Analytic Combinatorics, Information Theory, and Algorithmics: Precise Redundancy Rate Problem*

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

May 29, 2004

*Research supported by NSF Grant C-CR 9804760 and contract 1419991431A from sponsors of CERIAS at Purdue.

Outline

1. Basic Notation
2. Basic Facts of Source Coding
3. The Redundancy Rate Problem
4. Redundancy Rates for Known Sources
   (a) Average Redundancy (Huffman's Code)
   (b) Maximal Redundancy (Generalized Shannon's Code)
   (c) Sketch of Proofs (*sequences mod 1, Fourier analysis*)
5. Maximal Minimax Redundancy for a Class of Sources
   (a) Shtarkov's Bound and Its Generalizations
   (b) Memoryless Sources
   (c) Markov Sources
   (d) Renewal Sources
   (e) Sketch of Proofs (*analytic combinatorics, singularity analysis, Mellin transform, saddle point methods*)
6. Average Minimax Redundancy
   (a) Average Versus Maximal Minimax Redundancy
   (b) Sketch of Proof (*convexity theory*)
7. Conclusions and Open Problems

Basic Notation

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 $H_n(P) = -\sum_{x_1^n} P(x_1^n) \lg P(x_1^n)$,

The **basic problem** of **source coding** (part of *information theory* known also as *data compression*) is to find codes with shortest descriptions (lengths) either on *average* or for *individual sequences* when the source (i.e., statistics of the underlying probability distribution) is unknown.

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. An easy exercise on trees. Let $\ell_{\max}$ be the maximum codeword length. Since the number of descendants at level $\ell_{\max}$ of a codeword located at level ℓ_i is $2^{\ell_{\max}-\ell_i}$, we obtain

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

Shannon's Lower Bound

Lemma 2 (Shannon). *For any prefix 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

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

since the first term is a divergence and cannot be negative (or $\log x \leq x - 1$ for $0 < x \leq 1$) 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 cannot 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*

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

and therefore

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

Proof of Barron's Lemma

We argue as follows:

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

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

$$\begin{aligned} R_n(C_n, P; x_1^n) &= L(C_n, x_1^n) + \lg P(x_1^n) \\ \bar{R}_n(C_n) &= \mathbf{E}_{X_1^n}[R_n(C_n, P; X_1^n)] \\ &= \mathbf{E}[L(C_n, X_1^n)] - H_n(P) \geq 0 \end{aligned}$$

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

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

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

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

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* maximin 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).

Redundancy for Known Sources

We start with the simplest problem, that is, we assume that the source is known (i.e., $\mathcal{S} = \{P\}$ and the probability measure is given). Surprisingly enough, there still remains some open problems in this setting.

Mostly, information theory was concerned with finding an optimal code that minimizes the **average redundancy**, that is, a code solving the following problem

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

We recall that the well known Huffman code is the solution to this problem.

But there are other optimization criteria that are of interest. For example, what code minimizes the maximal redundancy? More precisely, we seek a prefix code C_n such that

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

We shall discuss it in the sequel.

Average Redundancy for Shannon Code

In order to give a glimpse into our approach used to derive the redundancy of Huffman code, we first illustrate it on a simpler code. Let us start with the **Shannon code** that assigns the length

