

Two Problems in Asymptotic Combinatorics

Rod Canfield

`erc@cs.uga.edu`

Department of Computer Science
University of Georgia

Emory Combinatorics Seminar
November 6, 2009

Thanks to Coauthors

Part I, Stirling numbers: Carl Pomerance

Part II, Integer matrices: Brendan McKay

Stirling Numbers

$S(n, k)$ (a Stirling number of the second kind) equals the number of partitions of an n -set into k (nonempty, pairwise disjoint) blocks.

Recursion:

$$S(n + 1, k) = kS(n, k) + S(n, k - 1)$$

Implies strict log-concavity:

$$S(n, k)^2 \geq \left(1 + \frac{3}{k}\right) S(n, k + 1) S(n, k - 1)$$

Location of Maximum

There is a unique index K_n

$$S(n, 1) < \cdots < S(n, K_n) \geq S(n, K_n + 1) > \cdots > S(n, n).$$

n					
1	1				
2	1	1			
3	1	3	1		
4	1	7	6	1	
5	1	15	25	10	1

Exceptional Integers

Define the *exceptional set* E :

$$E = \{n : S(n, K_n) = S(n, K_n + 1)\}$$

Problem.

$$E \stackrel{?}{=} \{2\}$$

First Upper Bound

Let $E(x)$ be the associated counting function:

$$E(x) = \{n : n \leq x \text{ and } n \in E\}$$

Using the asymptotic formula (Harper)

$$K_n \sim n / \log n$$

we can prove

$$E(x) = O(x / \log x)$$

Proof Sketch

Recursion + log-concavity imply

$$K_{n+1} \in \{K_n, K_n + 1\}$$

$$n \in E \implies K_{n+1} = K_n + 1$$

Harper's Formula

Roots real and negative:

$$\sum_{k=1}^n S(n, k)x^k = x \prod_{j=1}^{n-1} (x + r_j^{(n)})$$

X_1, \dots, X_n independent Bernoulli r.v.'s

$$B_n^{-1} S(n, k) = \text{Prob} \left\{ \sum_i X_i = k \right\}$$

By Darroch (1964) the mean and mode differ by at most 1!

The Mean

What is it

$$B_n^{-1} \sum_{k=1}^n kS(n, k) = ??$$

The recursion

$$S(n+1, k) = kS(n, k) + S(n, k-1)$$

implies (summing)

$$B_{n+1} = \sum_{k=1}^n kS(n, k) + B_n$$

$$\mu_n = B_{n+1}/B_n - 1$$

Estimating Bell

Exponential generating function:

$$\sum_{n=1}^{\infty} \frac{B_n}{n!} z^n = \exp(e^x - 1)$$

Cauchy integral formula:

$$[x^n]F(x) = \oint_{|z|=r} \frac{F(z)}{z^{n+1}} dz$$

Moser and Wyman (1955): Let $re^r = n$; then,

$$B_n \sim \frac{1}{e\sqrt{\log n}} \exp(nr - n + n/r)$$

Hayman's Generalisation

W.K. Hayman, A generalisation of Stirling's formula, 1956

$$[x^n] \exp(f(x)) = \frac{\exp(f(r))}{r^n \sqrt{2\pi B}} (1 + \dots)$$

$$f(re^{i\theta}) = f(r) + i\theta r f'(r) - B\theta^2/2 + \dots$$

$$r f'(r) = n$$

Hayman's Proof

$$\begin{aligned} [x^n] \exp(f(x)) &= \frac{1}{2\pi r^n} \int_{-\pi}^{\pi} \exp(f(re^{i\theta}) - ni\theta) d\theta \\ &= \frac{\exp(f(r))}{2\pi r^n} \int_{-\delta}^{\delta} e^{-B\theta^2/2} (1 + \dots) d\theta \\ &\quad + O(r^{-n}) \max_{\delta \leq |\theta| \leq \pi} |\exp(f(re^{i\theta}))| \end{aligned}$$

Property of E

$$k = e^r = k + O(1)$$

$$S(n, k) = \frac{(e^r - 1)^k}{k!} \frac{n!}{r^n} (2\pi k B)^{-1/2} (1 + \dots)$$

