Minimax Redundancy for Large Alphabets by Analytic Methods

Wojciech Szpankowski *
Purdue University
W. Lafayette, IN 47907

March 20, 2012

NSF STC Center for Science of Information

CISS, Princeton 2012

*Joint work with M. Weinberger

Outline

1. Source Coding: The Redundancy Rate Problem
2. Universal Memoryless Sources
   (a) Finite Alphabet
   (b) Unbounded Alphabet
3. Universal Renewal Sources

Source Coding and Redundancy

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

Known Source P: The pointwise and maximal redundancy are:

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

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

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

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

Source Coding and Redundancy

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

Known Source P: The pointwise and maximal redundancy are:

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

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

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

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

Shtarkov's Bound:

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

Maximal Minimax Redundancy R_n^*

For the maximal minimax redundancy define

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

the maximum likelihood distribution. Observe that (Shtarkov, 1976):

$$\begin{aligned}
R_n^*(\mathcal{S}) &= \min_{C_n\in\mathcal{C}} \sup_{P\in\mathcal{S}} \max_{x_1^n} (L(C_n, x_1^n) + \log P(x_1^n)) \\
&= \min_{C_n\in\mathcal{C}} \max_{x_1^n} \left(L(C_n, x_1^n) + \sup_{P\in\mathcal{S}} \log P(x_1^n) \right) \\
&= \min_{C_n\in\mathcal{C}} \max_{x_1^n} \left(L(C_n, x_1^n) + \log Q^*(x_1^n)\right) + \log \sum_{y_1^n\in\mathcal{A}^n} \sup_{P\in\mathcal{S}} P(y_1^n) \\
&= R_n^{GS}(Q^*) + \log \sum_{y_1^n\in\mathcal{A}^n} \sup_{P\in\mathcal{S}} P(y_1^n) = \log \sum_{y_1^n\in\mathcal{A}^n} \sup_{P\in\mathcal{S}} P(y_1^n) + O(1).
\end{aligned}$$

where $R_n^{GS}(Q^*)$ is the redundancy of the optimal generalized Shannon code. Also,

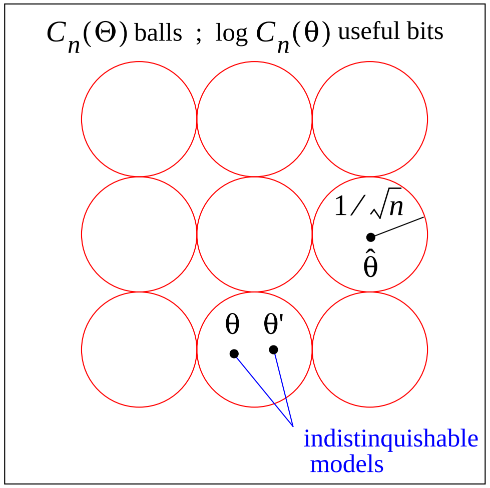
$$d_n(\mathcal{S}) = \log \left(\sum_{x_1^n\in\mathcal{A}^n} \sup_{P\in\mathcal{S}} P(x_1^n) \right) := \log D_n(\mathcal{S}).$$

Learnable Information and Redundancy

1. $\mathcal{S} := \mathcal{M}^k = \{P_\theta : \theta \in \Theta\}$ set of k-dimensional parameterized distributions. Let $\hat{\theta}(x^n) = \arg\max_{\theta\in\Theta} \log 1/P_\theta(x^n)$ be the ML estimator.

Learnable Information and Redundancy

1. $\mathcal{S} := \mathcal{M}^k = \{P_\theta : \theta \in \Theta\}$ set of k-dimensional parameterized distributions. Let $\hat{\theta}(x^n) = \arg\max_{\theta \in \Theta} \log 1/P_\theta(x^n)$ be the ML estimator.