$$L(C_n^S, x_1^n) = \lceil -\lg P(x_1^n) \rceil$$

to the source sequence x_1^n generated by a binary memoryless source such that

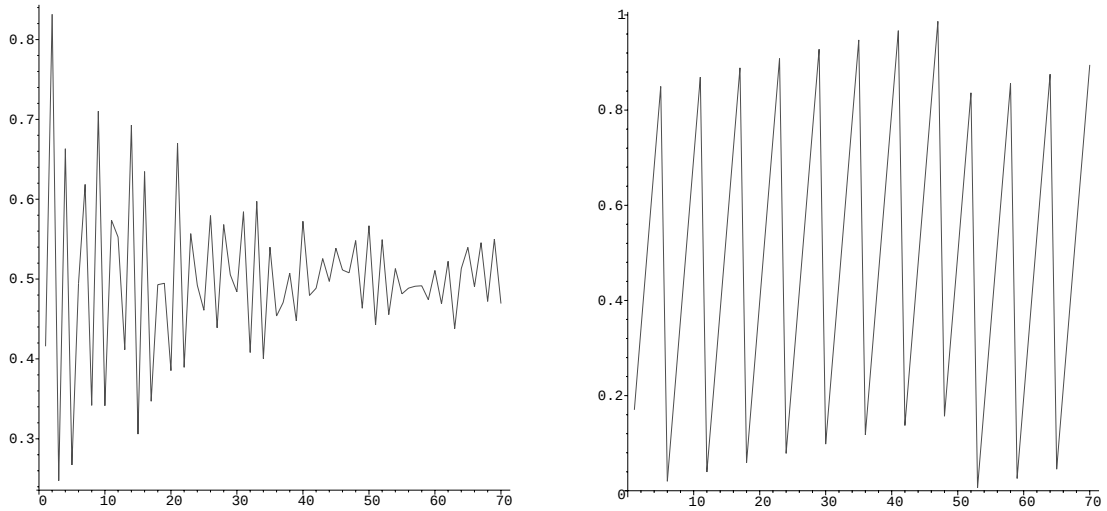
$$P(x_1^n) = p^k (1-p)^{n-k}$$

where p is **known** probability of generating 0 and k is the number of 0s. The Shannon code redundancy is

$$\begin{aligned} \bar{R}_n^S &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \Big(\lceil -\log_2 (p^k (1-p)^{n-k}) \rceil \\ &+ \log_2 (p^k (1-p)^{n-k}) \Big) \\ &= 1 - \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \langle \alpha k + \beta n \rangle \end{aligned}$$

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

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



(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 $p = 1/9$.

Sketch of Proof

The problem of evaluating the average redundancy of Shannon and Huffman codes can be reduced to an asymptotic estimate of the following sum (as $n \to \infty$)

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} f(\langle x_k + y \rangle)$$

for fixed p and some Riemann integrable function $f : [0,1] \to \mathbf{R}$ (uniformly over $y \in \mathbf{R}$).

It turns out that asymptotics of the above sum depends on the behavior of the sequence $\langle x_k + y \rangle \subset [0,1)$. For example, for $x_k = \alpha k$, two cases must be considered:

- α irrational;
- α rational

Uniformly Distributed Sequences Mod 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 1. *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) (cf. also Szpankowski (2000)).

Weyl's Criterion

Theorem 2 (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 $e^{2\pi i m x}$).

Shannon Code: The Irrational Case

Let us return to the Shannon code redundancy. As mentioned before we must consider two cases: α irrational and α rational. We first consider α **irrational**.

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

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

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**

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

Shannon Code: The Rational Case

Let now α be rational. The following simple result is easy to prove.

Lemma 4. *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*

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

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

Now, having the above two results we easily establish that

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

Average Redundancy of the Huffman Code

Now we can return to the Huffman code. It can be proved that the average redundancy 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.

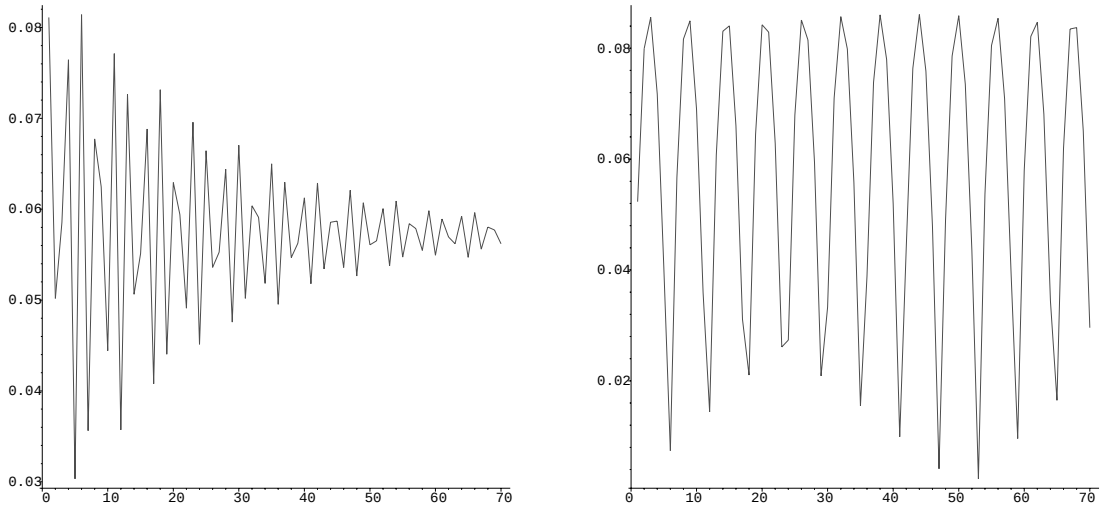
Theorem 3 (Szpankowski, 2000). *Consider the Huffman block code of length n over a binary memoryless source Binomial(n,p) and set*

