Information Theory by Analytic Methods: The Precise Minimax Redundancy*

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

March 1, 2001

*Research supported by NSF Grants NCR-9415491 and C-CR 9804760

Outline

1. Notations and Definitions
2. Shannon Lower Bound for the Average Code Length
3. Barron's Lemma
4. Definitions of Minimax Redundancy and Regret Functions
5. Analytic Information Theory
6. Shtarkov's Bound for the Maximal Minimax Redundancy
7. Making Shtarkov's Bound an Equality: Precise Minimax
   (a) Optimal Code Minimizing the Maximal Redundancy
   (b) Precise Maximal Minimax Result
   (c) Decomposition Theorem
8. The Average Minimax Redundancy
   (a) Gallager's Capacity Theorem
   (b) Optimal Prior
   (c) Conjecture on the Average Minimax Redundancy
9. Survey of Precise Results for the Redundancy Problem:
   (a) Shannon Code (average redundancy)
   (b) Generalized Shannon Code (maximal and average)
   (c) Huffman Code
   (d) Memoryless and Markovian Sources (minimax)
   (e) Renewal Processes
   (f) Lempel-Ziv Code
   (g) Tunstall's Code

10. Analytic Methods at Work:
   (a) Fourier Analysis and Distribution of Sequences Mod 1
       (*Redundancy of Shannon and Huffman Codes*)
   (b) Tree Generating Functions and Singularity Analysis
       (*Minimax Redundancy of a Memoryless Source*)
   (c) Mellin Transform and Saddle Point Method
       (*Redundancy of a Renewal Process*)
   (d) Analytic Depoissonization
       (*Lempel-Ziv Code*)

Notation and Definitions

The **redundancy-rate problem** OF **universal coding** FOR A CLASS OF SOURCES CONSISTS IN DETERMINING BY HOW MUCH THE ACTUAL CODE LENGTH EXCEEDS THE OPTIMAL (IDEAL) CODE LENGTH.

Definition: *A code*

$$C_n : \mathcal{A}^n \to \{0,1\}^*$$

is a mapping from the set $\mathcal{A}^n$ of all sequences of length n over the alphabet $\mathcal{A}$ to the set $\{0,1\}^*$ of binary sequences.

Given a probabilistic source model and a code C_n we let:

- $P(x_1^n)$ be the probability of the message $x_1^n = x_1 \dots x_n$,
- $L(C_n, x_1^n)$ be the code length for x_1^n,
- Entropy $H_n(P) = -\sum_{x_1^n} P(x_1^n) \lg P(x_1^n)$,

Information-theoretic quantities are expressed in binary logarithms written $\lg := \log_2$.

Kraft's Inequality

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

Lemma 1. [**Kraft's Inequality**] *For any prefix code (over a binary alphabet), the codeword lengths* $\ell_1, \ell_2, \ldots, \ell_m$ *satisfy the inequality*

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

Conversely, if codeword lengths satisfy this inequality, then one can build a prefix code.

Proof. This is an easy exercise on trees. Let $\ell_{\max}$ be the maximum codeword length. Observe that at level $\ell_{\max}$ some nodes are codewords, some are descendants of codewords, and some are neither. Since the number of descendants at level $\ell_{\max}$ of a codeword located at level ℓ_i is $2^{\ell_{\max}-\ell_i}$ (if root is at level 0 of the tree, then at level ℓ there are no more than 2^{ℓ} nodes), we obtain

$$\sum_{i=1}^{m} 2^{\ell_{\max}-\ell_i} \leq 2^{\ell_{\max}},$$

which is the desired inequality. The converse part is easy to prove, and we omit it here. ∎

Shannon's Lower Bound

Lemma 2. **[Shannon]** *For any code, the average code length* $\mathbf{E}[L(C_n, X_1^n)]$ *cannot be smaller than the entropy of the source* $H_n(P)$*, that is,*

$$\mathbf{E}[L(C_n, X_1^n)] \geq H_n(P).$$

Sketch of Proof: Let $K = \sum_{x_1^n} 2^{-L(x_1^n)} \leq 1$, and $L(C_n, x_1^n) := L(C_n)$. Then

$$\begin{aligned}
\mathbf{E}[L(C_n, X_1^n)] - H_n(P) &= \sum_{x_1^n \in \mathcal{A}^n} P(x_1^n) L(x_1^n) \\
&+ \sum_{x_1^n \in \mathcal{A}^n} P(x_1^n) \log P(x_1^n) \\
&= \sum_{x_1^n \in \mathcal{A}^n} P(x_1^n) \log \frac{P(x_1^n)}{2^{-L(x_1^n)}/K} - \log K \\
&\geq 0
\end{aligned}$$

since the first term is a divergence and cannot be negative (or $\log x \leq x - 1$) while $K \leq 1$ by Kraft's inequality.

Barron's Lemma

Observation: For every prefix code, there exists at least one source sequence $\tilde{x}_1^n$ such that

$$L(\tilde{x}_1^n) \geq -\log_2 P(\tilde{x}_1^n).$$

Indeed, if this is not true, then the Kraft inequality will not hold.

Lemma 3. [Barron] *Let $L(X_1^n)$ be the length of a fixed-to-variable codeword satisfying the Kraft inequality, where X_1^n is generated by a stationary ergodic source. For any sequence a_n of positive constants satisfying $\sum_n 2^{-a_n} < \infty$ the following holds*

$$\Pr\{L(X_1^n) \leq -\log P(X_1^n) - a_n\} \leq 2^{-a_n},$$

and conclude that

$$L(X_1^n) \geq -\log P(X_1^n) - a_n \qquad \text{(a.s)}.$$

Proof of Barron's Lemma

We argue as follows:

$$\begin{aligned}
\Pr\{L(X_1^n) < -\log_2 P(X_1^n) - a_n\} &= \sum_{x_1^n : P(x_1^n) < 2^{-L(x_1^n) - a_n}} P(x_1^n) \\
&\le \sum_{x_1^n : P(x_1^n) < 2^{-L(x_1^n) - a_n}} 2^{-L(x_1^n) - a_n} \\
&\le 2^{-a_n} \sum_{x_1^n} 2^{-L(x_1^n)} \\
&\le 2^{-a_n}.
\end{aligned}$$

The lemma follows from the Kraft inequality and the Borel-Cantelli Lemma.

Definitions of Redundancy

The **pointwise redundancy** $R_n(C_n, P; x_1^n)$ and the **average redundancy** $\bar{R}_n(C_n, P)$ are defined as

$$\begin{array}{rcl} R_n(C_n, P; x_1^n) & = & L(C_n, x_1^n) + \lg P(x_1^n) \ (\geq -a_n, \ \text{(a.s)}) \\ \bar{R}_n(C_n) & = & \mathbf{E}_{X_1^n}[R_n(C_n, P; X_1^n)] \\ & = & \mathbf{E}[L(C_n, X_1^n)] - H_n(P) \end{array}$$

where $\mathbf{E}$ denotes the expectation. The **maximal** redundancy is defined as

$$R^*(C_n, P) = \max_{x_1^n}\{R_n(C_n, P; x_1^n)\}(\geq 0).$$

The pointwise redundancy can be negative, maximal and average redundancy cannot (see next slide).

The **redundancy-rate problem** consists in determining for a class $\mathcal{S}$ of source models the rate of growth of the minimax quantities

$$\begin{array}{rcl} \bar{R}_n(\mathcal{S}) & = & \min_{C_n} \sup_{P \in \mathcal{S}} \mathbf{E}[L(C_n, x_1^n) + \lg P(x_1^n)] \\ R_n^*(\mathcal{S}) & = & \min_{C_n} \sup_{P \in \mathcal{S}} \max_{x_1^n}[L(C_n, x_1^n) + \lg P(x_1^n)] \end{array}$$

as $n \to \infty$.

Minimax Regret Functions

We should also point out that there are other measures of optimality for coding, gambling and prediction. We refer here to minimax regret functions defined as follows

$$
\begin{aligned}
\bar{r}_n &= \min_{C_n \in \mathcal{C}} \sup_{P \in \mathcal{S}} \sum_{x_1^n} P(x_1^n)[L_i + \lg \sup_P P(x_1^n)], \\
r_n^* &= \min_{C_n \in \mathcal{C}} \max_{x_1^n} [L_i + \lg \sup_P P(x_1^n)] \quad (= R_n^*)
\end{aligned}
$$

Also, we sometimes the maximin regret is of interest

$$
\underline{r}_n = \sup_{P \in \mathcal{S}} \min_{C_n \in \mathcal{C}} \sum_{x_1^n} P(x_1^n)[L_i + \lg \sup_P P(x_1^n)].
$$

We call $\bar{r}_n$ the *average* minimax regret, r_n^* the *maximal* minimax regret, and $\underline{r}_n$ the *average* maxmin regret.

One can look at the regret function as objective function for the following game theoretical problem: choose L to achieve for every x_1^n a value as good as the best for all players with hindsight, that is, $-\log \sup P(x_1^n)$ (i.e., the minimum code length over the whole set $\mathcal{S}$ of probability distributions).

Analytic Information Theory

The redundancy rate problem is typical of a situation where **second-order asymptotics** play a crucial role since the leading term of $L(C_n)$ is known to be nH, where H is the entropy rate. This problem is an ideal candidate for **analytic information theory** that applies analytic tools to information theory.

