GIBBS' MEASURES ON COMBINATORIAL OBJECTS AND THE CENTRAL LIMIT THEOREM FOR AN EXPONENTIAL FAMILY OF RANDOM TREES

J. MICHAEL STEELE

Program in Engineering Statistics
Princeton University
Princeton, New Jersey

A model for random trees is given which provides an embedding of the uniform model into an exponential family whose natural parameter is the expected number of leaves. The model is proved to be analytically and computationally tractable. In particular, a central limit theorem (CLT) for the number of leaves of a random tree is given which extends and sharpens Rényi's CLT for the uniform model. The method used is general and is shown to provide tractable exponential families for a variety of combinatorial objects.

1. INTRODUCTION

The theory of random combinatorial objects has shown a considerable devotion to the uniform distribution. This is a natural development, and a great number of interesting and useful results have been produced. Still, for the purpose of modeling and for technical applications, it seems valuable to be able to go beyond uniformity and the elementary extensions amounting to conditional uniformity.

47

A simple technique given here will show how one can sometimes construct a combinatorially rich family of nonuniform probability distributions on a class of combinatorial objects. More pointedly, one can still retain many of the attractive features of the uniform distribution. Those features include (1) a sufficient analytical tractability to permit limit results like the central limit theorem, (2) the possibility of economical computer simulation to obtain realizations from the distribution, (3) the presence of one or more parameters to permit tuning of the family distributions to the circumstances of modeling applications, and (4) an intuitive meaning of the underlying parameters of the distributions.

To illustrate the technique as concretely as possible, attention will be focused here on the construction of a family of random trees. It will be seen that the constructed family is in fact an exponential family in the usual sense of parametric families of distributions. Moreover, the sufficient statistics for the natural parameters have basic combinatorial interpretations. This exponential family of trees has the benefit of permitting an easy asymptotic analysis, and it also permits two general techniques for simulations of random samples.

The study of random trees was initiated in Rényi [20], where a model of random trees was built on the classical result of enumerative combinatorics that there are exactly n^{n-2} labeled trees with n vertices, (Cayley [3]).[1]

The basic object Rényi chose to study was the number of leaves (i.e., vertices of degree one) of a tree chosen at random from the set of all n^{n-2} labeled trees. If we let L_n denote the number of leaves of such a random tree, then Rényi's main results can be stated as follows:

$$E(L_n) \sim n/e \qquad \text{as } n \to \infty \tag{1.1}$$

$$\text{var}(L_n) \sim (e - 2)n/e^2 \qquad \text{as } n \to \infty \tag{1.2}$$

and, principally, the normalized number of leaves

$$\frac{L_n - n/e}{e^{-1}\sqrt{n(e - 2)}} \tag{1.3}$$

is asymptotically normal.

In the second section we establish some structural prerequisites to the study of random trees; and, in particular, we introduce the coding structure needed to construct random trees under the uniform model. The third section introduces an exponential family of random trees where the number of leaves is the sufficient statistic for the family's natural parameter. The measure-producing process given there starts with the notion of a general Gibbs distribution. It then quickly specializes to the construction of a Gibbs distribution on the set of labeled trees where there associated energy function for the Gibbs distribution is the number of leaves of the tree.

Sections four, five, and six pursue the probabilistic analysis of this new family of measures on labeled trees. The main analytic technique depends on

establishing a representation of a random variable as a sum of Bernoulli random variables by showing that the probability generating function has all real roots. This powerful idea was introduced in the work of Harper [9] on the normal law for Stirling numbers of the second kind. One benefit of Harper's method is that Rényi's central limit theorem can be sharpened to give a Berry-Essen type result on the rate of convergence to normality. The Bernoulli representation also lets one easily deduce some subtle qualitative features of random trees like the unimodularity of the probability mass function of L_n.

