. (1985) Exchangeability and related topics. *Lecture Notes in Mathematics* **1117**, 1–198, Springer-Verlag, Berlin.

B. FRISTEDT, *University of Minnesota*

Random partitions of large integers

Random partitions of integers are treated for the case where all partitions of an integer n are equally likely. The focus is on limit theorems as $n \to \infty$. In particular, as $n \to \infty$,, the decreasing sequence of large parts, beginning with the largest part, then the next largest part, etc. approaches, when appropriately normalized, a certain Markov chain which can be explicitly identified. The major tool is a simple construction of random partitions that treats the number being partitioned as a random variable in such a way that the numbers of parts of various sizes are independent random variables. The Markov result and other results using this construction are in Fristedt (1992).

FRISTEDT, B. (1992) The structure of random partitions of large integers. *Trans. Amer. Math. Soc.* To appear.

D. R. GAVELEK, *XonTech Inc.*

The height of elements in random mappings

Interest in the properties of random mappings was stimulated almost forty years ago by Metropolis and Ulam. Over the intervening years random mapping models have been used in applications ranging from random number generation and cryptology to the simulation of epidemic processes and tests of the intrinsic randomness of quantum mechanics. In this paper approximately twenty characteristics related to the height distributions of elements in the functional graphs or de Bruijn diagrams representing a random map were discussed. Some of these distributions are well known. Other results, such as the expected number of ancestors of an element of height H, and the average height of an orphan point, appear to be new. As an additional unifying factor it was shown that all of these parameters are naturally expressed in terms of the incomplete gamma function.

A. P. GODBOLE, *Michigan Technological University*

Some results on Poisson and compound Poisson approximation

Let $N = N(n, k)$ denote the number of non-overlapping occurrences of a fixed k-letter word obtained while generating n letters from a ξ-letter alphabet; the generation is performed either in an i.i.d. fashion or according to a positive transition matrix Γ. We obtain improved Poisson approximations for $\mathscr{L}(N)$ using the Stein–Chen method and eigenvalue bounds on convergence to stationarity for non-reversible chains. A Poisson approximation is also derived for the distribution of $X(n, k)$, which denotes the number of matches obtained, while sampling n times (with replacement) from an urn with m balls, and when only the k previously drawn balls are remembered. Finally, process versions of the Stein–Chen technique are employed to obtain a compound Poisson approximation in a non-i.i.d. urn problem related to the determination of the number X of winners of a lottery jackpot.

J. C. HANSEN, *Northeastern University*

Order statistics for random combinatorial structures

We consider labeled and unlabeled 'decomposable' combinatorial structures which are characterized by the following generating function equations. In the labeled case, $\hat{P}(z) = \exp \hat{C}(z)$ where $\hat{P}(z)$ is the exponential generating function for the number of structures of size n and $\hat{C}(z)$ is the exponential generating function for the number of connected structures of size n. In the unlabeled case, $P(z) = \exp(C(z) + C(z^2)/2 + \cdots)$ where $P(z)$ is the ordinary generating function for the number of structures of size n and $C(z)$ is the ordinary generating function for the number of 'connected' structures of size n. In both cases, we are interested in the measure induced $\nabla = \{\{x_i\} : x_1 \geqq x_2 \geqq \cdots \geqq 0, \sum x_i \leqq 1\}$ by the (decreasing) sequence of order statistics for the component sizes of a random structure of size n (normalized by n). We show that if the generating functions $\hat{C}(z)$, in the labeled case, and $C(z)$, in the unlabeled case, are logarithmic functions then the induced measures on ∇ converge in distribution to a Poisson–Dirichlet distribution on ∇. In the labeled case, this result unifies results known for particular examples such as random permutations and random mappings. In the unlabeled case, this gives new distributional results for examples such as factorization of polynomials over $GF(q)$.

L. HOLST, *Royal Institute of Technology, Stockholm*

On ménage problems

Consider n couples seated at a circular table with men and women taking alternating seats but otherwise completely random seating. Let W be the number of couples sitting next to each other. To determine the distribution of W is sometimes called the ménage problem; it is considered to be a tricky combinatorial problem. When n is large the distribution is approximately Po(2); the variation distance between the exact and the approximation is $\asymp n^{-1}$. Much better approximation is obtained by Bi $(2n, 1/n)$ giving the variation distance $\asymp n^{-2}$. The ménage problem can be generalized in various ways, e.g. letting there be many tables. The corresponding random variables are of the form $W = \sum c_{ij} I_{ij}$ with (c_{ij}) a given binary matrix and (I_{ij}) a random permutation matrix. Good explicit upper bounds between the variation distance of such distributions and Poisson with the same mean can be given.

HOLST, L. (1991) On the 'problème des ménages' from a probabilistic viewpoint. *Statist. Prob. Lett.* **11**, 225–231.

BARBOUR, A. D., HOLST, L. AND JANSON, S. (1992) *Poisson Approximation.* Oxford University Press.