As argued by Andrew Odlyzko: "*Analytic methods are extremely powerful and when they apply, they often yield estimates of unparalleled precision.*"

In *1997 Shannon Lecture*, Jacob Ziv presented compelling arguments for "backing off" to a certain degree from the (first-order) asymptotic analysis of information systems in order to predict the behaviour of real systems where we always face *finite* (and often small) lengths (of sequences, files, codes, etc.) One way of overcoming these difficulties is to **increase the accuracy of asymptotic analysis** and replace first-order analyses by more **complete asymptotic expansions**, thereby extending their range of applicability to smaller values while providing more accurate analyses (like constructive error bounds, large deviations, local or central limit laws).

Shtarkov's Bound for the Maximal Minimax

Shtarkov in 1978 proved that the minimax redundancy

$$\lg\left(\sum_{x_1^n}\sup_{P\in\mathcal{S}} P(x_1^n)\right) \le R_n^*(\mathcal{S}) \le \lg\left(\sum_{x_1^n}\sup_{P\in\mathcal{S}} P(x_1^n)\right) + 1.$$

Indeed, for the **lower bound** Shtarkov considered the following probability distribution

$$Q^*(x_1^n) := \frac{\sup_P P(x_1^n)}{\sum_{x_1^n}\sup_P P(x_1^n)}$$

By Kraft's inequality there exists $\tilde{x}_1^n$ such that (for uniquely decodable codes C_n)

$$L(C_n, \tilde{x}_1^n) \ge -\lg Q^*(\tilde{x}_1^n),$$

which implies the lower bound. For the **upper bound**, Shtarkov proposed a code C_n of length

$$L(\widetilde{C}_n) = \lceil \lg 1/Q^*(x_1^n)\rceil = \left\lceil \lg\Big(\sum_{x_1^n}\sup_P P(x_1^n)\Big) - \lg\sup_P P(x_1^n)\right\rceil,$$

which gives the desired upper bound.

Precise Maximal Redundancy

OUR GOAL IS MAKE THE SHTARKOV BOUND A PRECISE RESULT.

We start with a simpler problem. For **known** source P we want to find the optimal code for the maximal redundancy, that is,

$$R_n^*(P) = \min_{C_n \in \mathcal{C}} \max_{x_1^n} [L(C_n, x_1^n) + \log_2 P(x_1^n)].$$

We recall that the Huffman code minimizes

$$\bar{R}_n(P) = \min_{C_n \in \mathcal{C}} \mathbf{E}_{x_1^n} [L(C_n, x_1^n) + \log_2 P(x_1^n)].$$

WHAT IS YOUR GUESS?

Generalized Shannon Code

If P is given, then $\lceil \lg 1/P(x_1^n) \rceil$ defines the length of the Shannon code. A **generalized Shannon code** is such that for some $x_1^n \in \mathcal{L}$ we set the length to be $\lfloor \lg 1/P(x_1^n) \rfloor$ and for some $x_1^n \in \mathcal{U}$ we set the length to be $\lceil \lg 1/P(x_1^n) \rceil$ so that the Kraft inequality hold, that is,

$$\sum_{x_1^n \in \mathcal{L}} P(x_1^n) 2^{\langle -\lg P(x_1^n) \rangle} + \frac{1}{2} \sum_{x_1^n \in \mathcal{U}} P(x_1^n) 2^{\langle -\lg P(x_1^n) \rangle} \le 1,$$

where $\langle x \rangle = x - \lfloor x \rfloor$.

Theorem 1. *If the probability distribution P is dyadic, i.e. $\lg P(x_1^n) \in \mathbf{Z}$ for all $x_1^n \in \mathcal{A}^n$, then $R_n^*(P) = 0$. Otherwise, let $p_1, p_2, \ldots, p_{|\mathcal{A}|^n}$ be the probabilities $P(x_1^n)$, $x_1^n \in \mathcal{A}^n$, ordered in a way that*

$$0 \le \langle -\lg p_1 \rangle \le \langle -\lg p_2 \rangle \le \cdots \le \langle -\lg p_{|\mathcal{A}|^n} \rangle \le 1$$

and let j_0 be the maximal j such that

$$\sum_{i=1}^{j-1} p_i 2^{\langle -\lg p_i \rangle} + \frac{1}{2} \sum_{i=j}^{|\mathcal{A}|^n} p_i 2^{\langle -\lg p_i \rangle} \le 1,$$

where $\langle x \rangle = x - \lfloor x \rfloor$ is the fractional part of x. Then $R_n^*(P) = 1 - \langle -\lg p_{j_0} \rangle$, that is, the Generalized Shannon code with $\mathcal{L} = \{1, \ldots, j_0\}$ is optimal.

Proof of Theorem 1

By picture:

More precisely: assume that P is not dyadic and let C_n^* denote the set of optimal codes, i.e.

$$\mathcal{C}^* = \{C_n \in \mathcal{C} : R_n^*(C_n, P) = R_n^*(P)\}.$$

The idea of the proof is to extract several properties for which an optimal code exists. Especially we will show that there exists an optimal code $C_n^* \in \mathcal{C}^*$ with

(i) $\lfloor -\lg P(x_1^n) \rfloor \leq L(C_n^*, x_1^n) \leq \lceil -\lg P(x_1^n) \rceil$.

(ii) There exists $s_0 \in [0, 1]$ such that

$$L(C_n^*, x_1^n) = \lfloor \lg 1/P(x_1^n) \rfloor \quad \text{if} \quad \langle \lg 1/P(x_1^n) \rangle < s_0 \tag{1}$$

and

$$L(C_n^*, x_1^n) = \lceil \lg 1/P(x_1^n) \rceil \quad \text{if} \quad \langle \lg 1/P(x_1^n) \rangle \geq s_0 \tag{2}$$

Observe that w.l.o.g. we may assume that $s_0 = 1 - R_n^*(P)$. Thus, in order to compute $R_n^*(P)$ we just have to consider codes satisfying (1) and (2). It is clear that the condition in Theorem 1 is just Kraft's inequality for codes of that kind. The optimal choice is $j = j_0$ and consequently $R_n^*(P) = 1 - \langle -\lg p_{j_0} \rangle$.

Precise Maximal Minimax Redundancy

Theorem 2. *Let $\mathcal{S}$ be a system of probability distributions P on $\mathcal{A}^n$ and set*

$$Q^*(x_1^n) := \frac{\sup_{P\in\mathcal{S}} P(x_1^n)}{\sum_{y_1^n\in\mathcal{A}^n}\sup_{P\in\mathcal{S}} P(y_1^n)}.$$

Then

$$R_n^*(\mathcal{S}) = R_n^*(Q^*) + \lg\left(\sum_{x_1^n\in\mathcal{A}^n}\sup_{P\in\mathcal{S}} P(x_1^n)\right),$$

where $R_n^*(Q^*) = 1 - \langle -\lg q_{j_0}\rangle$ as in Theorem 1.

Proof. By definition we have

$$\begin{aligned}
R_n^*(\mathcal{S}) &= \min_{C_n\in\mathcal{C}}\sup_{P\in\mathcal{S}}\max_{x_1^n}(L(C_n,x_1^n)+\lg P(x_1^n)) \\
&= \min_{C_n\in\mathcal{C}}\max_{x_1^n}\left(L(C_n,x_1^n)+\sup_{P\in\mathcal{S}}\lg P(x_1^n)\right) \\
&= \min_{C_n\in\mathcal{C}}\max_{x_1^n}\left(L(C_n,x_1^n)+\lg Q^*(x_1^n)+\lg\left(\sum_{y_1^n\in\mathcal{A}^n}\sup_{P\in\mathcal{S}} P(y_1^n)\right)\right) \\
&= R_n^*(Q^*)+\lg\left(\sum_{y_1^n\in\mathcal{A}^n}\sup_{P\in\mathcal{S}} P(y_1^n)\right).
\end{aligned}$$

Thus, the precise maximal redundancy is established. ∎

Decomposition

Corollary 1. *The maximal minimax redundancy $R_n^*(\mathcal{S})$ can be decomposed into*

$$R_n^*(\mathcal{S}) = I_n(\mathcal{S}) + O_n(C_n^*)$$

where

$$\begin{aligned} I_n(\mathcal{S}) &= \lg\left(\sum_{x_1^n \in \mathcal{A}^n} \sup_{P\in\mathcal{S}} P(x_1^n)\right) := \lg c_n, \\ O_n &= 1 - \langle -\lg q_{j_0}\rangle. \end{aligned}$$

Moreover, $I_{n+1}(\mathcal{S}) \geq I_n(\mathcal{S})$ is a nondecreasing function of n that depends only on the "richness" of $\mathcal{S}$, while $O_n(C_n^*) = O(1)$ is potentially oscillating but bounded part that depends on the optimal code.

Proof. We only need to prove that I_n is nondecreasing. We have

$$\begin{aligned} \sup[P(x_1^n 0)] + \sup[P(x_1^n 1)] &\geq \sup\left[P(x_1^n 0) + P(x_1^n 1)\right] \\ &= \sup P(x_1^{n+1}). \end{aligned}$$

Thus

$$\sum_{x_1^{n+1}} \sup_{P\in\mathcal{S}} P(x_1^{n+1}) \geq \sum_{x_1^n} \sup_{P\in\mathcal{S}} P(x_1^n).$$

The Average Minimax Redundancy

OUR GOAL IS TO PROVIDE A PRECISE RESULT FOR THE AVERAGE MINIMAX REDUNDANCY. Some comments are in order.