Conceptually, the best thing about Harper's method is that the number of leaves of a tree chosen from the Gibbs' family is shown to be representable as a sum of independent Bernoulli random variables. The issue of random leaves is brought down to the problem of number of heads of (properly chosen) biased coins. The bias turns out to depend on the roots of certain simple polynomials. Curiously, there is still no combinatorial interpretation of the Bernoulli summands.

In section seven we see how to generate a random tree from our exponential family. Two methods are given and contrasted. One of these is based on the algorithms of Metropolis *et al.*, [14] for generating an observation from a Gibbs distribution. This method has come recently into wide circulation because of the so-called method of simulated annealing in combinatorial optimization, (see e.g., Kilpatrick *et al.*, [11]). A second method is given for generating random trees from the Gibbs measure which is based on more special features of our particular family of trees.

The final section discusses other combinatorial problems to which the program of Gibbs' measures and Harper's method can be applied. In particular, results are given for the choice of a random set from a set of size n, and to the choice of a random matching from the set of all matchings of a graph.

2. STRUCTURAL PREREQUISITES

Rényi's theory of random trees is based on a method of creating a one-to-one correspondence between a labeled tree on n vertices $P_1, P_2, \ldots, P_n$ and an $(n-2)$-tuple with elements from $\{1, 2, \ldots, n\}$. This correspondence is due to Prüfer [19] and is obtained as follows:

> First we find the leaf P_i with the highest value of the index i. We then delete P_i from the tree and record the index j of the unique vertex P_j to which P_i was adjacent. This process is repeated on the pruned tree until $n-2$ vertices have been deleted (and two vertices are left remaining).

The sequence of recorded adjacency indices becomes the Prüfer code for the tree. One consequence of this coding is that Cayley's formula for the number of trees on n vertices is now rendered obvious; our code has length $n-2$ and each place can be held by n values, so n^{n-2} is the cardinality of the set of labeled trees. For more detail on the Prüfer code as well as various generalizations and applications, one can consult Moon [15], Knuth [12], or Lovász [13].

With the Prüfer code in hand, it is easy to obtain a formula for the number of trees with k leaves. From the construction of the code, one has the fact that P_i is a leaf if and only if i does not appear in the Prüfer code. The number of ways to choose the k leaves which are not to appear in the code is $\binom{n}{k}$, and the number of functions from an $(n-2)$ set onto an $(n-k)$ set is $(n-k)!s(n-2, n-k)$ where $s(n,k)$ is the Stirling number of the second kind (i.e., the number of ways to partition an n set into k subsets); so, if we let $T(n,k)$ denote the number of labeled trees with n vertices and exactly k leaves, we then obtain, after Rényi, that

$$T(n,k) = n! \frac{s(n-2, n-k)}{k!} \quad .$$ (2.1)

By using the well-known recurrence relation $s(n,k) = s(n-1, k)k + s(n-1, k-1)$ together with Eq. (2.1), Rényi [20] obtained the basic identity on which he based his asymptotic analysis:

$$\sum_{k=2}^{n-1} \frac{T(n,k) \binom{x}{n-k}}{\binom{n}{n-k}} = x^{n-2} \quad .$$ (2.2)

One can easily obtain the asymptotic formula for the means, $EL_n \sim n/e$, just by letting $x = n - 1$ in Eq. (2.2). Similarly, the variance relation $\operatorname{var} L_n \sim n(e-2)/e^2$ is obtained by substituting $x = n - 2$.

It is worth noting that there is some subtlety in the formula for $\operatorname{var} L_n$. The naive notion that L_n behaves like the sum of n (or $n-2$) independent Bernoulli random variables with mean $p = 1/e$ is supported by $EL_n \sim n/e$, but it would suggest an asymptotic variance of $n(1 - 1/e)(1/e)$. The true asymptotic variance shows that the variance suggested by such a naive Bernoulli approximation is too big. The surprising fact which will be established shortly is that L_n does nevertheless have the same distribution as the sum of $n - 2$ independent Bernoulli variables. This fact is consistent with the asymptotics of EL_n and $\operatorname{var} L_n$ only because these Bernoulli variables are not identically distributed.

