Two Problems in Asymptotic Combinatorics

Rod Canfield
erc@cs.uga.edu

Department of Computer Science
University of Georgia

Ga Tech Combinatorics Seminar
April 2, 2010
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 $K_n$

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

<table>
<thead>
<tr>
<th>$n \backslash k$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>7</td>
<td>6</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>15</td>
<td>25</td>
<td>10</td>
<td>1</td>
</tr>
</tbody>
</table>
Define the *exceptional set* $E$:

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

Problem.

$$E \overset{?}{=} \{ 2 \}$$
Let $E(x)$ be the associated counting function:

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

Using the asymptotic formula (Harper)

$$K_n \sim \frac{n}{\log n}$$

we can prove

$$E(x) = O\left(\frac{x}{\log x}\right)$$
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:

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

Let \( B_n := \sum_k S(n, k) \), the \( n \)th Bell number.

\( X_1, \ldots, X_n \) 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

\[
\text{mean} = \mu_n = B_n^{-1} \sum_{k=1}^{n} kS(n, k) = \text{??}
\]

The recursion

\[
S(n + 1, k) = kS(n, k) + S(n, k - 1)
\]

implies (summing)

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

\[
\mu_n = \frac{B_{n+1}}{B_n} - 1
\]
Estimating Bell

Exponential generating function:

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

Cauchy integral formula:

$$[x^n] F(x) = \frac{1}{2\pi i} \oint_{|z|=r} \frac{F(z)}{z^{n+1}} dz$$

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

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

Let $r$ be defined (again) by $re^r = n$; then,

$$K_n \in \{[e^r - 1], \lceil e^r - 1 \rceil\}$$

for all sufficiently large $n$,

and for $1 \leq n \leq 1200$. 
Stirling Second Asymptotic

\[ \sum_{n=k}^{\infty} S(n, k) \frac{x^n}{n!} = \frac{(e^x - 1)^k}{k!} \]

\[ \frac{e^{re^{i\theta}} - 1}{e^r - 1} = \exp \left( Ai\theta - \frac{1}{2} B\theta^2 + O(r|\theta|^3) \right) , \]

\[ A = r + O(e^{-r}), \quad B = r + O(e^{-r}) \]
Property of $E$

\[ k = e^r + O(1) \]

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

For $n \in E$

\[ e^r = \text{INTEGER} + 1/2 + \frac{1/2}{r + 1} + R(r) e^{-r} + O_*(e^{-2r}) \]
Suppose $X$ is large, and

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

Say $n + \ell_i, i = 0, 1, 2$;

$$0 = \ell_0 < \ell_1 < \ell_2$$

Have also $r_i$

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

Likewise, $\frac{1/2}{r+1}, \ldots$
Consequence

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

\((\ell_1 - \ell_2), +\ell_2, -\ell_1\) combination:

\[ \frac{\ell_1 \ell_2^2 - \ell_2 \ell_1^2}{2(r + 1)^3 e^r} + O^*\left(\frac{\ell_2 \ell_1^3 + \ell_1 \ell_2^3}{e^{2r}}\right) \]

\[ = \text{INTEGER} + O^*\left(\frac{\ell_1 \ell_2^2 + \ell_2 \ell_1^2}{e^{2r}}\right) \]

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)
Erdős’ Theorem

\[ \Pi_{n,s} = \text{product 1 to } n, \text{ taken } s \text{ at time} \]

for prime \( p \) with \( n/(k+1) < p \leq n/k \) we have
\[ \Pi_{n,n-k} \not\equiv 0 \mod p, \text{ but} \]
\[ \Pi_{n,n-r} \equiv 0 \mod p \text{ for } r < k \]

P. Erdős, On a conjecture of Hammersley, (1953).
Definition. Let $m, s, n, t$ be integers with $ms = nt$

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

<table>
<thead>
<tr>
<th>Exact value</th>
<th>Confidence interval</th>
<th>Estimated value</th>
</tr>
</thead>
</table>

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: \((1 + x_j y_k + \cdots + (x_j y_k)^{\min(s, t)})\)

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

\[
\langle \alpha \rangle = \{1, \alpha, \ldots, \alpha^{q_1-1}\}
\]

Use symmetry
Confidence Interval

$X$ a random variable with mean $\mu$ variance $\sigma^2$

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 $\mu$

approximate $\sigma^2$ by the sample variance
A Trick

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

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

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

Sampling Algorithm

Number columns 0, \ldots, 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 $p_\omega$ of obtaining $\omega = x[i, j]$ equals ?
Three Estimates

Define the density $\lambda$ by

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

(It is the average entry in the matrix.)

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

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

$$M(m, n; s, t) \approx \left( \frac{m + t - 1}{t} \right)^n$$
A Good Estimate

\[ M(m, n; s, t) \approx \frac{(n+s-1)^m (m+t-1)^n}{\binom{mn(1+\lambda)-1}{\lambda mn}} \]
A Good Estimate

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

A Good Estimate

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


A Good Estimate

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


Conjecture 1

If \( m, n \to \infty \),

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

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

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

Then, $0 < \Delta(m, s; n, t) < 2$. 
Theorem. Conjecture 1 is correct provided that $m, n \to \infty$ 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