We recall that if $\mathcal{S}$ consists of one source P, then

$$\bar{R}_n^H(P) = \min_{C_n \in \mathcal{C}} \sum_{x_1^n} P(x_1^n)[L_i + \lg P(x_1^n)]$$

is the Huffman problem. Only recently we obtained (Szpankowski, 2000) a precise characterization of $R_n^H(P)$ for memoryless sources (see later survey of results).

From known results we often have $\bar{R}_n(\mathcal{S}) \sim R_n^*$. We conjecture the following.

Conjecture. *Under certain additional conditions*

$$\bar{R}_n = R_n^* + \Theta(1) = \lg\left(\sum_{x_1^n \in \mathcal{A}^n} \sup_{P \in \mathcal{S}} P(x_1^n)\right) + \Theta(1)$$

for all n. More generally, $\bar{R}_n(\mathcal{S}) = \lg c_n(\mathcal{S}) + o(\lg c_n)$ where

$$c_n(\mathcal{S}) := \sum_{x_1^n} \sup_{P \in \mathcal{S}} P(x_1^n).$$

Recent-Recent Result

Last week with M. Drmota we proved that

$$\bar{R}_n(\mathcal{S}) \leq \log_2 c_n - \inf_{P \in \mathcal{S}} \sum_{x_1^n} P(x_1^n) \lg \left(\frac{\sup_P P(x_1^n)}{P(x_1^n)} \right) + O(1),$$

$$\bar{R}_n(\mathcal{S}) \geq \log_2 c_n - \sup_{P \in \mathcal{S}} \sum_{x_1^n} P(x_1^n) \lg \left(\frac{\sup_P P(x_1^n)}{P(x_1^n)} \right) + O(1).$$

For example, it is known that for memoryless and Markov processes (in general, d-parameterized distributions)

$$\sum_{x_1^n} P(x_1^n) \lg \left(\frac{\sup_P P(x_1^n)}{P(x_1^n)} \right) = \frac{d}{2} + o(1).$$

The Average Minmax Regret

Theorem 3. *Suppose that $\mathcal{S}$ is a system of probability distributions P on $\mathcal{A}^n$. Then*

$$\bar{r}_n(\mathcal{S}) \leq \lg c_n + O(1).$$

Furthermore, if Q^* is contained in the convex hull of $\mathcal{S}$ then

$$\bar{r}_n(\mathcal{S}) = \lg c_n + O(1).$$

The Average Maxmin Regret

Theorem 4. *Let the distribution $Q^*(x_1^n)$ be defined as before and let $\bar{R}_n^H(P)$ be the average minimax for the Huffman code for the distribution P. Then*

$$\begin{aligned}
\underline{r}_n &= \lg\left(\sum_{x_1^n\in\mathcal{A}^n}\sup_{P\in\mathcal{S}} P(x_1^n)\right) + \sup_P[\bar{R}_n^H(P) - D(P||Q^*)] \\
&= \sup_{P\in\mathcal{S}}\left(\sum_{x_1^n} P(x_1^n)\lg\frac{\sup_{P\in\mathcal{S}} P(x_1^n)}{P(x_1^n)}\right) + O(1)
\end{aligned}$$

where $D(P||Q^*)$ is divergence (Kullback–Leibler distance) between P and Q^*.

Proof. We proceed as follows

$$\begin{aligned}
\underline{r}_n &= \sup_P \min_{C_n\in\mathcal{C}} \sum_{x_1^n} P(x_1^n)[L_i + \lg\sup_P P(x_1^n)] \\
&= \lg c_n + \sup_P \min_{C_n} \sum_{x_1^n} P(x_1^n)[L_i + \lg Q^*(x_1^n)] \\
&= \lg c_n + \sup_P \min_{C_n} \sum_{x_1^n} P(x_1^n)[L_i + \lg Q^*(x_1^n) + \lg P(x_1^n) - \lg P(x_1^n)] \\
&= \lg c_n + \sup_P \min_{C_n} \sum_{x_1^n} P(x_1^n)[L_i + \lg P(x_1^n)] - \sum_{x_1^n} P(x_1^n)\lg\left(\frac{P(x_1^n)}{Q^*(x_1^n)}\right) \\
&= \lg c_n + \sup_P[\bar{R}_n^H(P) - D(P||Q^*)],
\end{aligned}$$

as needed. ∎

Average Redundancy for Shannon Code

Shannon code assigns code of length $\lceil \lg P(x_1^n) \rceil$ to x_1^n (it is assumed that $P(x_1^n)$ is known).

Consider a binary memoryless source with p denoting the probability of generating 0. For a block of length n, the average redundancy $\bar{R}_n^S$ is

$$\begin{aligned} \bar{R}_n^S &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \left(\lceil -\log_2 (p^k (1-p)^{n-k}) \rceil \right. \\ &+ \left. \log_2 (p^k (1-p)^{n-k}) \right). \end{aligned}$$

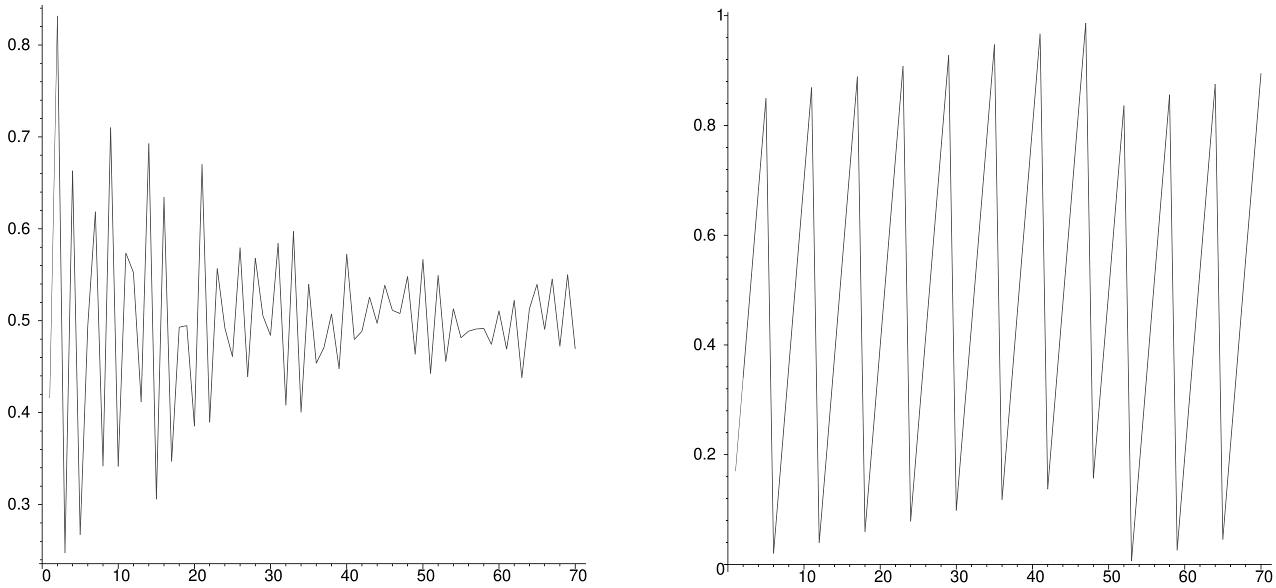
Let

$$\alpha = \log_2 \left(\frac{1-p}{p} \right), \quad \beta = \log_2 \left(\frac{1}{1-p} \right).$$

For the Shannon code we prove (Szpankowski, 2000) that its **average** redundancy is as $n \to \infty$

$$\bar{R}_n^S = \begin{cases} \frac{1}{2} & \alpha \quad \text{irrational} \\ \\ \frac{1}{2} - \frac{1}{M} \left(\langle M n \beta \rangle - \frac{1}{2} \right) & \alpha = \frac{N}{M} \end{cases}$$

where $\langle x \rangle = x - \lfloor x \rfloor$ is the fractional part of x, and N, M are integers such that $\gcd(N, M) = 1$.



(a) (b)

Figure 1: Shannon code redundancy versus block size n for: (a) irrational $\alpha = \log_2(1-p)/p$ with $p = 1/\pi$; (b) rational $\alpha = \log_2(1-p)/p$ with $p = 1/9$.

Maximal Redundancy for Generalized Shannon Code

We now consider the Generalized Shannon code that minimizes the maximal redundancy as observed in Theorem 1. We assume that a sequence is generated by a single memoryless source.

Theorem 5. *Suppose that* $\alpha = \lg \frac{1-p}{p}$ *is irrational. Then, as* $n \to \infty$*,*

$$R_n^*(P_p) = -\frac{\log \log 2}{\log 2} + o(1) = 0.5287\ldots + o(1).$$

If $\lg \frac{1-p}{p} = \frac{N}{M}$ *is rational and non-zero then, as* $n \to \infty$*,*

$$R_n^*(P_p) = -\frac{\lfloor M \lg(M(2^{1/M} - 1)) - \langle Mn \lg 1/(1-p) \rangle \rfloor + \langle Mn \lg 1/(1-p)}{M} + o(1).$$

Finally, if $\lg \frac{1-p}{p} = 0$ *then* $p = \frac{1}{2}$ *and* $R_n^*(P_{1/2}) = 0$.