The central limit theorem for L_n was obtained by Rényi by substituting $x = 1 - it/\sqrt{n}$ into his key identity Eq. (2.2). The asymptotic analysis which then led to Eq. (1.3) was considerably more subtle than that which will be required here to obtain a stronger and more general result.

3. GIBBS' MEASURE ON THE SET OF TREES

Before beginning a re-analysis of L_n under the uniform model, it will be useful to obtain two general perspectives on the distributions which can be placed

on the set of all labeled trees. First, we will consider a point of view common in statistical mechanics.

If S is any finite set and $f: S \to R$ is any real function on S, then we can define a parametric family of measures on S by

$$\mu_\beta(s) = \frac{e^{-f(s)\beta}}{\sum_{t \in S} e^{-f(t)\beta}} \ . \tag{3.1}$$

If S is thought of as the set of possible states of a physical system and if $f(s)$ is considered to be the energy of the state s, then $\mu_\beta(s)$ is called the Gibbs measure associated with f. The statistical mechanical interpretation of $\mu_\beta(s)$ is the probability of the state s. The expression $1/\beta$ is interpretable as a measure of *temperature*: so, in the first place, states of higher energy are less likely than states of lower energy. In the second place, as temperature tends to infinity, the Gibbs measure becomes maximally disordered, i.e., uniform on all states.

When temperature tends to zero, μ_β becomes concentrated on the states s of lowest energy $f(s)$; and, partly for that reason, the Gibbs measure has recently emerged as a tool in combinatorial optimization. The Gibbs measure is of basic importance in statistical mechanics, and for many models it has been studied in remarkable depth.

The formalism of Gibbs' measure has a convenient specialization to the set S of labeled trees on n vertices. In particular, if we let $f(s)$ denote the number of leaves of the tree $s \in S$ and let $Z(\beta)$ denote the denominator of Eq. (3.1), then we can express $Z(\beta)$ in terms of $T(n,k)$,

$$Z(\beta) = \sum_{2 \leq k \leq n-1} e^{-\beta k} T(n,k) \ . \tag{3.2}$$

In general, the denominator of the Gibbs' measure of Eq. (3.1) is called the partition function in the classical terminology, and $Z(\beta)$ is the traditional notation.

One should note from Eq. (3.2) that $Z(0) = n^{n-2}$; and, if P_β denotes the probability distribution on random trees given by the Gibbs' measure μ_β, then

$$P_\beta(L_n = k) = \frac{e^{-\beta k} T(n,k)}{Z(\beta)} \ . \tag{3.3}$$

If we let ϕ_n denote the probability generating function for L_n under the uniform model ($\beta = 0$) we have

$$\phi_n(x) = \sum_{2 \leq k \leq n-1} x^k P_0(L_n = k) \tag{3.4}$$

and there is also a simple relation between Z and ϕ,

$$Z(\beta) = \phi_n(e^{-\beta}) Z(0) \ . \tag{3.5}$$

4. REAL ROOTS OF THE LEAF POLYNOMIAL

We can now obtain an approximate representation for L_n under *both* the uniform and Gibbs models. A key point is that whenever Harper's method is successful there is an attendant extension to an appropriate Gibbs model which comes almost for free.

Harper [9] showed that if Y_n is the number of subsets in a partition of $\{1, 2, \ldots, n\}$ which is chosen at random from the set of all $B_n = \sum_k s(n, k)$ such partitions, then Y_n has the same distribution as the sum of n independent Bernoulli random variables. To recall Harper's method, and to establish some facts which we will use later, we let $\Psi_n(x)$ denote the probability generating function of Y_n, so

$$\Psi_n(x) = Ex^{Y_n} = B_n^{-1} \sum_{1 \leq k \leq n} s(n, k)x^k \ . \tag{4.1}$$

Using a variant of Rolle's theorem, Harper was able to give a rather easy proof that $\Psi_n(x)$ has all real roots. Since these roots must be nonpositive by the positivity of the coefficients of $\Psi_n(x)$, we see there are real numbers $r_i \geq 0$ such that

$$\Psi_n(x) = \frac{\displaystyle\prod_{i=1}^{n} (x + r_i)}{\displaystyle\prod_{i=1}^{n} (1 + r_i)} = \prod_{i=1}^{n} (p_i x + q_i) \tag{4.2}$$

where $p_i = (1 + r_i)^{-1}$ and $q_i = r_i/(1 + r_i)$.