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

Then for $p < \frac{1}{2}$ as $n \to \infty$

$$\bar{R}_n^H = \begin{cases} \frac{3}{2} - \frac{1}{\ln 2} + o(1) \approx 0.057304 & \alpha \text{ irrational} \\ \frac{3}{2} - \frac{1}{M}\left(\langle \beta M n \rangle - \frac{1}{2}\right) - \frac{1}{M(1-2^{-1/M})} 2^{-\langle n\beta M \rangle / M} + O(\rho^n) & \alpha = \frac{N}{M} \end{cases}$$

where N, M are integers such that $\gcd(N, M) = 1$ and $\rho < 1$.



(a) (b)

Figure 2: The average redundancy of Huffman codes 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 $p = 1/9$.

Generalized Shannon Code

We now turn our attention to the maximal redundancy, that is, we seek a prefix code C_n such that

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

Let us define a *generalized Shannon* code C_n^{GS} as

$$L(C_n^{GS}, x_1^n) = \begin{cases} \lfloor \lg 1/P(x_1^n) \rfloor & \text{if} \quad x_1^n \in \mathcal{L} \\ \lceil \lg 1/P(x_1^n) \rceil & \text{if} \quad x_1^n \in \mathcal{A}^n \setminus \mathcal{L} \end{cases}$$

where $\mathcal{L} \subset \mathcal{A}^n$, and the Kraft inequality holds. it is easy to see that a generalized Shannon code is the optimal code for the above problem.

Main Results

Theorem 4 (Drmota and Szpankowski, 2001). *Let* $p_1, p_2, \ldots, p_{|\mathcal{A}|^n}$ *be the probabilities* $P(x_1^n)$, $x_1^n \in \mathcal{A}^n$ *such that*

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

and let j_0 *be the maximal* j *such that*

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

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 for the maximal redundancy problem.*

Maximal Redundancy of the Generalized Shannon Code

Consider the Generalized Shannon code constructed for a source sequence x_1^n generated by a memoryless binary source such that

$$P(x_1^n) = p^k (1-p)^{n-k}$$

where p is **known** probability of generating 0s and k is the number of 0s.

Theorem 5 (Drmota and Szpankowski, 2001). *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 $n \to \infty$,

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

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

Minimax Redundancy for a Class of Sources

We now assume that a source sequence x_1^n is generated by a source from a **set** of sources $\mathcal{S}$ (i.e., a class of distributions $P \in \mathcal{S}$). Every source may have a different distribution P (within the class) and we must design the **best** code ($\min_{C_n}$, shortest length) for the **worst** source ($\sup_{P\in\mathcal{S}}$), i.e., $\min_{C_n} \sup_{P\in\mathcal{S}}[L(c_n, x_1^n) + \lg P(x_1^n n]$.

We consider the following classes of sources:

- **Memoryless sources** $\mathcal{M}_0$ over an m-ary (finite) alphabet, that is,

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

  with $k_1 + \cdots + k_m = n$, where p_i are **unknown**!
- **Markov sources** $\mathcal{M}_r$ over a finite alphabet of order r.
- **Renewal Sources** where an 1 is introduced after a run of 0s distributed according to some distribution.
- **Mixing Sources** where the probability distribution is mixing (i.e., $(1 - \psi(g))P(A)P(B) \leq P(AB) \leq (1 + \psi(g))P(A)P(B)$ where $A \in \mathcal{F}_{-\infty}^0$ and $B \in \mathcal{F}_g^\infty$).

Maximal Minimax Redundancy

Shtarkov in 1978 proved that the minimax redundancy

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

We can prove a precise result for the maximal minimax redundancy, as shown below.

Lemma 5. *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

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

where $R_n^{GS}(Q^*)=1-\langle-\lg q_{j_0}\rangle$ is the maximal redundancy of a Generalized Shannon code built for the (known) distribution Q^*.

Sketch of Proof

By definition we have

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

Thus, the precise maximal redundancy is established.