Proof of Theorem 3

Set $\alpha_p = \lg \frac{1-p}{p}$ and $\beta_p = \lg \frac{1}{1-p}$. Then

$$-\lg(p^k(1-p)^{n-k}) = \alpha_p k + \beta_p n.$$

We only prove the theorem for α_p irrational. Then we know (see INRIA talk in 1999, Szpankowski 2000, Drmota & Tichy, 1997) that for every Riemann integrable function $f : [0,1] \to \mathbf{R}$ we have

$$\lim_{n\to\infty} \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} f(\langle \alpha_p k + \beta_p n \rangle) = \int_0^1 f(x)\,dx.$$

Now set $f_{s_0}(x) = 2^x$ for $0 \le x < s_0$ and $f_{s_0}(x) = 2^{x-1}$ for $s_0 \le x \le 1$. Here we get

$$\lim_{n\to\infty} \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} f_{s_0}(\langle \alpha k + \beta n \rangle) = \frac{2^{s_0-1}}{\log 2} (= 1).$$

Especially for

$$s_0 = 1 + \frac{\log\log 2}{\log 2} = 0.4712\ldots$$

we get $\int_0^1 f(x)\,dx = 1$. This implies that

$$\lim_{n\to\infty} R_n^*(P_p) = 1 - s_0 = 0.5287\ldots$$

Average Redundancy for Generalized Shannon Code

Theorem 6. *Let $\alpha = \lg(1-p)/p$ be irrational. Then the average redundancy of the optimal generalized Shannon code for a memoryless source attains asymptotically*

$$\bar{R}_n^{GS} = \frac{3}{2} - 2\ln 2 + o(1) \approx 0..113705639\ldots \qquad (3)$$

as $n \to \infty$.

Sketch of Proof. Let

$$P_{\mathcal{L}} := \sum_{x_1^n \in \mathcal{L}} P(x_1^n),$$

where $P(x_1^n) = p^k(1-p)^{n-k}$. As before, we can prove that for α irrational

$$\lim_{n\to\infty} \sum_{k\in\mathcal{L}} \binom{n}{k} p^k (1-p)^{n-k} f(\langle \alpha k + \beta n \rangle) = P_{\mathcal{L}} \int_0^1 f(x)\, dx.$$

(Indeed, it suffices in our previous formula to replace f by $f(x)I_{\mathcal{L}}(x)$, where $I_A(b)$ is the characteristic function. Using this we estimate $P_{\mathcal{L}} = 2\ln 2 - 1 + o(1) \approx 0..386294361\ldots$, and observe that $\bar{R}_n^{GS} = 1 - P_{\mathcal{L}} - \bar{R}_n^S$. ∎

Average Redundancy of Huffman Code

As before we consider a binary memoryless source emitting 0 and 1 with probabilities p and $q = 1 - p$, respectively. For a block of length n, we construct its **Huffman code** (through the associated Huffman tree).

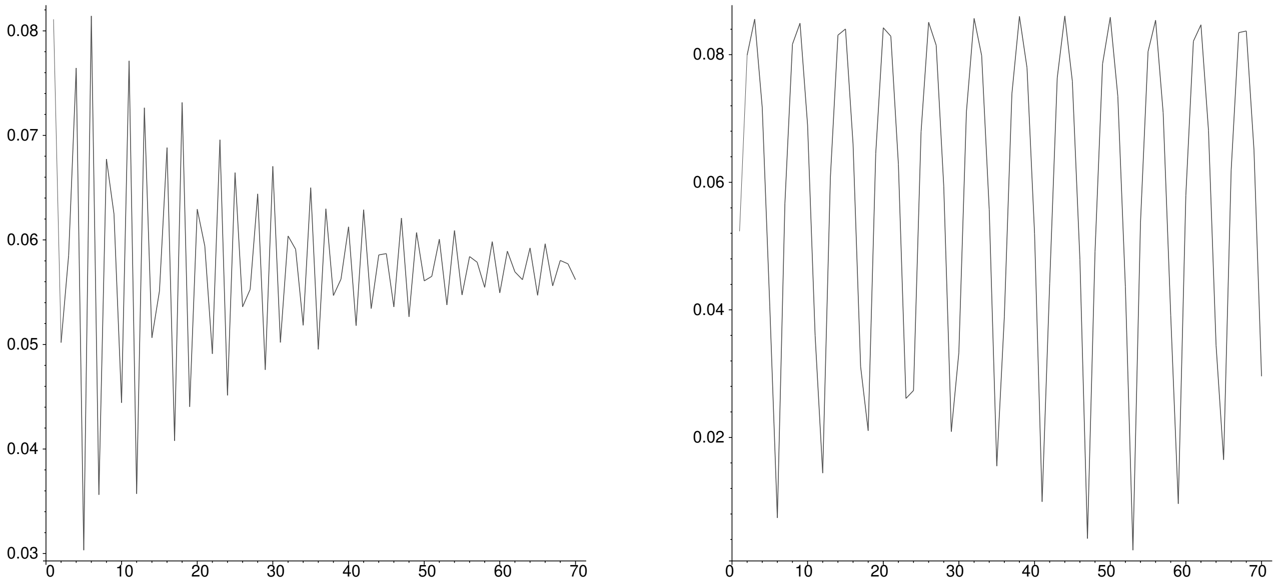
Using Stubley's result (1994), we conclude that the average redundancy $\bar{R}_n^H$ of the Huffman code is

$$\bar{R}_n^H = 1 + \bar{R}_n^S - 2\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} 2^{-\langle \alpha k + \beta n \rangle} + O(\rho^n)$$

where $\rho < 1$ and $\bar{R}_n^S$ is the average redundancy of the Shannon code. As $n \to \infty$ this becomes

$$R_n^H = \begin{cases} \frac{3}{2} - \frac{1}{\ln 2} = 0.057304\ldots & \alpha \text{ irrational} \\[2ex] \frac{3}{2} - \frac{1}{M}\left(\langle \beta M n \rangle - \frac{1}{2}\right) - \frac{1}{M(1-2^{-1/M})} 2^{-\langle n\beta M \rangle / M} & \alpha = \frac{N}{M} \end{cases}$$

with the notation as above.



(a) (b)

Figure 2: Huffman's code redundancy versus block size n for: (a) irrational $\alpha = \log_2(1-p)/p$ with $p = 1/\pi$; (b) rational $\alpha = \log_2(1-p)/p$ with $p = 1/9$.

The maximum Huffman redundancy is

$$\max\{\bar{R}_n^H\} = 1 - \frac{1 + \log\log 2}{\log 2} = \lg(2(\lg e)/e) = 0.08607\ldots,$$

as $n \to \infty$.

Maximal Minimax for a Class of Memoryless Sources

The next step is to consider a class of memoryless sources P_p such that p is contained in an interval $[a, b]$, i.e. we restrict on a quite special (but natural) case for $\mathcal{S}$. Here the result reads as follows.

Theorem 7. *Let* $0 \le a < b \le 1$ *be given and let* $\mathcal{S}_{a,b} = \{P_p : a \le p \le b\}$*. Then, as* $n \to \infty$*.*

$$R_n^*(\mathcal{S}_{a,b}) = \frac{1}{2}\lg n + \lg\, C_{a,b} - \frac{\log\log 2}{\log 2} + o(1),$$

where

$$C_{a,b} = \frac{1}{\sqrt{2\pi}} \int_a^b \frac{dx}{\sqrt{x(1-x)}} = \sqrt{\frac{2}{\pi}}(\arcsin\sqrt{b} - \arcsin\sqrt{a}).$$

Proof. See my talk INRIA 1999.

Finitely Parameterizable Class of Processes

If $\mathcal{M}$ is i.i.d. or the class of Markov chains, or more generally the process belongs to a finitely parameterizable class of dimension K, then Rissanen proved that the average redundancy $\bar{R}_n$ and the minimax redundancy $\bar{R}_n^*$

$$\bar{R}_n(\mathcal{M}) \sim R_n^*(\mathcal{M}) \sim \frac{K}{2} \lg n.$$

as $n \to \infty$. It was also found that the next term of $\bar{R}_n(\mathcal{S})$ and of $R_n^*(\mathcal{S})$ is $O(1)$.

We will prove a full asymptotic expansion of $R_n^*(\mathcal{M})$ for memoryless sources over an m-ary alphabet; e.g.,

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

where $\Gamma(x)$ is the Euler gamma function.

Tunstall's Code

Savari and Gallager 1997 and Savari 1998 analyzed Tunstall's variable-to-fixed codes for memoryless and Markovian sources. For memoryless binary source, it was proved that

$$\bar{R}_n(\mathcal{T}) = -\frac{H \lg H + 0.5 h_2}{\lg n}$$

provided $B = \frac{\log p}{\log q}$ is irrational, where $h_2 = p \log^2 p + q \log^2 q$. The case of B rational was not discussed, but one expects some fluctuation in this case.

Renewal Process

Csiszàr and Shields have studied order r Markov renewal sequences in which a 1 is inserted every $T_0, T_1, \ldots$ of 0's, where $\{T_i\}$ is either an i.i.d. or Markov renewal or r-order Markov renewal process. We denote such sources as $\mathcal{R}_r$.

Csiszàr and Shields proved that

$$\bar{R}_n(\mathcal{R}_r) = R^*(\mathcal{R}_r) = \Theta(n^{(r+1)/(r+2)})$$

for $r = 1, 2, \ldots$ which specializes to $\Theta(\sqrt{n})$ when $r = 0$.