This factorization is equivalent to saying that Y_n has the same distribution as the sum of n independent Bernoulli random variables. This was Harper's key probabilistic observation. It leads naturally to a central limit theorem for Y_n and to a sharper understanding of the qualitative behavior of the Stirling numbers. As an example of the latter, we note that the Bernoulli factorization of Eq. (4.2) implies that the Stirling numbers are unimodal. This is a fact which, though apparent in computed examples, is otherwise taxing to establish.

Because of the explicit formula in Eq. (2.1) for $T(n, k)$ in terms of Stirling numbers, it is natural to begin an analysis of the probability generating function for L_n by using what is known about Ψ_n. This requires only a little index shifting. We first note that by the defining relation of Eq. (4.1), we have

$$B_{n-2}\Psi_{n-2}(x) = \sum_{j=1}^{n-2} s(n-2, j)x^j$$

and, by obvious substitution and multiplication,

$$\sum_{j=1}^{n-2} s(n-2, k)y^{n-j} = y^n B_{n-2} \Psi_{n-2}\left(\frac{1}{y}\right) \ .$$

By the fact that Ψ_{n-2} has only real roots, we see that

$$\sum_{j=1}^{n-2} s(n-2, j)x^{n-j} = 0 \qquad (4.3)$$

must also have all real roots.

A remarkable fact about polynomials with real coefficients (c.f. Pólya and Szegö [18], p. 45) is that when an equation $a_0 + a_1 x + a_2 x^2 + \ldots a_n x^n = 0$ has only real roots, then the equation

$$a_0 + a_1 \frac{x}{1!} + a_2 \frac{x^2}{2!} + \ldots + a_n \frac{x^n}{n!} = 0$$

also has only real roots. Applying the result of Pólya and Szegö to Eq. (4.3), we see that

$$\sum_{j=1}^{n-2} s(n-2, j)\frac{x^{n-j}}{(n-j)!} = \sum_{k=2}^{n-1} s(n-2, n-k)\frac{x^k}{k!} \qquad (4.4)$$

must have only real zeros.

This last expression is just a positive multiple of the probability generating function of L_n which, using Eq. (2.1), can be written more explicitly than Eq. (3.4) as

$$\phi_n(x) = \sum_{2 \le k \le n-1} x^k \frac{n!}{n^{n-2}k!} s(n-2, n-k) \ . \qquad (4.5)$$

By comparison with Eq. (4.4) we see the probability generating function $\phi_n(x)$ has only real roots. This proves that L_n is equal to the sum of N independent Bernoulli random variables. One intriguing aspect of this representation is that there is (so far) no way to establish this representation except by the analytical method of analysis of roots. There is no graph theoretic interpretation of the Bernoulli summands.

5. MOMENT BOUNDS UNDER GIBBS' MODEL

The factorization of ϕ not only provides a representation of L_n under the uniform model, but it also yields a Bernoulli representation for L_n under all of the distributions in the Gibbs family. To make good use of the representation, we need to obtain some rudimentary bounds on the mean $E_\beta L_n$ and variance $\text{var}_\beta L_n$.

