Analytic Information Theory

W. Szpankowski
Department of Computer Science
Purdue University
W. Lafayette, IN 47907

December 8, 2011

Dedicated to PHILIPPE FLAJOLET

PARIS, 2011

Center for
Science of Information
NSF Science & Technology Center

Outline

1. Shannon Legacy
2. Analytic Combinatorics + IT = Analytic Information Theory
3. The Redundancy Rate Problem
   (a) Universal Memoryless Sources
   (b) Universal Renewal Sources

P. Flajolet and W.S., Analytic Variations on Redundancy Rates of Renewal Processes *IEEE Trans. Information Theory*, 48, 2911 -2921, 2002.

Outline

1. Shannon Legacy
2. Analytic Combinatorics + IT = Analytic Information Theory
3. The Redundancy Rate Problem
   (a) Universal Memoryless Sources
   (b) Universal Renewal Sources

P. Flajolet and W.S., Analytic Variations on Redundancy Rates of Renewal Processes *IEEE Trans. Information Theory*, 48, 2911 -2921, 2002.

P. Mathys and P. Flajolet, Q-ary collision resolution algorithms in random-access systems with free or blocked channel access. *IEEE Transactions on Information Theory*, IT-31,217-243, March 1985.

G. Fayolle, P. Flajolet, and M. Hofri, On a functional equation arising in the analysis of a protocol for a multi-access broadcast channel. *Advances in Applied Probability*, 18:441-472, 1986.

J. Kieffer, P. Flajolet, and EH. Yang, Data compression via binary decision diagrams. *2000 IEEE International Symposium on Information Theory*, 296. Sorento, 2000.

Outline

1. Shannon Legacy
2. Analytic Combinatorics + IT = Analytic Information Theory
3. The Redundancy Rate Problem
   (a) Universal Memoryless Sources
   (b) Universal Renewal Sources

P. Flajolet and W.S., Analytic Variations on Redundancy Rates of Renewal Processes *IEEE Trans. Information Theory*, 48, 2911 -2921, 2002.

P. Mathys and P. Flajolet, Q-ary collision resolution algorithms in random-access systems with free or blocked channel access. *IEEE Transactions on Information Theory*, IT-31,217-243, March 1985.

G. Fayolle, P. Flajolet, and M. Hofri, On a functional equation arising in the analysis of a protocol for a multi-access broadcast channel. *Advances in Applied Probability*, 18:441-472, 1986.

J. Kieffer, P. Flajolet, and EH. Yang, Data compression via binary decision diagrams. *2000 IEEE International Symposium on Information Theory*, 296. Sorento, 2000.

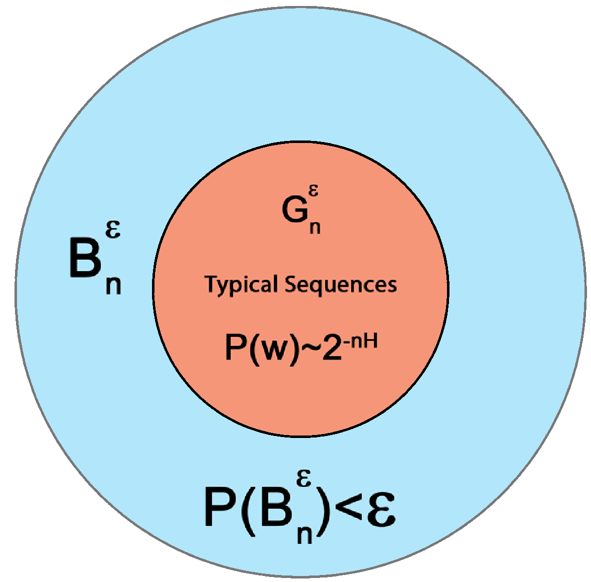
Algorithms:	are at the heart of virtually all computing technologies;
Combinatorics:	provides indispensable tools for finding patterns and structures;
Information:	permeates every corner of our lives and shapes our universe.