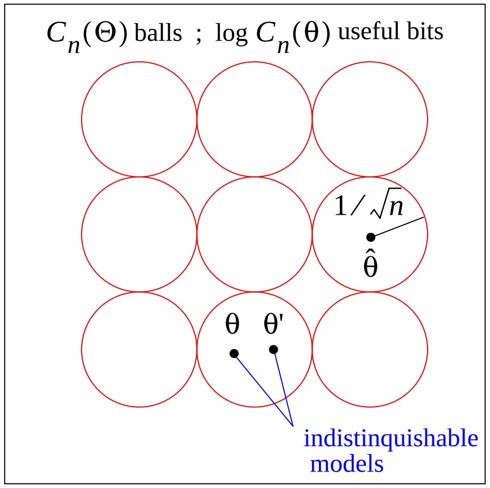


2. Two models, say $P_\theta(x^n)$ and $P_{\theta'}(x^n)$ are indistinguishable if the ML estimator $\hat{\theta}$ with high probability declares both models are the same.

3. The number of distinguishable distributions (i.e, $\hat{\theta}$), $C_n(\Theta)$, summarizes then learnable information, $I(\Theta) = \log_2 C_n(\Theta)$.

Learnable Information and Redundancy

1. $\mathcal{S} := \mathcal{M}^k = \{P_\theta : \theta \in \Theta\}$ set of k-dimensional parameterized distributions. Let $\hat{\theta}(x^n) = \arg\max_{\theta\in\Theta} \log 1/P_\theta(x^n)$ be the ML estimator.



2. Two models, say $P_\theta(x^n)$ and $P_{\theta'}(x^n)$ are indistinguishable if the ML estimator $\hat{\theta}$ with high probability declares both models are the same.

3. The number of distinguishable distributions (i.e, $\hat{\theta}$), $C_n(\Theta)$, summarizes then learnable information, $I(\Theta) = \log_2 C_n(\Theta)$.

4. Consider the following expansion of the Kullback-Leibler (KL) divergence

$$D(P_{\hat{\theta}}||P_\theta) := \mathbf{E}[\log P_{\hat{\theta}}(X^n)] - \mathbf{E}[\log P_\theta(X^n)] \sim \frac{1}{2}(\theta-\hat{\theta})^T I(\hat{\theta})(\theta-\hat{\theta}) \asymp d_I^2(\theta,\hat{\theta})$$

where $I(\theta) = \{I_{ij}(\theta)\}_{ij}$ is the *Fisher information matrix* and $d_I(\theta,\hat{\theta})$ is a rescaled Euclidean distance known as Mahalanobis distance.

5. Balasubramanian proved the number of distinguishable balls $C_n(\Theta)$ of radius $O(1/\sqrt{n})$ is asymptotically equal to the minimax redundancy:

$$\textbf{Learnable Information} = \log C_n(\Theta) = \inf_{\theta\in\Theta}\max_{x^n} \log \frac{P_{\hat{\theta}}}{P_\theta} = R_n^*(\mathcal{M}^k)$$

Outline Update

1. Source Coding: The Redundancy Rate Problem
2. Universal Memoryless Sources
   (a) Finite Alphabet
   (b) Unbounded Alphabet
3. Universal Renewal Sources

Maximal Minimax for Memoryless Sources

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

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

Then

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

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

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

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

We write

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

Let us introduce a tree-generating function

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

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

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

Then by the convolution formula

$$\boxed{D_m(z) = [B(z)]^m - 1.}$$

Asymptotics for FINITE m

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

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

By Cauchy's coefficient formula

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

Asymptotics for FINITE m

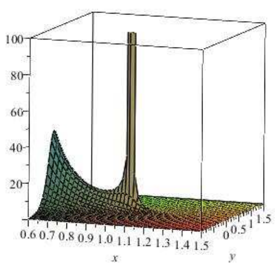
The function $B(z)$ has an algebraic singularity at $z = e^{-1}$, and

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

By Cauchy's coefficient formula

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

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



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

that finally yields (cf. Clarke & Barron, 1990, W.S., 1998)

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

Outline Update

1. Source Coding: The Redundancy Rate Problem
2. Universal Memoryless Sources
   (a) Finite Alphabet
   (b) Unbounded Alphabet
3. Universal Renewal Sources

Redundancy for LARGE m

Now assume that m is unbounded and may vary with n. Then

$$D_{n,m}(\mathcal{M}_0) = \sqrt{2\pi n}\frac{1}{2\pi i}\oint \frac{\beta(z)^m}{z^{n+1}}dz = \sqrt{2\pi n}\frac{1}{2\pi i}\oint e^{g(z)}dz$$

where $g(z) = m \ln \beta(z) - (n+1)\ln z$.