Under the Gibbs measure $\mu_\beta(s)$, where $f(s)$ is the number of leaves of the labeled tree s, the expression $P_\beta(L_n = k) = e^{-\beta k}T(n,k)/Z(\beta)$ lets us write the probability generating function $\phi_n(x,\beta) = E_\beta x^{L_n}$ in terms of the function $\phi_n(x) = \phi_n(x,0)$ via Eqs. (3.4) and (3.5) as

$$\phi_n(x,\beta) = \sum_{2 \le k \le n-1} x^k e^{-k\beta} T(k,n)/Z(\beta)$$
$$= \phi_n(xe^{-\beta})/\phi_n(e^{-\beta}) \ . \qquad (5.1)$$

The fact that Eq. (4.5) has only real roots implies the same for $\phi_n(x,\beta)$. If $L_{n,\beta}$ denotes a random variable with probability generating function $\phi_n(x,\beta)$, then again we can represent $L_{n,\beta}$ as the sum of n independent Bernoulli random variables X_i, $1 \le i \le n$. If the nonpositive numbers $(-r_i)$, $1 \le i \le n$, are the roots of $\phi_n(x,0)$, the relation $\phi_n(x,\beta) = \phi_n(xe^{-\beta},0)/\phi_n(e^{-\beta})$ shows that the roots of $\phi_n(x,\beta)$ are just $(-e^\beta r_i)$, $1 \le i \le n$. This gives us an easy way to relate moments and other properties of $L_{n,\beta}$ to the results obtained in the uniform case of L_n.

Factoring $\phi_n(x,\beta)$ as we did $\Psi_n(x)$ in Eq. (4.2), we see that $L_{n,\beta}$ is a sum of independent Bernoulli's X_i, $1 \le i \le n$, where $P(X_i = 1) = (1 + e^\beta r_i)^{-1}$, $1 \le i \le n$, and $(-r_i)_{i=1}^n$ are the n nonpositive real roots of $\phi_n(x)$. We have therefore that

$$EL_{n,\beta} = \sum_{i=1}^n (1 + e^\beta r_i)^{-1} \tag{5.1a}$$

and

$$\operatorname{var} L_{n,\beta} = \sum_{i=1}^n e^\beta r_i (1 + e^\beta r_i)^{-2} . \tag{5.1b}$$

We can now show $\operatorname{var}_\beta L_n$ diverges at a linear rate since for all $-\infty < \beta < \infty$, we have the crude bound

$$\operatorname{var}_\beta L_n = \sum_{i=1}^n e^\beta r_i (1 + e^\beta r_i)^{-2} \ge e^{-|\beta|} \sum_{i=1}^n r_i (1 + r_i)^{-2} . \tag{5.2}$$

This inequality is enough to permit us to prove a linear rate of divergence of $\operatorname{var}_\beta L_n$ because of the relation

$$\sum_{i=1}^n r_i (1 + r_i)^{-2} = \operatorname{var}_0 L_n \sim n(e - 2)e^{-2}$$

which comes from Eq. (1.2).

It is natural to expect that as n tends to infinity we have

$$E_\beta L_n \sim c(\beta)n \quad \text{and} \quad \operatorname{var}_\beta L_n \sim v(\beta)n \tag{5.3}$$

for some constants $c(\beta)$ and $v(\beta)$. These are natural conjectures and progress toward them would seem to begin with generalizing Rényi's Eq. (2.2). For the present we content ourselves with just having shown the linear divergence of $\operatorname{var}_\beta L_n$ since that is the essential result required for a central limit theorem.

6. CONSEQUENCES OF THE BERNOULLI APPROXIMATION

For a sum S_n of Bernoulli random variables, the Lindeberg condition becomes simply the condition that $\operatorname{var} S_n \to \infty$. From the linear divergence of $\operatorname{var}_\beta L_n$, we can then conclude that $(L_n - E_\beta L_n)(\operatorname{var}_\beta L_n)^{-1/2}$ is asymptotically normal.

To do a bit better, we first note that if X_i is a Bernoulli random variable

then $|X_i - EX_i| < 1$, so $E|X_i - EX_i|^3 < \text{var } X_i$. By the Berry–Essen theorem, we then have for all $n \geq 1$,

$$\left| P_\beta\left(\frac{(L_n - E_\beta L_n)}{\sqrt{\text{var } L_n}} \leq t \right) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2} \, dx \right| \leq A e^{|\beta|^{1/2}} n^{-1/2} \qquad (6.1)$$

where A is a universal constant (see e.g., Petrov [17], p. 111 or Feller [5], p. 544)..