For $n \in E$

$$e^r = \text{INTEGER} + 1/2 + \frac{1/2}{r+1} + R(r)e^{-r} + O_*(e^{-2r})$$

Three Together

Suppose X is large, and

$$\left| [X, X + X^{1/3-\epsilon}] \cap E \right| \geq 3$$

Say $n + l_i, i = 0, 1, 2$;

$$0 = l_0 < l_1 < l_2$$

Have also r_i

$$e^{r_i} = e^r + \frac{l_i}{r+1} - \frac{l_i^2}{2(r+1)^3 e^r} + O_*(l_i^3 e^{-2r})$$

Likewise, $\frac{1/2}{r_i+1}, \dots$

Consequence

$$e^{r_i} = \text{INTEGER} + 1/2 + \frac{1/2}{r_i + 1} + R(r_i)e^{-r_i} + O_*(e^{-2r})$$

$(l_1 - l_2), +l_2, -l_1$ combination:

$$\begin{aligned} & \frac{l_1 l_2^2 - l_2 l_1^2}{2(r+1)^3 e^r} + O_*\left(\frac{l_2 l_1^3 + l_1 l_2^3}{e^{2r}}\right) \\ &= \text{INTEGER} + O_*\left(\frac{l_1 l_2^2 + l_2 l_1^2}{e^{2r}}\right) \end{aligned}$$

Contradiction

Theorems, 2001

$$E(x) = O(x^{2/3+\epsilon})$$

$$E(10^6) = 1$$

$$E(x) = O(x^{3/5+\epsilon})$$

Huxley, Integer points close to a curve (1999)

Better Theorem

Via $S(x, y)$ for complex x, y

Kemkes, Merlini, and Richmond (2008)

$$E(x) = O(x^{1/2+\epsilon})$$

Bombieri & Pila, The number of integral points on arcs and ovals (1989)

Part II, Integer Matrices

Definition. Let m, s, n, t be integers with $ms = nt$

$$M(m, s; n, t) = \#\{A : \begin{array}{l} A \text{ is an } m \times n \text{ matrix over } \{0, 1, \dots\}, \\ \sum_{j=1}^n A_{ij} = s \text{ for all } i, \\ \sum_{i=1}^m A_{ij} = t \text{ for all } j \end{array}\}.$$

Example

Let $m = 4, s = 15, n = 6, t = 10$

$$\begin{bmatrix} 2 & 3 & 0 & 0 & 10 & 0 \\ 1 & 2 & 3 & 4 & 0 & 5 \\ 5 & 3 & 1 & 1 & 0 & 5 \\ 2 & 2 & 6 & 5 & 0 & 0 \end{bmatrix}$$

Estimated count = $2.36e + 11$

99 % confidence interval = $[2.01e + 11, 2.71e + 11]$

Exact value: $M(4, 15; 6, 10) = 234, 673, 404, 860$

Three Items

Exact value

Confidence interval

Estimated value

Exact Value

$$M(m, s; n, t) = [(x_1 \cdots x_m)^s (y_1 \cdots y_n)^t] \prod_{j=1}^m \prod_{k=1}^n (1 - x_j y_k)^{-1}$$

Make right side poly: $(1 + x_j y_k + \cdots + (x_j y_k)^{\min(s,t)})$

$$[\vec{x}^{\vec{s}} \vec{y}^{\vec{t}}] Poly(\vec{x}, \vec{y}) = \frac{1}{q_1^m q_2^n} \sum_{\vec{x} \in \langle \alpha \rangle^m} \sum_{\vec{y} \in \langle \beta \rangle^n} \frac{Poly(\vec{x}, \vec{y})}{\vec{x}^{\vec{s}} \vec{y}^{\vec{t}}}$$

$$\langle \alpha \rangle = \{1, \alpha, \dots, \alpha^{q_1-1}\}$$

Use symmetry

Confidence Interval

X a random variable with mean μ variance σ^2

sample mean $X_N = N^{-1}(x_1 + \cdots + x_N)$

normal approximation: $[X_N \pm 3\sigma/N^{1/2}]$ is a 99% confidence interval for μ

approximate σ^2 by the sample variance

A Trick

Population: $\Omega = \{\omega_1, \dots, \omega_\ell\}$, with probabilities p_i

$$X(\omega_i) = \frac{1}{p_i}$$

$$E(X) = |\Omega|$$

Application to Enumeration of contingency tables: Chen, Diaconis, Holmes, Liu *JASA*(2005)

Sampling Algorithm

Number columns $0, \dots, n - 1$; choose one at the time

When choosing $x[i, j], 0 \leq i < m$

- j columns are complete
- target row sums are $r_i = \text{target}$

Set $x[i, j]$ as if it were the first part in a random composition of r_i with $n - j$ parts

If $\sum_i x[i, 1] = t$, accept and proceed to next column; else, try again

Probability p_ω of obtaining $\omega = x[i, j]$ equals ?