We will prove here (Flajolet & Szpankowski 1998) a precise asymptotic expansion of $R_n^*(\mathcal{R}_0)$ for the renewal processes, namely

$$R_n^*(\mathcal{R}_0) = \frac{2}{\log 2}\sqrt{cn} - \frac{5}{8}\lg n + \frac{1}{2}\lg\log n + O(1)$$

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

Lempel-Ziv Code

Louchard and Szpankowski 1997, Savari 1997, Wyner 1998, and Jacquet and Szpankowski 1995 proved that the Lempel-Ziv codes in the class of i.i.d. and Markov processes have either rate

- $\Theta(n/\log n)$ for LZ'78
- $\Theta(n \log\log n/\log n)$ for LZ'77 code.

More precisely, for LZ'78 Louchard and Szpankowski 1997 showed that (binary alphabet with 0's generated with probability p)

$$\begin{aligned}\bar{R}_n(\mathcal{LZ}) &= H\left(2-\gamma-\frac{1}{2H}h_2+\omega-\delta(n)\right)\frac{n}{\log n} \\ &+ O\left(\frac{n\log\log n}{\log^2 n}\right)\end{aligned}$$

where $H = -p\log p - q\log q > 0$ is the entropy rate, $\gamma = 0.577\ldots$ is the Euler constant, $h_2 = p\log^2 p + q\log^2 q$, and

$$\omega = -\sum_{k=1}^{\infty}\frac{p^{k+1}\log p + q^{k+1}\log q}{1-p^{k+1}-q^{k+1}} \,.$$

The function $\delta(x)$ that fluctuates with mean zero and a tiny amplitude for $\log p/\log q$ rational, but satisfies $\lim_{x\to\infty}\delta(x) = 0$ otherwise.

Shields' Result

Shields proved that there is no function $\rho(n) = o(n)$ which is a weak redundancy rate bound for the class of all ergodic processes.

ANALYTIC METHODS: Fourier Analysis and Sequence Distribution Modulo 1

We consider here redundancy of the **Shannon** block code and the **Huffman** block code for a memoryless source generating a block of length n with the binomial distribution.

Let $p(k) = p^k(1-p)^{n-k}$, where p is the probability of generating 0. Redundancy of the **Shannon** code is

$$\begin{aligned}\bar{R}_n^S &= \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \left(\lceil -\lg p(k) \rceil + \lg p(k)\right) \\ &= 1 - \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \langle \alpha k + \beta n \rangle\end{aligned}$$

$\langle x \rangle = x - \lfloor x \rfloor$ is the fractional part of x and

$$\begin{aligned}\alpha &= \lg\left(\frac{1-p}{p}\right), \\ \beta &= \lg\left(\frac{1}{1-p}\right).\end{aligned}$$

Shannon Code – Irrational Case

Throughout, we shall use the following Fourier series; for real x

$$\begin{aligned}\langle x\rangle &= \frac{1}{2} - \sum_{m=1}^{\infty} \frac{\sin 2\pi m x}{m\pi} \\ &= \frac{1}{2} - \sum_{m\in \mathbf{Z}-\{0\}} c_m e^{2\pi i m x}, \qquad c_m = -\frac{i}{2\pi m},\end{aligned}$$

where $\mathbf{Z}$ is the set of all integers. Hereafter, we shall write $\sum_{m\neq 0} := \sum_{m\in\mathbf{Z}-\{0\}}$.

Irrational Case:

$$\begin{aligned}\bar{R}_n^S &= \frac{1}{2} + \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \sum_{m\neq 0} c_m e^{2\pi i m(\alpha k+\beta n)} \\ &= \frac{1}{2} + \sum_{m\neq 0} c_m e^{2\pi i m\beta n} \left(p e^{2\pi i m\alpha} + q\right)^n .\end{aligned}$$

How to prove that the last term is $o(1)$ for α irrational?

Bernoulli Uniformly Distributed Sequences Modulo 1

Definition 1. [B-u.d. mod 1] *A sequence $x_n \in \mathbf{R}$ is said to be Bernoulli uniformly distributed modulo 1 (in short: B-u.d. mod 1) if for $0 < p < 1$*

$$\lim_{n\to\infty} \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \chi_I(\langle x_k \rangle) = \lambda(I)$$

holds for every interval $I \subset \mathbf{R}$, where $\chi_I(x_n)$ is the characteristic function of I (i.e., it equals to 1 if $x_n \in I$ and 0 otherwise) and $\lambda(I)$ is the Lebesgue measure of I.

Theorem 8. *Let $0 < p < 1$ be a fixed real number and suppose that the sequence x_n is B-uniformly distributed modulo 1. Then for every Riemann integrable function $f : [0,1] \to \mathbf{R}$ we have*

$$\lim_{n\to\infty} \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} f(\langle x_k + y \rangle) = \int_0^1 f(t)\, dt,$$

where the convergence is uniform for all shifts $y \in \mathbf{R}$.

Proof. Standard; cf. Drmota and Tichy (1997) or Kuipers and Niederreiter (1974).

Weyl's Criterion

Theorem 9. [Weyl's Criterion] *A sequence x_n is B-u.d. mod 1 if and only if*

$$\lim_{n\to\infty} \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} e^{2\pi i m x_k} = 0$$

holds for all non-zero $m \in \mathbf{Z} - \{0\}$.

Proof. The proof again is standard. Basically, it is based on the fact that by Weierstrass's *approximation theorem* every Riemann integrable function f of period 1 can be uniformly approximated by a trigonometric polynomial (i.e., a finite combination of functions of the type $e^{2\pi i m x}$).

Finishing the Irrational Case

In our case, we must show that $\langle \alpha k \rangle$ is B-u.d. mod 1. By Weyl's criterion

$$\begin{aligned}\lim_{n\to\infty}\sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}e^{2\pi i m(k\alpha)} &= \lim_{n\to 0}\left(pe^{2\pi i m\alpha}+q\right)^n \\ &= 0\end{aligned}$$

provided α is irrational. Hence, by the previous theorem, with $f(t) = t$ and $y = \beta n$, we immediately obtain

$$\lim_{n\to\infty}\sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}\langle \alpha k + \beta n\rangle = \int_0^1 t dt = \frac{1}{2}.$$

This proves that for α irrational

$$R_n^S = \frac{1}{2} + o(1).$$

It can be proved (thanks to M. Drmota) that for almost all irrational α the rate of convergence in the above is $O\left(\frac{\log^{1+\delta} n}{\sqrt{n}}\right)$ for some $\delta > 0$.

Shannon Redundancy – Rational Case

Now we assume that $\alpha = N/M$ where N, M are integers such that $\gcd(N, M) = 1$. Denote $p_{n,k} = \binom{n}{k} p^k q^{n-k}$. We proceed as follows:

$$
\begin{aligned}
S_n &= \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \left\langle k\frac{N}{M} + \beta n \right\rangle \\
&= \sum_{\ell=0}^{M-1} \sum_{m:\ k=\ell+mM \le n} p_{n,k} \left\langle \ell\frac{N}{M} + N + \beta n \right\rangle \\
&= \sum_{\ell=0}^{M-1} \left\langle \frac{\ell}{M} + \beta n \right\rangle \sum_{m:\ k=\ell+mM \le n} p_{n,k}.
\end{aligned}
$$

We will prove that the last sum is well approximated by $1/M$.

Useful Lemma

Lemma 4. *For fixed $\ell \leq M$ and M, there exist $\rho < 1$ such that*

$$\sum_{m:\ k=\ell+mM\leq n} \binom{n}{k} p^k (1-p)^{n-k} = \frac{1}{M} + O(\rho^n).$$

Proof. Let $\omega_k = e^{2\pi ik/M}$ for $k = 0, 1, \ldots, M-1$ be the Mth root of unity. It is well known that

$$\frac{1}{M}\sum_{k=0}^{M-1} \omega_k^n = \begin{cases} 1 & \text{if } M|n \\ 0 & \text{otherwise.} \end{cases}$$

where $M|n$ means that M divides n. In view of this, we can write

$$\begin{aligned} \sum_{m:\ k=\ell+mM\leq n} \binom{n}{k} p^k q^{n-k} &= \frac{1 + (p\omega_1 + q)^{n-\ell} + .. + (p\omega_{M-1} + q)^{n-\ell}}{M} \\ &= \frac{1}{M} + O(\rho^n), \end{aligned}$$

since $|(p\omega_r + q)| = p^2 + q^2 + 2pq\cos(2\pi r/M) < 1$ for $r \neq 0$.

Finishing the Rational Case

Continuing the derivation and using the above lemma we obtain

$$
\begin{aligned}
S_n &= \frac{1}{M}\sum_{\ell=0}^{M-1}\left(\frac{1}{2} - \sum_{m\neq 0} c_m e^{2\pi i m(\ell/M+\beta n)}\right) \\
&= \frac{1}{2} - \sum_{m\neq 0} c_m e^{2\pi i m n\beta}\frac{1}{M}\sum_{\ell=0}^{M-1} e^{2\pi i m\frac{\ell}{M}} \\
&= \frac{1}{2} - \frac{1}{M}\sum_{m=kM\neq 0} c_{kM} e^{2\pi i k M\beta n} \\
&= \frac{1}{2} - \frac{1}{M}\left(\frac{1}{2} - \langle \beta n M\rangle\right).
\end{aligned}
$$

Huffman Redundancy

We only need to analyze

$$T_n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} 2^{-\langle \alpha k + \beta n \rangle}.$$

The **irrational** case is easy since a direct application of our previous result, with $f(t) = 2^{-t}$ a nd $y = \beta n$, yields

$$\lim_{n\to\infty} T_n = \int_0^1 2^{-t} dt = \frac{1}{2 \log 2}$$

for α irrational.