Three Theorems of Shannon

Theorem 1 & 3. (**Shannon 1948; Lossless & Lossy Data Compression**)

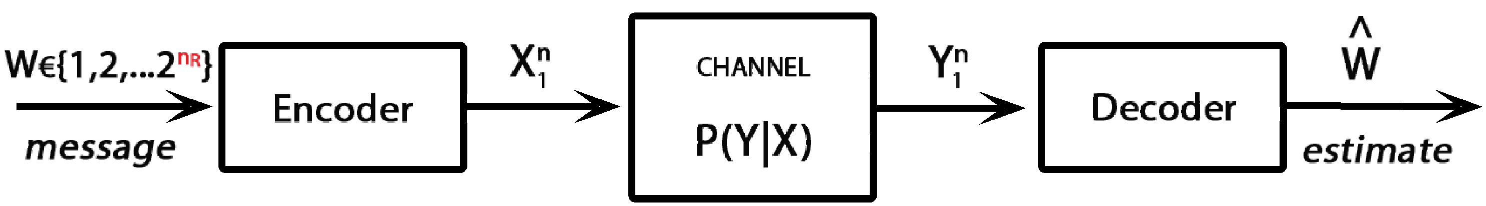
compression bit rate $\geq$ source entropy $H(X)$

for distortion level D:
lossy bit rate $\geq$ rate distortion function $R(D)$



Theorem 2. (**Shannon 1948; Channel Coding**)
In Shannon's words:

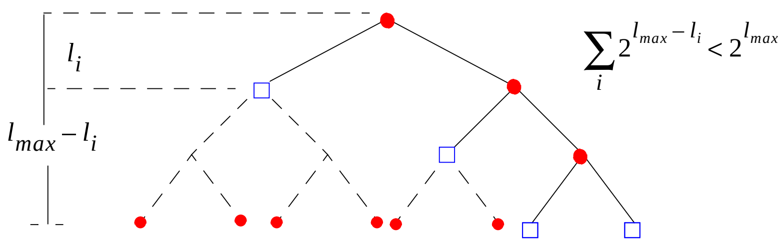
It is possible to send information at the capacity through the channel with as small a frequency of errors as desired by proper (**long**) encoding. This statement is not true for any rate greater than the capacity.



Theorem 1: Fundamental Limit

Prefix code is such that no codeword is a prefix of another codeword.

Kraft's Inequality: A prefix code iff lengths $\ell_1, \ldots, \ell_N$ satisfy[1]



$$\sum_{i=1}^{N} 2^{-\ell_i} \leq 1.$$

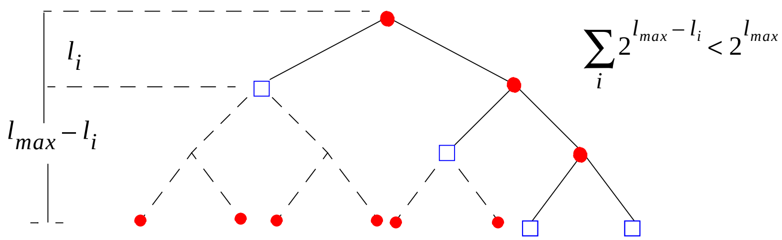
Shannon First Theorem: For any **prefix code** the average code length $\mathbf{E}[L(C,X)]$ cannot be smaller than the entropy $H(P)$:

$$\mathbf{E}[L(C,X)] \geq H(P) = -\sum_{x \in \mathcal{A}^*} P(x) \log P(x).$$

[1] Flajolet and Prodinger, "Level number of sequences for trees", *Disc. Math.*, 1987.