The Bernoulli representation of L_n can be applied to give numerous additional inequalities of the central limit and large deviation type. In particular, one can apply the results of Bender [2] or Canfield [4]. We are content to stop with the basic bound of Eq. (6.1) because it is the most likely to be used and because it completes the program of providing an effortless refinement of the original central limit theorem of Rényi [20].

7. GENERATING RANDOM SAMPLES

To generate a tree which satisfies the law

$$P_\beta(L_n = k) = e^{-\beta k} T(n,k)/Z(\beta), \qquad 2 \leq k \leq n - 1 \qquad (7.1)$$

for the number of leaves, we can adapt a general technique of Metropolis *et al.*, [14] for generating an observation from a Gibbs distribution.

If $\beta = 0$, then the task of generating a tree satisfying Eq. (7.1) is easy; one can just choose the $n - 2$ coordinate of the Prüfer code by making uniform random selections from $\{1, 2, \ldots, n\}$. To expand this beyond $\beta = 0$, we let S denote the set of Prüfer codes; and, for $s \in S$, we define $f(s)$ as the number of leaves of the tree associated with s. In terms of the Prüfer code, we can equivalently express the defining property of $f(s)$ by noting that $n - f(s) - 2$ is the number of distinct integers in the code s.

We make S into a graph by considering an edge to exist between a pair of Prüfer codes s and s' if the codes are equal except in one coordinate. The degree of any vertex of the resulting graph is thus $(n - 1)(n - 2)$ since there are $n - 2$ coordinates which can be changed to $n - 1$ other values.

We now let $\alpha(s,s') = \min\{1, e^{-\beta\Delta}\}$, where $\Delta = f(s') - f(s)$. Further, let $N(s)$ denote the set of neighbors of s in the graph S, and define a Markov transition matrix on S by

$$p(s,s') = \begin{cases} \alpha(s,s')/(n-1)(n-2) & \text{if } s' \in N(s) \\ 1 - \displaystyle\sum_{s' \in N(s)} p(s,s') & \text{if } s = s' \\ 0 & \text{otherwise} \end{cases} \qquad (7.2)$$

It is easy to verify that the Markov chain on S determined by $\{p(s,s')\}$ has stationary distribution $\mu_\beta(s) = e^{-\beta f(s)}/Z(\beta)$; one just checks the reversibility condition

$$\mu_\beta(s)p(s,s') = \mu_\beta(s')p(s',s) \ .$$

The suggestion of Metropolis *et al.*, [14] is that one generates an $s \in S$ in accordance with the probability law μ_β by proceeding as follows:

Begin with an arbitrary s_o and make N random transitions under $\{p(s,s')\}$. If N is sufficiently large, then the ergodic property of Markov chains would say that the terminal state s is reached with probability which is close to $\mu_\beta(s)$.

The charm of this method is that one never has to calculate $Z(\beta)$; and, in our case, one does not even have to make a random choice from a set larger than $(n-1)(n-2)$. The main theoretical drawback to the method is that it is hard to be confident of the proper choice of N. A minimal value should at least be a substantial multiple of the diameter of the graph S. For the current problem, we suggest an N at least as large as $n \log n$, but one might more properly choose $N = n^2$. A better understanding of the principles which might more authoritatively guide the choice of N can be found in Aldous [1].

A fortunate property of our problem of generating random trees from the Gibbs distribution is that there is an alternative method which is fast and exact. A benefit of this method is that it may provide a tool for pursuing a better empirical understanding of the Metropolis method.