Huffman Redundancy – Rational Case

We could use Fourier series again, but instead we generalize our previous approach in the following theorem (proposed by M. Drmota):

Theorem 10. *Let $0 < p < 1$ be a fixed real number and suppose that $\alpha = \frac{N}{M}$ is a rational number with $\gcd(N, M) = 1$. Then, for every bounded function $f : [0,1] \to \mathbf{R}$ we have*

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} f(\langle k\alpha + y \rangle) \quad = \quad \frac{1}{M} \sum_{l=0}^{M-1} f\left(\frac{l}{M} + \frac{\langle My \rangle}{M}\right) + O(\rho^n)$$

uniformly for all $y \in \mathbf{R}$ and some $\rho < 1$.

Setting $f(t) = 2^{-t}$ we obtain

$$\begin{aligned} T_n &= \frac{1}{M} \sum_{l=0}^{M-1} 2^{-(l/M)-(\langle M\beta n \rangle/M)} + O(\rho^n) \\ &= \frac{1}{M} 2^{-\langle M\beta n \rangle/M} \frac{1-2^{-M/M}}{1-2^{-1/M}} + O(\rho^n) \\ &= \frac{1}{2M(1-2^{-1/M})} 2^{-\langle M\beta n \rangle/M} + O(\rho^n) \end{aligned}$$

for $\rho < 1$.

ANALYTIC METHODS: Tree Generating Function and Singularity Analysis

We consider here the minimax redundancy for a memoryless source over an m-ary alphabet $\mathcal{A}(m)$. Shtarkov's result implies that

$$R_n^* = \log D_n(m)$$

where $D_n(m)$ satisfies

$$D_n(m) = \sum_{i=1}^{m} \binom{m}{i} D_n^*(i)$$

where $D_0^*(1) = 0$, $D_n^*(1) = 1$ for $n \geq 1$, and for $i > 1$ we have

$$D_n^*(i) = \sum_{k=1}^{n} \binom{n}{k} \left(\frac{k}{n}\right)^k \left(1 - \frac{k}{n}\right)^{n-k} D_{n-k}^*(i-1) \ .$$

Derivation of the Recurrence on $D_n^*(m)$

Observe first that one can write

$$\begin{aligned} D_n(m) &= \sum_{i=1}^{m} \binom{m}{i} \sum_{x^n \in \mathcal{A}(i)} P^*(x^n; \omega_i) \\ &= \sum_{i=1}^{m} \binom{m}{i} D_n^*(i) \end{aligned}$$

where $\mathcal{A}(i)$ represents a subset of $\mathcal{A}$ consisting of i symbols. Indeed, we count separately sequences consisting of only symbols form $\mathcal{A}(i)$.

To derive the recurrence of $D_n^*(i)$ we argue as follows: Consider an alphabet $\mathcal{A}(i-1)$ and assume that these $i-1$ symbols of $\mathcal{A}(i-1)$ occur on $n-k$ positions of x^n. Thus, we deal with $D_{n-k}^*(i-1)$. On the remaining k positions we place the ith symbol with the (optimal) probability

$$\sup_{\omega} P^*(x_1^n, \omega) = \left(\frac{k}{n}\right)^k \left(1 - \frac{k}{n}\right)^{n-k}$$

and establishes the recurrence:

$$D_n^*(i) = \sum_{k=1}^{n} \binom{n}{k} \left(\frac{k}{n}\right)^k \left(1 - \frac{k}{n}\right)^{n-k} D_{n-k}^*(i-1) \ .$$

Main Result

Theorem 11. *For fixed $m \geq 1$ the quantity $D_n^*(m)$ attains the following asymptotics*

$$\begin{aligned} D_n^*(m) &= \frac{\sqrt{\pi}}{\Gamma(\frac{m}{2})}\left(\frac{n}{2}\right)^{\frac{m}{2}-\frac{1}{2}} - \frac{\sqrt{\pi}}{\Gamma(\frac{m}{2}-\frac{1}{2})}\left(\frac{2m}{3}\right)\left(\frac{n}{2}\right)^{\frac{m}{2}-1} \\ &+ \frac{\sqrt{\pi}}{\Gamma(\frac{m}{2})}\left(\frac{n}{2}\right)^{\frac{m}{2}-\frac{3}{2}}\left(\frac{3+m(m-2)(8m-5)}{72}\right) \\ &+ O(n^{\frac{m}{2}-2}) \end{aligned}$$

for large n.

Corollary 2. *For $m \geq 2$*

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

for large n.

Sketch of Proof of Theorem 1

1. Let us introduce a new sequence $\widehat{D}^*_n(m)$ defined as

$$\begin{aligned}\widehat{D}^*_n(m) &= \sum_{k=0}^{n} \binom{n}{k} \left(\frac{k}{n}\right)^k \left(1-\frac{k}{n}\right)^{n-k} D^*_{n-k}(m-1) \\ &= D^*_n(m) + D^*_n(m-1).\end{aligned}$$

2. Observe that

$$\frac{n^n}{n!}\widehat{D}^*_n(m) = \sum_{k=0}^{n} \frac{k^k}{k!} \cdot \frac{(n-k)^{n-k}}{(n-k)!} \widehat{D}^*_{n-k}(m-1) \ .$$

This is a convolution of two sequences $\{k^k/k!\}$ and $\{k^k/k! D^*_k(m-1)$.

3. Define
$$\widehat{D}^*_m(z) = \sum_{k=0}^{\infty} \frac{k^k}{k!} z^k \widehat{D}^*_k(m).$$
The **tree function** $T(z)$ is defined as a solution to

$$T(z) = ze^{T(z)}$$

which is also
$$T(z) = \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} z^k \ .$$
The function $T(z)$ is called the "tree function" since it enumerates rooted labeled trees. It is also related to Lambert's W-function defined as a solution of $W(x)\exp(W(x)) = x$ and which can be called from MAPLE. (In fact, $T(z) = -W(-z)$.) Furthermore, it can be obtained from the Ramanujan's Q-function which finds many applications in hashing, random mappings, and memory conflict.
Related to the tree function is

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

which we need in the analysis.

4. From the recurrence and our definitions we obtain

$$\widehat{D}^*_m(z) = B(z) D^*_{m-1}(z) \ .$$

Thus

$$D^*_m(z) = (B(z)-1)^{m-1} D^*_1(z) = (B(z)-1)^m$$

since $D^*_1(z) = B(z) - 1$. So

$$D^*_n(m) = \frac{n!}{n^n} [z^n] \left((B(z)-1)^m \right)$$

where $[z^n]f(z)$ is the standard notation for the coefficient of $f(z)$ at z^n.

5. Properties of $T(z)$. The tree function has an algebraic singularity at $z = e^{-1}$. This can be seen if we view the functional equation of $T(z)$ is a definition of an implicit function of

$$z(T) = Te^{-T}.$$

This function achieves its maximum value $z = e^{-1}$ at $T = 1$, and by the **implicit-function theorem** it can not be inverted. Thus, it has an algebraic singularity at this point.

6a. Asymptotics. We know that

$$\begin{aligned} T(z) - 1 &= -\sqrt{2(1-ez)} + \frac{2}{3}(1-ez) - \frac{11\sqrt{2}}{36}(1-ez)^{3/2} \\ &+ \frac{43}{135}(1-ez)^2 + O((1-ez)^{5/2}) \ . \end{aligned}$$

Then, $B(z)$ can also be expanded around $z = e^{-1}$ leading to

$$B(z) = \frac{1}{\sqrt{2(1-ez)}} + \frac{1}{3} - \frac{\sqrt{2}}{24}\sqrt{(1-ez)} + \frac{4}{135}(1-ez) + \cdots$$

Singularity analysis of Flajolet and Odlyzko, allows to compute separately the coefficients for every function involved in the above asymptotic expansion. For example,

$$\begin{aligned} [z^n]\left(\frac{1}{\sqrt{1-ez}}\right) &= \frac{e^n}{\sqrt{\pi n}}\left(1 - \frac{1}{8n} + O(1/n^2)\right) \ , \\ [z^n]\left(\sqrt{1-ez}\right) &= -\frac{e^n}{\sqrt{\pi n^3}}\left(\frac{1}{2} + \frac{3}{16n}\right) \\ [z^n]\left(\frac{1}{1-ez}\right) &= e^n \ , \\ \frac{n!}{n^n} &= e^{-n}\sqrt{2\pi n}\left(1 + \frac{1}{12n} + O(1/n^2)\right) \ . \end{aligned}$$

6b. Thus

$$
\begin{aligned}
e^{-n}[z^n](B(z)-1)^m &= \frac{n^{\frac{m}{2}-1}}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})} - \frac{n^{\frac{m}{2}-\frac{3}{2}}}{2^{\frac{m}{2}-\frac{1}{2}}}\left(\frac{2m}{3\Gamma(\frac{m}{2}-\frac{1}{2})}\right) \\
&+ \frac{n^{\frac{m}{2}-2}}{2^{\frac{m}{2}}}\left(\frac{m(m-2)(8m-5)}{36\Gamma(\frac{m}{2})}\right) \\
&+ O(n^{\frac{m}{2}-\frac{5}{2}})
\end{aligned}
$$

as desired.