Decomposition

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

$$R_n^*(\mathcal{S}) = D_n(\mathcal{S}) + R_n^{GS}(Q_n^*)$$

where

$$\begin{aligned} D_n(\mathcal{S}) &= \lg \left(\sum_{x_1^n \in \mathcal{A}^n} \sup_{P \in \mathcal{S}} P(x_1^n) \right) := \lg d_n(\mathcal{S}), \\ R_n^{GS}(Q_n^*) &= 1 - \langle -\lg q_{j_0} \rangle. \end{aligned}$$

Moreover, $D_{n+1}(\mathcal{S}) \geq D_n(\mathcal{S})$ is a nondecreasing function of n that depends only on the "richness" of $\mathcal{S}$, while $R_n^{GS}(Q_n^*) = O(1)$ is potentially fluctuating but bounded part that depends on the optimal code.

Class of Memoryless Sources

We first consider a class of memoryless sources over a finite m-ary alphabet, and derive precise asymptotics using a combination of **combinatorial and analytic tools** (e.g., generating functions and singularity analysis)

We first deal only with the term $d_n(\mathcal{M}_0)$ of the maximal minimax redundancy $R_n^*(\mathcal{M}_0)$. It easy to see that

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

since

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

where

$$k_1 + \cdots + k_m = n.$$

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

We write

$$d_n(\mathcal{M}_0) = \frac{n!}{n^n} \sum_{k_1+\cdots+k_m=n} \frac{k_1^{k_1}}{k_1!} \cdots \frac{k_m^{k_m}}{k_m!}$$

Let us introduce a *tree-generating function*

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

where $T(z)$ satisfies $T(z) = ze^{T(z)}$ ($= -W(-z)$, **Lambert's** W-function) and also

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

enumerates all rooted labeled trees. Let now

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

Then

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

Asymptotics

The function $B(z)$ has an algebraic singularity at $z = e^{-1}$ (it becomes a multi-valued function) and one finds

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

The singularity analysis yields (cf. Clarke & Barron, 1990, Szpankowski, 1998)

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

To derive asymptotics of $R_n^*(\mathcal{M}_0)$ we need $\bar{R}_n^{GS}(Q^*)$ that we just proved to be (cf. Drmota & Szpankowski, 2001)

$$R_n^{GS}(Q^*) = -\frac{\ln\frac{1}{m-1}\ln m}{\ln m} + o(1),$$

In general, the term $o(1)$ can not be improved. Thus

$$R_n^*\mathcal{M}_0) = \frac{m-1}{2}\log\left(\frac{n}{2}\right) - \frac{\ln\frac{1}{m-1}\ln m}{\ln m} + \log\left(\frac{\sqrt{\pi}}{\Gamma(\frac{m}{2})}\right) + o(1).$$

Markov Sources

Let $\mathcal{M}_1$ be a class of Markov sources of order 1 over an m-ary alphabet. Let $P = \{p_{ij}\}_{i,j=1}^{m}$ be the transition probability. To simplify we assume that the initial state of the Markov process is fixed. Then

$$P(x_1^n) = p_{11}^{k_{11}} \cdots p_{mm}^{k_{mm}}$$

where k_{ij} is the number of **pair symbols** ij, that is, i followed by j in x_1^n. Observe that k_{ij} are **not** completely independent. (Some results see Rissanen 1996, and Atteson 1999.)

We only consider **circular** strings (i.e., after the n symbol we re-visit the first symbol of x_1^n). Then k_{ij} satisfy

$$\begin{aligned} \sum_{1 \le i,j \le m} k_{ij} &= n, \\ \sum_{j=1}^{m} k_{ij} &= \sum_{j=1}^{m} k_{ji}, \ \forall\, i \ \ (\textbf{conservation flow property}) \end{aligned}$$

We denote these constraints as $\mathcal{K}_n$.

Statistics and Combinatorics

It is not difficult to see that the non-fluctuating part $d_n(\mathcal{M}_1)$ of the maximal minimax redundancy for Markov sources over m-ary alphabet is

$$d_n(\mathcal{M}_1) = \sum_{k_{ij} \in [k]} N_{[k]} \left(\frac{k_{11}}{k_1}\right)^{k_{11}} \cdots \left(\frac{k_{m,m}}{k_m}\right)^{k_{m,m}},$$

where $k_i = \sum_{j=1}^{m} k_{ij}$, the matrix $[k] = \{k_{ij}\}_{i,j=1}^{m}$ is an integer matrix whose (i,j)-th coefficients satisfy:

- $\sum_{1 \le i,j \le m} k_{ij} = n - 1$;
- the flow conservation property: $\sum_{j=1}^{m} k_{ij} = \sum_{j=1}^{m} k_{ji}$

The quantity $N_{[k]}$ is the number of string x_1^n generated over $\mathcal{A}$ having k_{ij} positions in x_1^n where j follows i (known as *frequency count*).