Although $\mu_\beta(s) = e^{-\beta f(s)}/Z(\beta)$ was developed in the general framework of Gibbs' measures, the Gibbs family of trees is really much simpler than the general case where $Z(\beta)$ is typically intractable. The fact that $T(n,k) = s(n-2, n-k)n!/k!$ and the classical recursion $s(n,k) = ks(n-1, k) + s(n-1, k-1)$ for Stirling numbers of the second kind gives us an easy recursion for $T(n,k)$, namely,

$$kT(n,k)/n = kT(n-1, k) + (n-k)T(n-1, k-1) \qquad (7.3)$$

and this relation makes the calculation of $Z(\beta)$ easy.

To generate a Prüfer code satisfying the probability law μ_β, we first choose a random element k of $\{2, 3, \ldots, n-1\}$ according to discrete distribution $\{e^{-\beta k}T(n,k)/Z(\beta)\}_{k=2}^{n-1}$. We then choose a random k-subset $A \subset \{1, 2, \ldots, n\}$ from the uniform distribution on k-sets (see e.g., Nijenhuis and Wilf [16]). Finally, for each of the $n-2$ coordinates of the Prüfer code s, we choose an element at random from the uniform distribution on $\{2, 3, \ldots, n-1\} - A$.

8. OTHER GIBBS MODELS AND CONCLUDING REMARKS

The Gibbs model method of using a potential function to produce an exponential family of probability measures is a rich and powerful tool. It brings physical intuition and statistical technique into focus on problems which might otherwise seen amenable only to methods of enumeration. The Gibbs approach is one which we can expect to be applied in many combinatorial contexts. It seems particularly important in the construction of discrete models where an underlying combinatorial structure requires a probability law which cannot be easily constructed using independent random variables.

The most ready targets for the full program which was applied here for random trees are those combinatorial problems for which it is already known that an associated generating function has only real roots.

The best known candidate, much better known than Harper's Stirling numbers, is the problem of choosing an element at random from the power set of a finite set. There, of course, the Gibbs procedure just boils down to the extension of the binomial distribution with $p = \frac{1}{2}$ to the case of general p.

Let us consider in more detail the first nontrivial case, the Stirling numbers of the second kind. We recall that $s(n,k)$ is the number of ways of partitioning an n set into k non-empty subsets. We define a potential function f on the set S of all partitions of $\{1, 2, \ldots, n\}$ by letting $f(s)$ be the number of elements of the partition s. We note that the Gibbs measure $\mu_\beta(s)$ then yields a probability law for T_n, the number of classes in a partition s chosen in accordance with μ_β. The probability generating function for T_n is

$$\phi_{n,\beta}(x) = Ex^{T_n} = \sum_{k=1}^{n} e^{-\beta k} x^k S(n,k)/Z(\beta)$$

where the partition function $Z(\beta)$ is determined by the relation $\phi_{n,\beta}(1) = 1$. Since Harper [9] showed $\phi_{n,o}(x)$ has only real roots, the whole program which was applied to L_n, the number of leaves of a random tree, can be applied to T_n. As a result, we have an exponential family of combinatorial objects which have a central limit theory and which are amenable to practical simulation.

The most far-reaching class of combinatorial objects to which the Gibbs program is immediately applicable is probably that of matching in a graph. Briefly, a matching of G is a set of edges, no pair of which share a vertex. A *k-matching* is a matching with k edges, and the number of k-matchings of G is denoted by $p(G,k)$ with $p(G,0) = 1$ by convention. The cardinality of the largest matching in G is denoted by $v(G)$, and the matching polynomial of G is

$$\mu(G,x) = \sum_{k=0}^{v(G)} (-1)^k p(G,k) x^{n-2k} \tag{8.1}$$

An amazing number of classical numbers and polynomials are related to $p(G,k)$ and $\mu(G,x)$ for special choices of G. For example, there is a G such that $p(G,k)$ becomes the Stirling number $s(n, n-k)$. The Hermite polynomial of degree n is the matching polynomial of the complete graph K_n, and even the Laguerre and Chebyshev polynomials are matching polynomials (see Godsil and Gutman [8], Godsil [6], and Godsil [7]).