Z.-X. HU, *University of Illinois, Urbana*

On ([n], P)-partitions

We introduce the concept of ([n], P)-partitions which are generalizations of partitions of [n]. (It was inspired by R. P. Stanley's (P, ω)-partitions)) where P is a poset and $[n]=\{1, 2, \cdots, n\}$. Many mathematical models can be considered as special cases of ([n], P)-partitions. One interesting application of ([n], P)-partitions is the special way of realizing a finite poset P by the complete graph K_n on [n]. A new parameter $n(P)$ (called the norm of P) is naturally derived. $n(P)$ gives an indication of the complexity of P by reflecting both $|P|$ and order relations in P. We also introduce some new results about ([n], P)-partitions and $n(P)$.

HU, Z.-X. (1990) On generalized partitions of an N-set. *Congressus* **76**, 55–62.

STANLEY, R. P. (1972) Ordered structures and partitions. *Amer. Math. Memoires* **119**, 1–104.

J. JAWORSKI, *Adam Mickiewicz University, Poznan*

The evolution of a random mapping

A random mapping $(T_n; q)$ of a finite set $V=\{1, \cdots, n\}$ into itself assigns independently to each $i \in V$ its unique image $j=T(i) \in V$ with probability q for $i=j$ and with probability $P=(1-q)/(n-1)$ for $i \neq j$. We study the evolution of a random digraph $G_{T_n}(q)$, representing $(T_n; q)$, as its arc-occurrence probability $P=P(n)$ increases from 0 to $1/(n-1)$. The structure of functional digraphs enables asymptotic studies of exact discrete distributions of many characteristics related to G_{T_n}. For example, we consider the number of predecessors of m given vertices, the quasi-binomially distributed random variable associated with a particular epidemic process. Finally, let $(T_n; M)$ be a random element of a family of all loopless digraphs on n vertices with exactly M vertices of outdegree 1 and $n-M$ vertices of outdegree 0. Clearly, there is an equivalence between $(T_n; q)$ and $(T_n; M)$. Moreover, $(T_n; M)$ can be treated as the Mth stage of the 'regular' random digraph process $\{G_{T_n}(M)\}_{M=0}^{n(n-1)}$. We study the appearance of the first cycle in such a process and the structure of the digraph near the critical point $M=n$.

P. J. JOYCE, *University of Idaho*

Poisson limit laws for dependent random permutations

The process of cycle counts $(C_1, C_2, \cdots, C_n, 0, 0, \cdots)$ for a random permutation $\mathbf{\Pi}$ of length n distributed according to the Ewens sampling formula converges to a Poisson process with independent coordinates. This result is extended to a vector of random permutations $\mathbf{\Pi}=(\Pi_1, \cdots, \Pi_d)$ in the following way. Let $\boldsymbol{Y}=(Y_1, \cdots, Y_d)$ be an integer-valued random vector with $\sum_{i=1}^d Y_i=n$. Conditional on $\boldsymbol{Y}, \Pi_i$ is a permutation of length Y_i distributed according to a Ewens sampling formula. For $i=1, \cdots, d$ define $\boldsymbol{C}_i=(C_{i1}, \cdots, C_{in}, 0, \cdots)$, where C_{ij} is the number of cycles of size j in permutation $\mathbf{\Pi}_i$. It can be shown that for a wide class of distributions for $\boldsymbol{Y}$, the $\boldsymbol{C}_i$ convergence to independent Poisson processes. Total variation techniques are used to establish the result. The work is motivated by a problem in population genetics. This is joint work with Simon Tavaré.

V. I. KHOKHLOV, *Steklov Mathematical Institute, Moscow*

On the structure of a non-uniformly distributed random graph

We consider a random graph $G_{N,T}$ with N labelled vertices and T edges. These T edges are obtained by T independent trials. In each trial the edge between vertices i and j occurs with the probability $2p_ip_j$, and the loop at the vertex i occurs with the probability p_i^2; $i, j = 1, \cdots, N, p_1, \cdots, p_N \geqq 0, p_1 + \cdots + p_N = 1$.

Let $N \to \infty$, $2T/N \to \lambda$, $p_i = a_i/N$, $a_i = a_i(N)$, $i = 1, 2, \cdots, N$, and suppose there exists a limit

$$a^2 = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} a_i^2$$

and positive constants ε and E such that $\varepsilon \leqq a_i \leqq E$, $i = 1, 2, \cdots, N$. Then, under the additional condition $\lambda a^2 < 1$, the graph $G_{N,T}$, with probability approaching 1, does not contain components with more than one cycle and tree-components that have more than $c \log N$ vertices, where c is a constant. Moreover, under these conditions the distribution of the number of cycles in the graph converges to the Poisson distribution with parameter $\Lambda = -\frac{1}{2} \ln (1 - \lambda a^2)$.

KOLCHIN, V. F. AND KHOKHLOV, V. I. (1990) On the number of cycles in a non-equiprobable random graph. *Diskretnaya Matematika* **2,** 137–145 (in Russian).

KOLCHIN, V. F. AND KHOKHLOV, V. I. (1991) On the structure of a random graph with nonuniform distribution. In *New Trends in Probability and Statistics,* Vol. 1, pp. 445–456, VSP/Mokslas.