Theorem 1: Fundamental Limit

Prefix code is such that no codeword is a prefix of another codeword.
Kraft's Inequality: A prefix code iff lengths $\ell_1, \ldots, \ell_N$ satisfy[1]

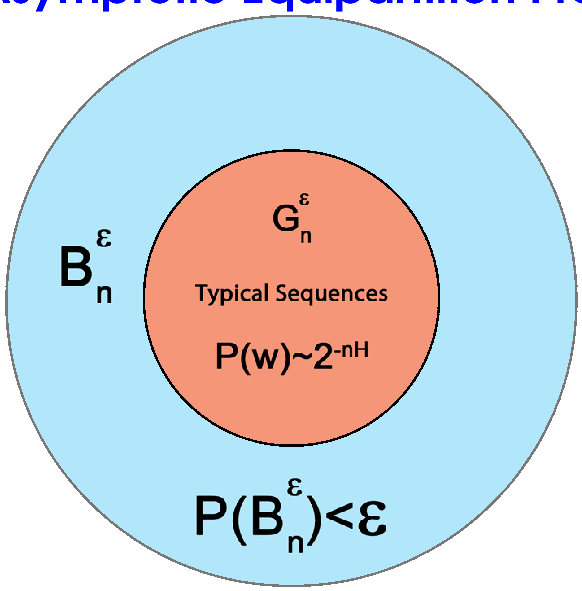


$$\sum_{i=1}^{N} 2^{-\ell_i} \leq 1.$$

Shannon First Theorem: For any **prefix code** the average code length $\mathbf{E}[L(C,X)]$ cannot be smaller than the entropy $H(P)$:

$$\mathbf{E}[L(C,X)] \geq H(P) = -\sum_{x \in \mathcal{A}^*} P(x) \log P(x).$$

Asymptotic Equipartition Property:



Shannon-McMilan-Breiman:

$$-\frac{1}{n} \log P(X_1^n) \to H(X) \quad \text{(pr.)}$$

$H(X)$ is the entropy rate.

Code Length: $\lceil -\log P(X_1^n) \rceil \sim nH(X)$.

AEP: Good set G_n^ε: $P(w) \sim 2^{-nH(X)}$

Post-Shannon Challenges

1. **Back off from infinity** (Ziv'97): Extend Shannon findings to finite size data structures (i.e., sequences, graphs), that is, develop information theory of various data structures beyond first-order asymptotics.

> **Claim**: Many interesting information-theoretic phenomena appear in the second-order terms.

2. **Science of Information**:[2] Information Theory needs to meet new challenges of current applications in

biology, communication, knowledge extraction, economics, . . .

to understand new aspects of information in:

structure, time, space, and semantics,

and

dynamic information, limited resources, complexity, representation-invariant information, and cooperation & dependency.

---

[2] Philippe's email about an article in the CACM (Feb. 2011) about Sol: "*That's a great one. The thing is intelligently done, does not over-sell. The writer is good too. That would almost make be believe in the project. :-) Congrats also for the photograph (without a cap—good!).*"

Outline Update

1. Shannon Legacy
2. Analytic Combinatorics + IT = Analytic Information Theory
3. The Redundancy Rate Problem

Analytic Combinatorics+IT=Analytic Information Theory

- In the **1997 Shannon Lecture** Jacob Ziv presented compelling arguments for "backing off" from first-order asymptotics in order to predict the behavior of real systems with finite length description.

- To **overcome** these difficulties, one may replace first-order analyses by non-asymptotic analysis, however, we propose to develop full asymptotic expansions and more precise analysis (e.g., large deviations, CLT).

- Following **Hadamard's precept**[3], we study information theory problems using techniques of complex analysis such as generating functions, combinatorial calculus, Rice's formula, Mellin transform, Fourier series, sequences distributed modulo 1, saddle point methods, analytic poissonization and depoissonization, and singularity analysis.

- This program, which applies complex-analytic tools to information theory, constitutes **analytic information theory**.[4]

- **Philippe** was the midwife and active participant of analytic information theory since mid 90's.

---

[3] The shortest path between two truths on the real line passes through the complex plane.

[4] Andrew Odlyzko: "Analytic methods are extremely powerful and when they apply, they often yield estimates of unparalleled precision."