Three Estimates

Define the *density* λ by

$$\frac{s}{n} = \lambda = \frac{t}{m}.$$

(It is the average entry in the matrix.)

$$M(m, n; s, t) \approx \binom{mn(1 + \lambda) - 1}{\lambda mn}$$

$$M(m, n; s, t) \approx \binom{n + s - 1}{s}^m$$

$$M(m, n; s, t) \approx \binom{m + t - 1}{t}^n$$

A Good Estimate

$$M(m, n; s, t) \approx \frac{\binom{n+s-1}{s}^m \binom{m+t-1}{t}^n}{\binom{mn(1+\lambda)-1}{\lambda mn}}$$

A Good Estimate

$$M(m, n; s, t) \approx \frac{\binom{n+s-1}{s}^m \binom{m+t-1}{t}^n}{\binom{mn(1+\lambda)-1}{\lambda mn}}$$

I. J. Good, *Probability and the Weighing of Evidence*, Charles Griffin, London, 1950.

A Good Estimate

$$M(m, n; s, t) \approx \frac{\binom{n+s-1}{s}^m \binom{m+t-1}{t}^n}{\binom{mn(1+\lambda)-1}{\lambda mn}}$$

I. J. Good, *Probability and the Weighing of Evidence*, Charles Griffin, London, 1950.

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

A Good Estimate

$$M(m, n; s, t) \approx \frac{\binom{n+s-1}{s}^m \binom{m+t-1}{t}^n}{\binom{mn(1+\lambda)-1}{\lambda mn}}$$

I. J. Good, *Probability and the Weighing of Evidence*, Charles Griffin, London, 1950.

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

I. J. Good and J. F. Crook, The enumeration of arrays and a generalization related to contingency tables, *Discrete Math.* **19** (1977) 23–45.

Conjecture 1

If $m, n \rightarrow \infty$,

$$M(m, n; s, t) = \frac{\binom{n+s-1}{s}^m \binom{m+t-1}{t}^n}{\binom{mn(1+\lambda)-1}{\lambda mn}} \times \exp\left(\frac{1}{2} + o(1)\right)$$

Conjecture 2

Define $\Delta(m, s; n, t)$ by

$$\begin{aligned} M(m, n; s, t) &= \frac{\binom{n+s-1}{s}^m \binom{m+t-1}{t}^n}{\binom{mn(1+\lambda)-1}{\lambda mn}} \\ &\times \left(\frac{m+1}{m}\right)^{(m-1)/2} \left(\frac{n+1}{n}\right)^{(n-1)/2} \\ &\times \exp\left(-\frac{1}{2} + \frac{\Delta(m, s; n, t)}{m+n}\right). \end{aligned}$$

Then, $0 < \Delta(m, s; n, t) < 2$.

Evidence

Theorem. Conjecture 1 is correct provided that $m, n \rightarrow \infty$ in such a way that

$$\frac{(1 + 2\lambda)^2}{4\lambda(1 + \lambda)} \left(1 + \frac{5m}{6n} + \frac{5n}{6m} \right) \leq a \log n, \quad a < 1/2.$$

Proof begins with

$$M(m, s; n, t) = \frac{1}{(2\pi i)^{m+n}} \times \oint \cdots \oint \frac{\prod_{j,k} (1 - x_j y_k)^{-1}}{x_1^{s+1} \cdots x_m^{s+1} y_1^{t+1} \cdots y_n^{t+1}} dx_1 \cdots dx_m dy_1 \cdots dy_n$$

More Evidence

Exact Calculations

Several thousands of $(m, s; n, t)$ with $m, n \leq 30$

Statistical Calculations

High degree of confidence for many larger values