Problem. Let $[k] = \{k_{ij}\}_{i,j=1}^{m}$ be a given matrix satisfying the above constraint $\mathcal{K}_n$. How many strings, $N_{[k]}$, can be generated over $\mathcal{A}$ having k_{ij} pairs (i,j) such that j follows i in x_1^n? (Whittle, 1956)

Example

Example: Let $\mathcal{A} = \{0, 1\}$ and

$$[k] = \begin{bmatrix} 1 & 2 \\ 2 & 2 \end{bmatrix}$$

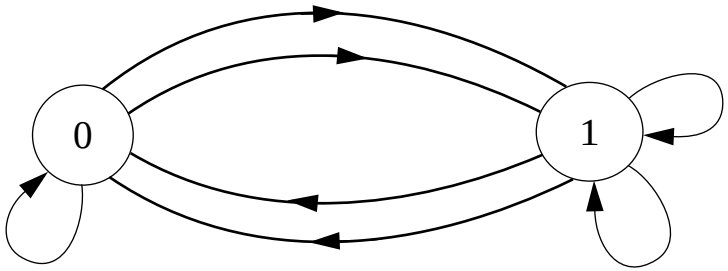


Figure 3: The directed multigraph for a binary alphabet $\mathcal{A} = \{0, 1\}$ with the matrix $[k]$ as above. $N_{[k]}$ is equal the number of string x_1^7 with matrix K_2 and it is also equal to the number of Eulerian cycles in such graph.

We can prove that

$$N_{[k]} = \alpha_{[k]} \binom{k_1}{k_{11}, \ldots k_{1m}} \cdots \binom{k_m}{k_{m1} \ldots k_{mm}}$$

where $k_i = \sum_j k_{ij}$ and $\alpha_{[k]}$ is a complicated function of the matrix $[k]$, however, it does not contribute significantly to the asymptotics of $N_{[k]}$.

Some Preliminary Results

Theorem 6. *Let $\mathcal{M}_1$ be a class of Markov sources over a finite alphabet $\mathcal{A}$ of size m. Then*

$$D_n(\mathcal{M}_1) = \left(\frac{n}{2\pi}\right)^{m(m-1)/2} A_m \times \left(1 + O\left(\frac{1}{n}\right)\right)$$

wth

$$A_m = \int_{\mathcal{K}(1)} m F_m(y_{ij}) \prod_i \frac{\sqrt{\sum_j y_{ij}}}{\prod_j \sqrt{y_{ij}}} d[y_{ij}]$$

where $\mathcal{K}(1) = \{y_{ij} : \sum_{ij} y_{ij} = 1\}$ and $F_m(\cdot)$ is a polynomial expression of degree $m-1$.

In particular, for $m = 2$ $A_2 = 16 \times$ Catalan where Catalan is Catalan's constant $\sum_i \frac{(-1)^i}{(2i+1)^2} \approx 0.915965594$.

Theorem 7. *Let $\mathcal{M}_r$ be a class of Markov sources of order r over a finite alphabet $\mathcal{A}$ of size m. Then*

$$D_n(\mathcal{M}_r) = \left(\frac{n}{2\pi}\right)^{m^r(m-1)/2} A_m^r \times \left(1 + O\left(\frac{1}{n}\right)\right)$$

where A_m^r is a constant defined in a similar fashion as A_m above.

Renewal Sources

Csiszár and Shields studied redundancy of the **renewal process** defined as follows:

- Let $T_1, T_2 \ldots$ be a sequence of i.i.d. positive-valued random variables with distribution $Q(j) = \Pr\{T_i = j\}$ 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 a renewal process we associate a **binary renewal sequence** in which the positions of the 1's are at the renewal epochs $T_0, T_0+T_1, T_0+T_1+T_2, \ldots$.
- We start with $x_0 = 1$.