Some Successes of Analytic Information Theory

- Wyner-Ziv Conjecture concerning the longest match in the WZ'89 compression scheme (W.S., 1993).
- Ziv's Conjecture on the distribution of the number of phrases in the LZ'78 (Jacquet & W.S., 1995, 2011).
- Redundancy of the LZ'78 (Savari, 1997, Louchard & W.S., 1997).
- Steinberg-Gutman Conjecture regarding lossy pattern matching compression (Luczak & W.S., 1997; Kieffer, Flajolet, Yang, 1998; Kontoyiannis, 2003).
- Precise redundancy of Huffman's Code (W.S., 2000) and redundancy of a fixed-to-variable no prefix free code (W.S. & Verdu, 2010).
- **Minimax Redundancy** for memoryless sources (Xie &Barron, 1997; W.S., 1998; W.S. & Weinberger, 2010), Markov sources (Risannen, 1998; Jacquet & W.S., 2004), and renewal sources (Flajolet & W.S., 2002; Drmota & W.S., 2004).
- Analysis of variable-to-fixed codes such as Tunstall and Khodak codes (Drmota, Reznik, Savari, & W.S., 2006, 2008, 2010).
- Entropy of hidden Markov processes and the noisy constrained capacity (Jacquet, Seroussi, & W.S., 2004, 2007, 2010; Han & Marcus, 2007).
- . . .

Outline Update

1. Shannon Legacy
2. Analytic Information Theory
3. Source Coding: The Redundancy Rate Problem
   (a) Universal Memoryless Sources
   (b) Universal Renewal Sources

Source Coding and Redundancy

Source coding aims at finding codes $C : \mathcal{A}^* \to \{0,1\}^*$ of the shortest length $L(C,x)$, either on *average* or for *individual sequences*.

Known Source P: The pointwise and maximal redundancy are:

$$\begin{aligned} R_n(C_n, P; x_1^n) &= L(C_n, x_1^n) + \log P(x_1^n) \\ R^*(C_n, P) &= \max_{x_1^n}\{R_n(C_n, P; x_1^n)\}(\geq 0). \end{aligned}$$

where $P(x_1^n)$ is the probability of $x_1^n = x_1 \cdots x_n$.

Unknown Source P: Following Davisson, the maximal minimax redundancy $R_n^*(\mathcal{S})$ for a family of sources $\mathcal{S}$ is:

$$R_n^*(\mathcal{S}) = \min_{C_n} \sup_{P \in \mathcal{S}} \max_{x_1^n} [L(C_n, x_1^n) + \log P(x_1^n)].$$

Shtarkov's Bound:

$$d_n(\mathcal{S}) := \log \sum_{x_1^n \in \mathcal{A}^n} \sup_{P \in \mathcal{S}} P(x_1^n) \leq R_n^*(\mathcal{S}) \leq \underbrace{\log \sum_{x_1^n \in \mathcal{A}^n} \sup_{P \in \mathcal{S}} P(x_1^n)}_{D_n(\mathcal{S})} + 1$$

Outline Update

1. Shannon Legacy
2. Analytic Information Theory
3. The Redundancy Rate Problem
   (a) Universal Memoryless Sources
   (b) Universal Renewal Sources

Maximal Minimax for Memoryless Sources

For a **memoryless source** over the alphabet $\mathcal{A} = \{1, 2, \ldots, m\}$ we have

$$P(x_1^n) = p_1{}^{k_1} \cdots p_m{}^{k_m}, \qquad k_1 + \cdots + k_m = n.$$

Then

$$\begin{aligned}
D_n(\mathcal{M}_0) &:= \sum_{x_1^n} \sup_{P(x_1^n)} P(x_1^n) \\
&= \sum_{x_1^n} \sup_{p_1,\ldots,p_m} p_1^{k_1} \cdots p_m^{k_m} \\
&= \sum_{k_1+\cdots+k_m=n} \binom{n}{k_1, \ldots, k_m} \sup_{p_1,\ldots,p_m} p_1^{k_1} \cdots p_m^{k_m} \\
&= \sum_{k_1+\cdots+k_m=n} \binom{n}{k_1, \ldots, k_m} \left(\frac{k_1}{n}\right)^{k_1} \cdots \left(\frac{k_m}{n}\right)^{k_m}.
\end{aligned}$$