Sketch of Proof of Corollary 1

From our definition we have

$$D_m(z) = \sum_{i=1}^{m} \binom{m}{i} D_i^*(z) = \sum_{i=1}^{m} \binom{m}{i} (B(z)-1)^i \ .$$

Thus

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

Then, $D_n(m) = \frac{n!}{n^n}[z^n](B^m(z)-1)$. We found:

$$\begin{aligned} e^{-n}[z^n]B^m(z) \quad &= \quad \frac{n^{\frac{m}{2}-1}}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})} + \frac{n^{\frac{m}{2}-\frac{3}{2}}}{2^{\frac{m}{2}-\frac{1}{2}}}\left(\frac{m}{3\Gamma(\frac{m}{2}-\frac{1}{2})}\right) \\ &+ \quad \frac{n^{\frac{m}{2}-2}}{2^{\frac{m}{2}}}\left(\frac{m(m-2)(2m+1)}{36\Gamma(\frac{m}{2})}\right) + O(n^{\frac{m}{2}-\frac{5}{2}}) \ . \end{aligned}$$

Finally, we additionally observe that

$$\log(1 + a\sqrt{x} + bx + cx^{3/2}) = a\sqrt{x} + (b - \frac{1}{2}a^2)x + O(x^{3/2})$$

as $x \to 0$, and this completes the proof.

General Recurrence

Our method allows to solve a general recurrence of the following form:

$$x_n^m = a_n + \sum_{i=0}^{n} \binom{n}{i} \left(\frac{i}{n}\right)^i \left(1 - \frac{i}{n}\right)^{n-i} (x_i^{m-1} + x_{n-i}^{m-1}) \ ,$$

where a_n is a given sequence (the so called *additive* term), and m is an additional parameter.

Indeed, the above leads to

$$X_m(z) = A(z) + 2B(z)X_{m-1}(z)$$

where

$$X_m(z) = \sum_{k=0}^{\infty} \frac{k^k}{k!} z^k x_k^m \ , \qquad A(z) = \sum_{k=0}^{\infty} \frac{k^k}{k!} z^k a_k \ .$$

This last recurrence can be solved by telescoping in terms of m, and then the singularity analysis will provide an asymptotic expansion, as discussed above.

ANALYTIC METHODS: Mellin Transform and Saddle Point Approach

Csiszár and Shields studied the minimax redundancy of the **renewal process** defined as:
Let $T_1, T_2 \ldots$ be a sequence of i.i.d. positive-valued random variables with distribution $Q(j) = \Pr\{T_i = j\}$ over nonnegative integers $j \geq 0$. The process $T_0, T_0 + T_1, T_0 + T_1 + T_2, \ldots$ is called the renewal process which is stationary if T_0 is chosen properly. With such a renewal process we associate a **binary renewal sequence** in which the positions of the 1's are at the renewal epochs $T_0, T_0 + T_1, T_0 + T_1 + T_2, \ldots$.

By Shtarkov's method, to study the redundancy R_n^* we should evaluate

$$R_n = \lg \left(\sum_{x_1^n} \sup_Q P(x_1^n) \right)$$

A Simple Lemma

Lemma 5. *Define* $r_n = 2^{R_n}$. *Then*

$$
\begin{aligned}
r_n &= \sum_{k=0}^{n} r_{n,k} \\
r_{n,k} &= \sum_{\mathcal{P}(n,k)} \binom{k}{k_0 \cdots k_{n-1}} \left(\frac{k_0}{k}\right)^{k_0} \left(\frac{k_1}{k}\right)^{k_1} \cdots \left(\frac{k_{n-1}}{k}\right)^{k_{n-1}}
\end{aligned}
$$

where $\mathcal{P}(n,k)$ *denotes the partition of* n *into* k *terms, that is*

$$
\begin{aligned}
n &= k_0 + 2k_1 + \cdots + nk_{n-1}, \\
k &= k_0 + \cdots + k_{n-1}.
\end{aligned}
$$

Proof. Observe that the renewal sequence x_1^n can be represented as

$$x_1^n = 0^{\alpha_1} 1 0^{\alpha_2} 1 \cdots 1 0^{\alpha_n} 1 \underbrace{0 \cdots 0}_{k^*}$$

where $0 \leq \alpha_i \leq n$ for $i = 1, \ldots, n$. Let k_m be the number of i such that $\alpha_i = m$, where $m = 0, 1, \ldots, n-1$.

Then

$$P(x_1^n) = Q^{k_0}(0)Q^{k_1}(1)\cdots Q^{k_{n-1}}(n-1)Q^*(k^*)$$

subject to $Q(0) + Q(1) + \cdots + Q(n-1) \le 1$, where

$$k_0 + 2k_1 + \cdots + nk_{n-1} + k^* = n,$$

and $Q^*(k^*) = \Pr\{T_1 \ge k^*\}$. This is a simple optimization problem with constraints that can be easily solved leading to

$$\sup_Q P(x_1^n) = \left(\frac{k_0}{k_0 + \cdots + k_{n-1}}\right)^{k_0} \cdots \left(\frac{k_{n-1}}{k_0 + \cdots + k_{n-1}}\right)^{k_{n-1}}$$

which proves the lemma.

Re-Formulation of the Problem

A difficulty of finding asymptotics of r_n stems from the factor $k!/k^k$ present in the definition of $r_{n,k}$. We circumvent this problem by analyzing a related pair of sequences, namely s_n and $s_{n,k}$ that are defined as

$$\begin{cases} s_n & = & \Sigma_{k=0}^{n} s_{n,k} \\ s_{n,k} & = & e^{-k} \Sigma_{\mathcal{P}(n,k)} \frac{k^{k_0}}{k_0!} \cdots \frac{k^{k_{n-1}}}{k_{n-1}!}. \end{cases}$$

The translation from s_n to r_n is most conveniently expressed in probabilistic terms. Introduce the random variable K_n whose probability distribution is $s_{n,k}/s_n$, that is,

$$\varpi_n \ : \qquad \Pr\{K_n = k\} = \frac{s_{n,k}}{s_n},$$

where ϖ_n denotes the distribution. Then Stirling's formula yields

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

Thus, the problem of finding r_n reduces to asymptotic evaluations of s_n, $\mathbf{E}[\sqrt{K_n}]$ and $\mathbf{E}[K_n^{-\frac{1}{2}}]$.

Fundamental Lemmas

The heart of the matter is the following lemma which provides the necessary estimates.

Lemma 6. *Let $\mu_n = \mathbf{E}[K_n]$ and $\sigma_n^2 = \mathbf{Var}(K_n)$, where K_n has the distribution ϖ_n defined above. The following holds*

$$\begin{aligned} s_n &\sim \exp\left(2\sqrt{cn} - \frac{7}{8}\log n + d + o(1)\right) \\ \mu_n &= \frac{1}{4}\sqrt{\frac{n}{c}}\log\frac{n}{c} + o(\sqrt{n}) \\ \sigma_n^2 &= O(n\log n) = o(\mu_n^2), \end{aligned}$$

where $c = \pi^2/6 - 1$, $d = -\log 2 - \frac{3}{8}\log c - \frac{3}{4}\log\pi$.

By Chebyshev's we also have:

Lemma 7. *For large n*

$$\begin{aligned} \mathbf{E}[\sqrt{K_n}] &= \mu_n^{1/2}(1 + o(1)) \\ \mathbf{E}[K_n^{-\frac{1}{2}}] &= o(1). \end{aligned}$$

where $\mu_n = \mathbf{E}[K_n]$.

Main Result

In summary, r_n and s_n are related by

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

This leads to

Theorem 12. [Flajolet and Szpankowski 1998] *Consider the class of renewal processes as defined above. The minimax redundancy* ρ_n *attains the following asymptotics*

$$R_n^*(\mathcal{R}_0) = \frac{2}{\log 2}\sqrt{cn} - \frac{5}{8}\lg n + \frac{1}{2}\lg\log n + O(1)$$

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

Proof of the Fundamental Lemma

1. We start by introducing the well-known "tree function" $T(z)$ defined as the solution of

$$T(z) = ze^{T(z)}$$

that is analytic at 0. The function $T(z)$ satisfies, by the Lagrange inversion theorem,

$$T(z) = \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} z^k \ .$$

2. Next define the function $\beta(z)$ as

$$\beta(z) = \sum_{k=0}^{\infty} \frac{k^k}{k!} e^{-k} z^k.$$

One has (e.g., by Lagrange inversion again or otherwise)

$$\beta(z) = \frac{1}{1 - T(ze^{-1})}.$$

3. The quantities s_n and $s_{n,k}$ have generating functions,

$$S_n(u) = \sum_{k=0}^{\infty} s_{n,k} u^k, \qquad S(z,u) = \sum_{n=0}^{\infty} S_n(u) z^n.$$

Then, since $s_{n,k}$ involves convolutions of sequences of the form $k^k/k!$, we have

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

4. To compute the moments μ_n and $\mathbf{E}[K_n(K_n-1)]$ we use the following formulas

$$\begin{aligned} s_n &= [z^n]S(z,1), \\ \mu_n &= \frac{[z^n]S'_u(z,1)}{[z^n]S(z,1)}, \\ \mathbf{E}[K_n(K_n-1)] &= \frac{[z^n]S''_{uu}(z,1)}{[z^n]S(z,1)} \end{aligned}$$

where $[z^n]F(z)$ denotes the coefficient at z^n of $F(z)$, $S'_u(z,1)$ and $S''_{uu}(z,1)$ represent the first and the second derivative of $S(z,u)$ at $u=1$.

Mellin Asymptotics

5. We deal here with

$$S(z,1) = \prod_{i=1}^{\infty} \beta(z^i).$$

The behaviour of the generating function $S(z,1)$ as $z \to 1$ is an essential ingredient of the analysis.