Observe that the renewal process is not a Markovian process, and as a matter of fact may not be a mixing process.

We shall analyze the maximal minimax redundancy $R_n(\mathcal{R}_0)$ over the renewal process $\mathcal{R}_0$. Csiszár and Shields proved that $R_n(\mathcal{R}_0) = \Theta(\sqrt{n})$. We will provide a more precise estimate.

Some Preliminary Estimates

For a sequence

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

where $0 \le \alpha_i \le n$ 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

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

It can be proved that

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

where

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

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

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

Main Results

Theorem 8 (Flajolet and Szpankowski 1998). *Consider the class of renewal processes as defined above. The quantity r_n attains the following asymptotics*

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

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

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

It can also be observed that the quantity r_n has an intrinsic meaning by its own. Let $\mathcal{W}_n$ denote the set of all n^n sequences of length n over the alphabet $\{0, \ldots, n-1\}$. For a sequence w, take k_j to be the number of letters j in w. Then each sequence w carries a "maximum likelihood probability"

$$\pi_{ML}(w) = \left(\frac{k_0}{k}\right)^{k_0} \cdots \left(\frac{k_{n-1}}{k}\right)^{k_{n-1}}.$$

This is the probability that w gets assigned in the Bernoulli model that makes it most likely. The quantity r_n is also $r_n = \sum_{w\in\mathcal{W}_n} \pi_{ML}(w)$.

Asymptotics: Overview

As in the Markov case, we observe that the difficulty in extracting asymptotics of d_n lies in a complicated combinatorial structure of the summation index $\mathcal{P}(n,k)$. Again, combinatorics works hand-in-hand with analytic tools to solve this problem of information theory.

To give a glimpse into the analysis required for this problem, let us mention that we:

- first, we transform r_n into another quantity s_n that we know how to handle and (using a **probabilistic technique**) 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 for extracting the asymptotics of the coefficients;
- finally, estimate $R_n^*(\mathcal{R}_0)$ by the **saddle point method**.

Asymptotics: Some Details

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

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

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

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

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

Thus

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

Generating Functions

1. 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)

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

2. Define

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

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

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

We need to compute $s_n = [z^n]S(z,1)$ where $[z^n]F(z)$ denotes the coefficient at z^n of $F(z)$.

Mellin Asymptotics

3. 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.

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

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

This leads to

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

What's Next?

5. 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

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

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

To extract asymptotic we need to apply the **saddle point method**.

Average vs Maximal Minimax Redundancy

1. The average minimax redundancy $\bar{R}_n(\mathcal{S})$ is much harder to estimate than the maximal minimax redundancy $R_n^*(\mathcal{S})$.

2. We have good understanding how to estimate the maximal minimax redundancy, that is,

$$R_n^*(\mathcal{S}) = \log d_n(\mathcal{S}) + O(1)$$

where $d_n(\mathcal{S}) = \sum_{x_1^n} \sup P(x_1^n)$.

3. From known results (for memoryless and Markov sources) we observed that (see Barron & Clark, Rissanen)

$$\bar{R}_n(\mathcal{S}) \sim R_n^*(\mathcal{S}), \qquad n \to \infty.$$

4. We want to test the following conjecture:

Conjecture. For some class of sources $\mathcal{S}$ (which one?)

$$\begin{aligned} \bar{R}_n(\mathcal{S}) &= R_n^*(\mathcal{S}) + o(\log d_n(\mathcal{S}) \sim \lg \left(\sum_{x_1^n \in \mathcal{A}^n} \sup_{P \in \mathcal{S}} P(x_1^n) \right), \\ \bar{r}_n(\mathcal{S}) &\sim \underline{r}_n(\mathcal{S}) \sim \log d_n(\mathcal{S}). \end{aligned}$$

Some General Results – Redundancy

Theorem 9 (Drmota & Szpankowski, 2002). *Suppose that $\mathcal{S}$ is a system of probability distributions P on $\mathcal{A}^n$. Then*

$$\bar{R}_n(\mathcal{S}) \le \lg d_n(\mathcal{S}) - \inf_{P\in\mathcal{S}} \left(\sum_{x_1^n} P(x_1^n) \lg \frac{\sup_{P\in\mathcal{S}} P(x_1^n)}{P(x_1^n)} \right) + O(1).$$

Furthermore, if the maximal likelihood distribution Q^* is contained in the convex hull[1] *of $\mathcal{S}$ then*