P. MATTHEWS, *University of Maryland Baltimore County, Baltimore*

A lower bound on the probability of conflict under non-uniform access in database systems

We consider an N item database and t transactions, each of which will independently request a subset of the items. For $i = 1, \cdots, t$ transaction i will request n_i items according to some probability distribution on the $\binom{N}{n_i}$ sets of n_i items. We say there are no conflicts if the t chosen sets are all disjoint. In probabilistic language t complexes of balls are being allocated independently to N urns, and we are considering the probability that no urn receives two or more balls. If each of the complexes is of size 1, the problem is the generalization of the birthday problem to birthdays with a non-uniform probability distribution. If the transactions choose simple random samples, then the probability of no conflicts is $\binom{N}{n_1, \cdots, n_t}\left(\prod_{i=1}^{t} \binom{N}{n_i}\right)^{-1}$. We give a class of sampling schemes of practical interest and show that, within this class, the probability of no conflicts is no larger than that for simple random sampling. This supports a long-standing conjecture in the database community that uniform access minimizes the probability of conflicts. This is joint work with K. Humenik, A. B. Stephens, and Y. Yesha.

E. SCHMUTZ, *Drexel University*

Part size statistics on general partition families

Put a uniform probability distribution on the set of partitions of the integer n into parts that are elements of a certain set A. If $A_1, A_2, \cdots, A_d$ are disjoint sets whose union is A, let $P_i(\lambda)$ denote the number of part sizes that the partition λ has in A_i. Under certain conditions, due essentially to Meinardus, the random vector $P = (P_1, P_2, \cdots, P_d)$ is asymptotically normally distributed.

MEINARDUS, G. (1954) Asymptotische Aussagen über Partitionen. *Math. Z.* **59,** 225–241.

W. STADJE, *University of Osnabrück, Germany*

On sets of integers with prescribed gaps

For a fixed set I of positive integers we consider the set of paths $(p_0, \cdots, p_k)$ of arbitrary length satisfying $p_l - p_{l-1} \in I$ for $l = 2, \cdots, k$ and $p_0 = 1$, $p_k = n$. Equipping it with the uniform distribution, the random path length T_n is studied. Asymptotic expansions of the moments of T_n are derived and its asymptotic normality is proved. The step lengths $p_l - p_{l-1}$ are seen to follow asymptotically a restricted geometrical distribution. Analogous results are given for the free boundary case in which the values of p_0 and p_k are not specified. In the special case $I = \{m+1, m+2, \cdots\}$ (for some fixed $m \in \mathbb{N}$) we derive the exact distribution of a random 'm-gap' subset of $\{1, \cdots, n\}$ and exhibit some connections to the theory of representations of natural numbers. A simple mechanism for generating a random m-gap subset is also presented. This is joint work with Y. Baryshnikov.

J. M. STEELE, *University of Pennsylvania*

Long common subsequence problems

If X_i and Y_i are independent random variables with values in the same alphabet, the variable L_n that we investigate is defined as the maximal m such that there are subsequences $i_1, i_2, \cdots, i_m$ and $j_1, j_2, \cdots, j_m$ of $\{1, 2, \cdots, n\}$ such that $X_{i_k} = Y_{j_k}$ for all $1 \leqq k \leqq m$. This talk briefly reviews recent progress on the tightness of concentration and other properties of this variable.

P. TETALI*, *DIMACS Center, Rutgers University*

Covering with latin transversals

Given an $n \times n$ matrix $A = [a_{ij}]$, a *transversal* of A is a set of elements, one from each row and one from each column. A transversal is a *latin transversal* if no two elements are the same. There have been more conjectures than theorems on latin transversals in the literature. Recently, Erdős and Spencer (1990) showed that there always exists a latin transversal in any $n \times n$ matrix in which no element appears more than s times, for $s \leqq (n-1)/16$. Here we show that, in fact, all the elements of the matrix can be partitioned into latin transversals, provided n is a power of 2 and no element appears more than εn times for some fixed $\varepsilon > 0$.

Theorem. Let n be 2^m. Any $n \times n$ matrix in which no element appears more than s times contains n disjoint latin transversals provided $s \leqq \varepsilon n$ (for ε, an absolute constant $\ll 1$).

The assumption that n is a power of 2 can be weakened, but at the moment we are unable to prove the theorem for all values of n. On the other hand, our proof can be easily modified to prove the existence of many pairwise disjoint transversals in any $n \times n$ matrix in which no entry appears more than εn times, without any restriction on n. Therefore our method implies a strengthening of the result of Erdős and Spencer for any n, (apart from the actual value of the constant ε). The proof of the theorem involves *random partitioning* and the *Lovász local lemma*. This is joint work with N. Alon and J. Spencer.

ERDŐS, P. AND SPENCER, J. (1990) Lopsided Lovász local lemma and latin transversals. *Disc. Appl. Math.* **30**, 151–154.

* Now at AT&T Bell Laboratories, Murray Hill.