since the (unnormalized) **likelihood distribution** is

$$\sup_{P(x_1^n)} P(x_1^n) = \sup_{p_1,\ldots,p_m} p_1^{k_1} \cdots p_m^{k_m} = \left(\frac{k_1}{n}\right)^{k_1} \cdots \left(\frac{k_m}{n}\right)^{k_m}$$

Generating Function for $D_n(\mathcal{M}_0)$

We write

$$D_n(\mathcal{M}_0) = \sum_{k_1+\cdots+k_m=n} \binom{n}{k_1,\ldots,k_m} \left(\frac{k_1}{n}\right)^{k_1} \cdots \left(\frac{k_m}{n}\right)^{k_m} = \frac{n!}{n^n} \sum_{k_1+\cdots+k_m=n} \frac{k_1^{k_1}}{k_1!} \cdots \frac{k_m^{k_m}}{k_m!}$$

Let us introduce a tree-generating function

$$B(z) = \sum_{k=0}^{\infty} \frac{k^k}{k!} z^k = \frac{1}{1-T(z)}, \qquad T(z) = \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} z^k$$

where $T(z) = ze^{T(z)}$ $(= -W(-z)$, **Lambert's** W-function) that enumerates all rooted labeled trees. Let now

$$D_m(z) = \sum_{n=0}^{\infty} z^n \frac{n^n}{n!} D_n(\mathcal{M}_0).$$

Then by the convolution formula

$$D_m(z) = [B(z)]^m - 1.$$

Asymptotics for FINITE m

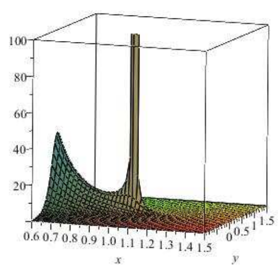
The function $B(z)$ has an algebraic singularity at $z = e^{-1}$, and

$$\beta(z) = B(z/e) = \frac{1}{\sqrt{2(1-z)}} + \frac{1}{3} + O(\sqrt{(1-z)}).$$

By Cauchy's coefficient formula

$$D_n(\mathcal{M}_0) = \frac{n!}{n^n}[z^n][B(z)]^m = \sqrt{2\pi n}(1 + O(1/n))\frac{1}{2\pi i}\oint \frac{\beta(z)^m}{z^{n+1}}dz.$$

For finite m, the singularity analysis of Flajolet and Odlyzko implies[5]



$$[z^n](1-z)^{-\alpha} \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)}, \quad \alpha \notin \{0, -1, -2, \ldots\}$$

$$\begin{aligned} R_n^*(\mathcal{M}_0) &= \frac{m-1}{2}\log\left(\frac{n}{2}\right) + \log\left(\frac{\sqrt{\pi}}{\Gamma(\frac{m}{2})}\right) + \frac{\Gamma(\frac{m}{2})m}{3\Gamma(\frac{m}{2}-\frac{1}{2})}\cdot\frac{\sqrt{2}}{\sqrt{n}} \\ &+ \left(\frac{3+m(m-2)(2m+1)}{36} - \frac{\Gamma^2(\frac{m}{2})m^2}{9\Gamma^2(\frac{m}{2}-\frac{1}{2})}\right)\cdot\frac{1}{n} + \cdots \end{aligned}$$

[5] Flajolet and Odlyzko, "Singularity Analysis of Generating Functions", *SIAMCOMP*, 1990.