Heilmann and Lieb [10] established the remarkable result that for any G the polynomial $\mu(G,x)$ has only real roots, and hence the matching generating function of G

$$\beta(G,x) = \sum_{k=0}^{v(G)} p(G,k) x^k \tag{8.2}$$

must also have only real roots. Godsil [6] built on the result of Heilmann and Lieb and established that if Z_n is the cardinality of a matching chosen uniformly from the set of all matchings of G_n, then Z_n is asymptotically normal, provided only that $\{G_n\}_{n=1}^{\infty}$ is a sequence of graphs, each regular with degree d, and the number of vertices of G_n increases with n.

The Gibbs family which was developed here for labeled trees can be developed just as easily for the number of matchings in a graph. Since this construction brings tools of statistical mechanics, the exponential family, and the Monte Carlo method into focus on problems as *a priori* remote as the coefficient of the Hermite polynomial and the number of leaves of a random tree, the Gibbs method would seem to merit much further study in combinatorics.

Acknowledgments

Note

1. Here, of course, a tree is a connected graph without cycles and a labeled tree is one with labeled vertices. Also, to avoid trivial exceptions we will always assume $n \geq 3$.

References

1. Aldous, D. (1986). On the Markov chain simulation method for uniform combinatorial distributions and simulated annealing. Technical Report No. 60, Department of Statistics, University of California, Berkeley, California.
2. Bender, E. A. (1973). Central and local limit theorems applied to asymptotic enumeration. *J. Combinatorial Theory A* 15: 91-111.
3. Cayley, A. (1889). A theorem on trees. *Quarterly Journal of Pure and Applied Mathematics* 23: 376-378.
4. Canfield, E. R. (1980). Application of the Berry-Essen inequality to combinatorial estimates. *J. Combinatorial Theory A* 28: 17-25.
5. Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, Vol. II, Second Edition, Wiley, New York, p. 544.
6. Godsil, C. D. (1981a). Hermite polynomials and a duality relation for matching polynomials. *Combinatorica* 1: 257-262.
7. Godsil, C. D. (1981b). Matching behavior is asymptotically normal. *Combinatorica* 1: 369-376.
8. Godsil, C. D. and Gutman, I. (1981). On the theory of matching polynomials. *J. Graph Theory* 5: 137-144.
9. Harper, L. H. (1967). Stirling behavior is asymptotically normal. *Ann. Math. Statis.* 38: 410-414.
10. Heilmann, O. J. and Lieb, E. H. (1972). The theory of monomer-dimer systems. *Comm. Math. Phys.* 22: 190-232.
11. Kilpatrick, S., Gelatt, C., and Vecchi, M. (1983). Optimization by simulated annealing. *Science* 220: 671-680.
12. Knuth, D. (1968). Another enumeration of trees. *Canadian J. of Math.* 20: 1077-1086.
13. Lovász, L. (1979). *Combinatorial Problems and Exercises*, North-Holland, Amsterdam, The Netherlands, p. 29.

14. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equations of state calculation by fast computing machines. *J. Chem. Physics* 21: 1087–1092.
15. Moon, J. W. (1970). *Counting Labeled Trees* Canadian Mathematical Monographs, No. 1, Alberta, Canada, p. 5.
16. Nijenhuis, A. and Wilf, H. (1978). *Combinatorial Algorithms*, Academic Press, New York, p. 35–39.
17. Petrov, V. V. (1975). *Sums of Independent Random Variables*, Springer-Verlag, New York, p. 111.
18. Pólya, G. and Szegö, G. (1976). *Problems and Theorems in Analysis II*, Springer-Verlag, New York, p. 45.
19. Prüfer, A. (1918). Neuer Beweis eines Satzes uber Permutationen. *Archiv für Mathematik und Physik* 27: 142–144.
20. Rényi, A. (1959). Some remarks on the theory of trees. *MTA Mat. Kut. Int. Kozl.* 4: 73–85.