5a. The singularity of the tree function $T(z)$ at $z = e^{-1}$ is of the square-root type, that is, near $z = 1$, $\beta(z)$ admits the singular expansion :

$$\beta(z) = \frac{1}{\sqrt{2(1-z)}} + \frac{1}{3} - \frac{\sqrt{2}}{24}\sqrt{(1-z)} + O(1-z).$$

5b. We now turn to the infinite product asymptotics as $z \to 1^-$, with z real. Let $L(z) = \log S(z,1)$ and $z = e^{-t}$, so that

$$L(e^{-t}) = \sum_{k=1}^{\infty} \log \beta(e^{-kt}).$$

Mellin transform techniques provide an expansion of $L(e^{-t})$ around $t = 0$ (or equivalently $z = 1$) since the sum falls under the *harmonic sum* paradigm.

Mellin Properties

(M1) Direct and Inverse Mellin Transforms. Let c belong to the *fundamental strip* defined below.

$$f^*(s) := \mathcal{M}(f(x); s) = \int_0^\infty f(x) x^{s-1} dx$$

then

$$f(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} f^*(s) x^{-s} ds.$$

(M2) Fundamental Strip. The Mellin transform of $f(x)$ exists in the *fundamental strip* $\Re(s) \in (-\alpha, -\beta)$, where

$$f(x) = O(x^\alpha) \qquad (x \to 0), \qquad\qquad f(x) = O(x^\beta) \qquad (x \to \infty).$$

(M3) Harmonic Sum Property. By linearity and the scale rule $\mathcal{M}(f(ax); s) = a^{-s}\mathcal{M}(f(x); s)$,

$$f(x) = \sum_{k \geq 0} \lambda_k g(\mu_k x)$$

then

$$f^*(s) = g^*(s) \sum_{k \geq 0} \lambda_k \mu_k^{-s}.$$

(M4) Mapping Properties (Asymptotic expansion of

$f(x)$ and singularities of $f^*(s)$).

$$f(x) = \sum_{(\xi,k)\in A} c_{\xi,k} x^{\xi} (\log x)^k + O(x^M)$$

then

$$f^*(s) \asymp \sum_{(\xi,k)\in A} c_{\xi,k} \frac{(-1)^k k!}{(s+\xi)^{k+1}}.$$

(i) *Direct Mapping.* Assume that $f(x)$ admits as $x \to 0^+$ the asymptotic expansion of the above for some $-M < -\alpha$ and $k > 0$. Then for $\Re(s) \in (-M, -\beta)$, the transform $f^*(s)$ satisfies the singular expansion of above.

(ii) *Converse Mapping.* Assume that $f^*(s) = O(|s|^{-r})$ with $r > 1$, as $|s| \to \infty$ and that $f^*(s)$ admits the singular expansion above for $\Re(s) \in (-M, -\alpha)$. Then $f(x)$ satisfies the asymptotic expansion of above at $x = 0^+$.

Continuation of the Proof

6. The Mellin transform $L^*(s) = \mathcal{M}(L(e^{-t}); s)$ of $L(e^{-t})$ is computed by the harmonic sum property (M3). For $\Re(s) \in (1, \infty)$, the transform evaluates to

$$L^*(s) = \zeta(s)\Lambda(s)$$

where $\zeta(s) = \Sigma_{n\geq 1} n^{-s}$ is the Riemann zeta function, and

$$\Lambda(s) = \int_0^\infty \log \beta(e^{-t}) t^{s-1} dt.$$

7. By the direct mapping property (M4), the expansion of $\beta(z)$ at $z = 1$ implies

$$\log \beta(e^{-t}) = -\frac{1}{2} \log t - \frac{1}{2} \log 2 + O(\sqrt{t}),$$

so that, collecting local expansions,

$$\Lambda(s) \asymp \Big(\Lambda(1)\Big)_{s=1} + \left(\frac{1}{2}\frac{1}{s^2} - \frac{1}{2}\frac{\log 2}{s}\right)_{s=0}.$$

Then

$$L^*(s) \asymp \left(\frac{\Lambda(1)}{s-1}\right)_{s=1} + \left(-\frac{1}{4s^2} - \frac{\log \pi}{4s}\right)_{s=0}.$$

8. An application of the converse mapping property (M4) allows us to come back to the original function,

$$L(e^{-t}) = \frac{\Lambda(1)}{t} + \frac{1}{4}\log t - \frac{1}{4}\log \pi + O(\sqrt{t}),$$

which translates in

$$L(z) = \frac{\Lambda(1)}{1-z} + \frac{1}{4}\log(1-z) - \frac{1}{4}\log \pi - \frac{1}{2}\Lambda(1) + O(\sqrt{1-z}).$$

where

$$\begin{aligned} c = \Lambda(1) \quad &= \quad -\int_0^1 \log(1 - T(x/e))\frac{dx}{x} \\ &= \quad \frac{\pi^2}{6} - 1. \end{aligned}$$

9. In summary, we just proved that, as $z \to 1^-$,

$$S(z,1) = e^{L(z)} = a(1-z)^{\frac{1}{4}} \exp\left(\frac{c}{1-z}\right)(1 + o(1)),$$

where $a = \exp(-\frac{1}{4}\log \pi - \frac{1}{2}c)$.

10. It remains to collect the information gathered on $S(z,1)$ and recover $s_n = [z^n]S(z,1)$ asymptotically. The inversion is provided by the Cauchy coefficient formula, that is,

$$s_n = \frac{1}{2\pi i}\oint \frac{S(z,1)}{z^{n+1}}dz$$

where the integration path is any simple loop around 0 inside the unit disk.

11. To estimate s_n we use the following lemma that is based on an application of the saddle point method summarized on the next few slides.

Lemma 8. *For positive $A > 0$, and reals B and C, define $f(z) = f_{A,B,C}(z)$ as*

$$f(z) = \exp\left(\frac{A}{1-z} + B\log\frac{1}{1-z} + C\log\left(\frac{1}{z}\log\frac{1}{1-z}\right)\right).$$

Then, the nth Taylor coefficient of $f_{A,B,C}(z)$ satisfies asymptotically, for large n,

$$\begin{aligned}[z^n]f_{A,B,C}(z) \quad = \quad & 2\sqrt{An} + \frac{1}{2}\left(B - \frac{3}{2}\right)\log n \\ & + C\log\log\sqrt{\frac{n}{A}} \\ & - \frac{1}{2}\log\left(4\pi e^{-A}/\sqrt{A}\right) + o(1).\end{aligned}$$

Saddle Point Method

Input: A function $g(z)$ analytic in $|z| < R$ $(0 < R < +\infty)$ with nonnegative Taylor coefficients and "fast growth" as $z \to R^-$. Let $h(z) := \log g(z) - (n+1)\log z$.

Output: The asymptotic formula for $g_n := [z^n]g(z)$ derived from the Cauchy coefficient integral

$$g_n = \frac{1}{2i\pi}\int_\gamma g(z)\,\frac{dz}{z^{n+1}} = \frac{1}{2i\pi}\int_\gamma e^{h(z)}\,dz$$

where γ is a loop around $z = 0$.

<u>(S1). Saddle point contour.</u> *Require that $g'(z)/g(z) \to +\infty$ as $z \to R^-$.* Let $r = r(n)$ be the unique positive root of the saddle point equation

$$h'(r) = 0 \qquad \text{or} \qquad \frac{rg'(r)}{g(r)} = n+1,$$

so that $r \to R$ as $n \to \infty$. The integral above is evaluated on $\gamma = \{z \mid |z| = r\}$.

(S2). BASIC SPLIT. *Require that* $h'''(r)^{1/3} h''(r)^{-1/2} \to 0$. Define $\varphi = \varphi(n)$ called the "range" of the saddle point by

$$\varphi = \left| h'''(r)^{-1/6} h''(r)^{-1/4} \right|,$$

so that $\varphi \to 0$, $h''(r)\varphi^2 \to \infty$, and $h'''(r)\varphi^3 \to 0$. Split $\gamma = \gamma_0 \cup \gamma_1$, where $\gamma_0 = \{z \in \gamma \mid |\arg(z)| \leq \varphi\}$, $\gamma_1 = \{z \in \gamma \mid |\arg(z)| \geq \varphi\}$.

(S3) ELIMINATION OF TAILS. *Require that* $|g(re^{i\theta})| \leq |g(re^{i\varphi})|$ *on* γ_1. Then, the tail integral satisfies the trivial bound,

$$\left| \int_{\gamma_1} e^{h(z)}\, dz \right| = O\left(\left| e^{-h(re^{i\varphi})} \right| \right).$$

(S4) LOCAL APPROXIMATION. *Require that* $h(re^{i\theta}) - h(r) - \frac{1}{2}r^2\theta^2 h''(r) = O(|h'''(r)\varphi^3|)$ *on* γ_0. Then, the central integral is asymptotic to a complete Gaussian integral, and

$$\frac{1}{2i\pi}\int_{\gamma_0} e^{h(z)}\,dz = \frac{g(r)r^{-n}}{\sqrt{2\pi h''(r)}}\left(1 + O(|h'''(r)\varphi^3|)\right).$$

(S5) COLLECTION. Requirements $(S1), (S2), (S3), (S4)$, imply the estimate:

$$[z^n]g(z) = \frac{g(r)r^{-n}}{\sqrt{2\pi h''(r)}}\left(1 + O(|h'''(r)\varphi^3|)\right) \sim \frac{g(r)r^{-n}}{\sqrt{2\pi h''(r)}}.$$