Outline Update

1. Shannon Legacy
2. Analytic Information Theory
3. The Redundancy Rate Problem
   (a) Universal Memoryless Sources
   (b) Universal Renewal Sources

Renewal Sources

The **renewal process** $\mathcal{R}_0$ (introduced in 1996 by Csiszár and Shields) defined as follows:

- Let $T_1, T_2 \ldots$ be a sequence of i.i.d. positive-valued random variables with distribution $Q(j) = \Pr\{T_i = j\}$.
- In a **binary renewal sequence** the positions of the $1'$s are at the renewal epochs $T_0, T_0 + T_1, \ldots$ with runs of zeros of lengths $T_1 - 1, T_2 - 1, \ldots$.

Renewal Sources

The **renewal process** $\mathcal{R}_0$ (introduced in 1996 by Csiszár and Shields) defined as follows:

- Let $T_1, T_2 \ldots$ be a sequence of i.i.d. positive-valued random variables with distribution $Q(j) = \Pr\{T_i = j\}$.
- In a **binary renewal sequence** the positions of the 1's are at the renewal epochs $T_0, T_0 + T_1, \ldots$ with runs of zeros of lengths $T_1 - 1, T_2 - 1, \ldots$.

For a sequence

$$x_0^n = 10^{\alpha_1}10^{\alpha_2}1 \cdots 10^{\alpha_n}1\underbrace{0 \cdots 0}_{k^*}$$

define k_m as the number of i such that $\alpha_i = m$. Then

$$P(x_1^n) = [Q(0)]^{k_0}[Q(1)]^{k_1} \cdots [Q(n-1)]^{k_{n-1}} \Pr\{T_1 > k^*\}.$$

Renewal Sources

The **renewal process** $\mathcal{R}_0$ (introduced in 1996 by Csiszár and Shields) defined as follows:

- Let $T_1, T_2 \ldots$ be a sequence of i.i.d. positive-valued random variables with distribution $Q(j) = \Pr\{T_i = j\}$.
- In a **binary renewal sequence** the positions of the 1's are at the renewal epochs $T_0, T_0 + T_1, \ldots$ with runs of zeros of lengths $T_1 - 1, T_2 - 1, \ldots$.

For a sequence

$$x_0^n = 10^{\alpha_1}10^{\alpha_2}1 \cdots 10^{\alpha_n}1\underbrace{0 \cdots 0}_{k^*}$$

define k_m as the number of i such that $\alpha_i = m$. Then

$$P(x_1^n) = [Q(0)]^{k_0}[Q(1)]^{k_1} \cdots [Q(n-1)]^{k_{n-1}}\Pr\{T_1 > k^*\}.$$

Theorem 2 (Flajolet and W.S., 1998). [6] *Consider the class of renewal processes. Then*

$$R_n^*(\mathcal{R}_0) = \frac{2}{\log 2}\sqrt{cn} + O(\log n).$$

where $c = \frac{\pi^2}{6} - 1 \approx 0.645$.

---

[6] Flajolet and W.S., "Analytic Variations on Redundancy Rates of Renewal Processes" *IEEE IT*, 2002.

Maximal Minimax Redundancy

It can be proved that

$$r_{n+1} - 1 \le D_n(\mathcal{R}_0) \le \sum_{m=0}^{n} r_m$$

where $r_n = \sum_{k=0}^{n} r_{n,k}$ and

$$r_{n,k} \quad = \quad \sum_{\mathcal{P}(n,k)} \binom{k}{k_0 \ldots k_{n-1}} \left(\frac{k_0}{k}\right)^{k_0} \left(\frac{k_1}{k}\right)^{k_1} \cdots \left(\frac{k_{n-1}}{k}\right)^{k_{n-1}}$$