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

Remark. The upper bound is quite trivial, while the lower bound relies of the following lemma:

Lemma 8. *Suppose that $\mathcal{S}$ is a subset of probability distributions P on a finite set X. Then for all probability distributions $\tilde{Q}$ contained in the convex hull of $\mathcal{S}$ we have*

$$\inf_{Q} \sup_{P\in\mathcal{S}} \left(\sum_{x\in X} P(x) \lg \frac{\tilde{Q}(x)}{Q(x)} \right) = 0.$$

[1] We assume no topology on the set of all probability measures on X. Therefore the convex hull of $\mathcal{S}$ is just the set of all finite convex combinations of elements of $\mathcal{S}$.

Special Sources

1. For memoryless and Markov sources we can prove that

$$\sup_{P\in\mathcal{S}}\left(\sum_{x_1^n} P(x_1^n)\lg\frac{\sup_{P\in\mathcal{S}} P(x_1^n)}{P(x_1^n)}\right) = O(1)\ \left(\le \frac{m-1}{\ln 2}\right).$$

2. Actually, it is known that for memoryless and Markov processes (in general, d-parameterized distributions)

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

3. However, for renewal processes it seems that

$$\bar{R}_n(\mathcal{R}_0) \sim c_1\sqrt{n}, \qquad R_n^*(\mathcal{R}_0) \sim c_2\sqrt{n}$$

but $c_1 \neq c_2$.

Some General Results – Regrets

We recall that

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

Theorem 10. *Suppose that $\mathcal{S}$ is a system of probability distributions P on $\mathcal{A}^n$. Then*

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

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

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

Maximin Regret

Theorem 11. *Suppose that $\mathcal{S}$ is a system of probability distributions P on $\mathcal{A}^n$. Let $\bar{R}_n^H(P)$ be the average minimax redundancy for the Huffman code for the distribution P. Then*

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

where $D(P||Q^*)$ denotes the Kullback–Leibler distance (relative entropy) between P and Q^*.

Why Multidisciplinary Research?

The history of scientific discovery is a testimony to the success of multidisciplinary research:

- The recent proof of Fermat's last theorem by A. Wiles provides us with a prime example of this. Recall that Wiles' success followed on the heels of the works of Frey and Ribert that connected the seemingly unrelated Taniyama-Shimura conjecture (of modular forms) and Fermat's last theorem.

- Feynman's fiddling with the axial wobble of a cafeteria plate at Cornell led him to his version of quantum mechanics. Feynman's "sum over histories" approach was picked up by mathematician Mark Kac who wrapped it in the now-famous Feynman-Kac path integral.

- The prime number conjecture had irked mathematicians for almost two thousand years until Hadamard applied complex analysis to prove it.

- The assembly of the human genome by a group of biologists, computer scientists, mathematicians and physicists is the ultimate evidence supporting multidisciplinary research.

Conclusions

- Many open problems in **analytic information theory**:
  1. Generalize Huffman redundancy to Markov sources.
  2. What code minimizes the r-th redundancy R_n^r defined for $1 < r < \infty$ as

$$R_n^{[r]} = \left(\sum_{x_1^n} P(x_1^n) \left[L(C_n, x_1^n) + \log P(x_1^n) \right]^r \right)^{1/r}.$$

  3. Compute redundancy rate(s) for mixing sources.
  4. Prove or disprove that for renewal source $\bar{R}_n(\mathcal{R}_0) \sim R_n^*(\mathcal{R}_0) \sim \frac{2}{\log 2}\sqrt{\left(\frac{\pi^2}{6} - 1\right) n}$.
- Probabilistic, analytic and combinatorial methods must work hand in hand to produce precise results of analytic information theory.
- While information theory proved to be quintessential to communications, in our opinion a non-trivial application of information theory to biology and information security still awaits us. The same applies to discrete mathematics.

The methods discussed are explained in my recent book: *Average Case Analysis of Algorithms on Sequences*, Wiley, New York, 2001.

Appendix A: 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 behavior 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).

Appendix B: 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 $x = 0^+$.

Appendix C: 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 $z = 0$.

(S1). SADDLE POINT CONTOUR. *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 $\gamma_0 = \{z \in \gamma \mid |\arg(z)| \le \varphi\}, \quad \gamma_1 = \{z \in \gamma \mid |\arg(z)| \ge \varphi\}.$

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

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

(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:

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