where $\mathcal{P}(n,k)$ is is the integer partition of n into k terms, i.e.,

$$n = k_0 + 2k_1 + \cdots + nk_{n-1}, \quad k = k_0 + \cdots + k_{n-1}.$$

But we shall study $s_n = \sum_{k=0}^{n} s_{n,k}$ where

$$s_{n,k} = e^{-k} \sum_{\mathcal{P}(n,k)} \frac{k^{k_0}}{k_0!} \cdots \frac{k^{k_{n-1}}}{k_{n-1}!}, \qquad \frac{r_{n,k}}{s_{n,k}} = \frac{k!}{k^k e^{-k}}$$

since

$$S(z,u) = \sum_{k,n} s_{n,k} (u/e)^k z^n = \sum_{\mathcal{P}_{n,k}} z^{1k_0 + 2k_1 + \cdots} \left(\frac{u}{e}\right)^{k_0 + \cdots + k_{n-1}} \frac{k^{k_0}}{k_0!} \cdots \frac{k^{k_{n-1}}}{k_{n-1}!} = \prod_{i=1}^{\infty} \beta(z^i u)$$

Refined Main Results

Theorem 3 (Flajolet and W.S., 1998). *We have the following asymptotics* ($c = \frac{\pi^2}{6} - 1 \approx 0.645$)

$$s_n \sim \exp\left(2\sqrt{cn} - \frac{7}{8}\log n + O(1)\right),$$

$$\log r_n = \frac{2}{\log 2}\sqrt{cn} - \frac{5}{8}\log n + \frac{1}{2}\log\log n + O(1).$$

Asymptotic analysis is sophisticated and follows these steps:

- first, we transform r_n into s_n that we know how to handle and we know how to read back results for r_n from s_n;
- use combinatorial calculus to find the generating function of s_n, which turns out to be an infinite product of tree-functions $B(z)$ defined above;
- transform this product into a harmonic sum that can be analyzed asymptotically by the Mellin transform;
- obtain an asymptotic expansion of the generating function around $z = 1$ which is the starting point to get asymptotics of the coefficients;
- finally, estimate $R_n^*(\mathcal{R}_0)$ by the saddle point method.

Translating s_n into r_n

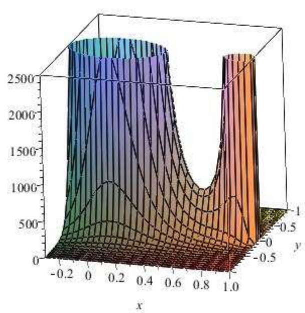
To compare s_n to r_n, we introduce the random variable K_n as follows

$$\Pr\{K_n = k\} = \frac{s_{n,k}}{s_n}.$$

Stirling's formula yields

$$\frac{r_n}{s_n} = \sum_{k=0}^{n} \frac{r_{n,k}}{s_{n,k}} \frac{s_{n,k}}{s_n} = \mathbf{E}[(K_n)! K_n^{-K_n} e^{K_n}] = \mathbf{E}[\sqrt{2\pi K_n}] + O(\mathbf{E}[K_n{}^{-\frac{1}{2}}]).$$

Lemma 1. *Let* $\mu_n = \mathbf{E}[K_n]$ *and* $\sigma_n^2 = \mathbf{Var}(K_n)$.



(by saddle point method)

$$s_n = [z^n] S(z,1) = [z^n] \exp\left(\tfrac{c}{1-z} + a \log \tfrac{1}{1-z}\right)$$

and $\mu_n = \frac{1}{4}\sqrt{\frac{n}{c}} \log \frac{n}{c} + o(\sqrt{n})$ *while* $\sigma_n^2 = O(n \log n) = o(\mu_n^2)$, *where* $c = \pi^2/6 - 1$, $d = -\log 2 - \frac{3}{8} \log c - \frac{3}{4} \log \pi$.

Thus

$$r_n = s_n \mathbf{E}[\sqrt{2\pi K_n}](1 + o(1)) = s_n \sqrt{2\pi \mu_n}(1 + o(1)).$$

Thank you, Philippe . . .

. . . for long standing support, friendship, and sharing your knowledge!

Merci au bon docteur Flajolet. We will miss you!