Redundancy for LARGE m

Now assume that m is unbounded and may vary with n. Then

$$D_{n,m}(\mathcal{M}_0) = \sqrt{2\pi n}\frac{1}{2\pi i}\oint \frac{\beta(z)^m}{z^{n+1}}dz = \sqrt{2\pi n}\frac{1}{2\pi i}\oint e^{g(z)}dz$$

where $g(z) = m \ln \beta(z) - (n+1)\ln z$.

The saddle point z_0 is a solution of $g'(z_0) = 0$, that is,

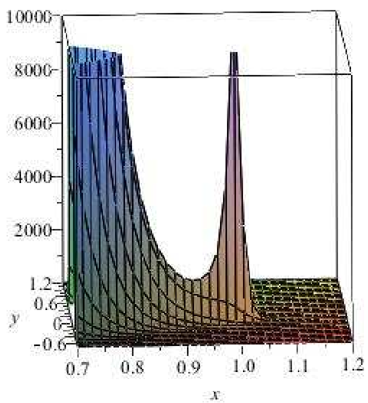
$$g(z) = g(z_0) + \frac{1}{2}(z - z_0)^2 g''(z_0) + O(g'''(z_0)(z - z_0)^3).$$

Under mild conditions satisfied by our $g(z)$ (e.g., z_0 is real and unique), the saddle point method leads to:

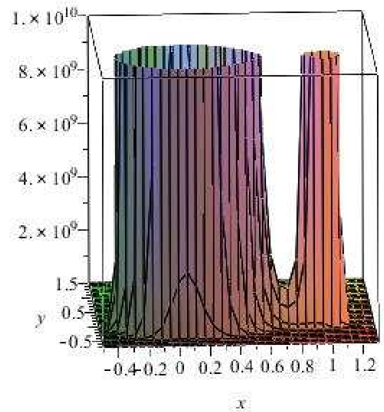
$$D_{n,m}(\mathcal{M}_0) = \frac{e^{g(z_0)}}{\sqrt{2\pi|g''(z_0)|}} \times \left(1 + O\left(\frac{g'''(z_0)}{(g''(z_0))^\rho}\right)\right),$$

for some $\rho < 3/2$.

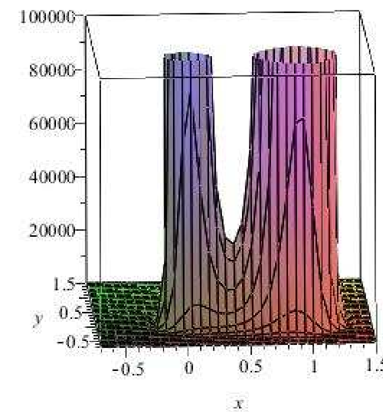
Saddle Points



$m = o(n)$



$m = n$



$n = o(m)$

Main Results – Large Alphabet

Theorem 1 (Orlitsky and Santhanam, 2004, and W.S. and Weinberger, 2010). *For memoryless sources $\mathcal{M}_0$ over an m-ary alphabet, $m \to \infty$ as n grows, we have:*

(i) *For* $m = o(n)$

$$R^*_{n,m}(\mathcal{M}_0) = \frac{m-1}{2}\log\frac{n}{m} + \frac{m}{2}\log e + \frac{m\log e}{3}\sqrt{\frac{m}{n}}\frac{1}{2} - O\left(\sqrt{\frac{m}{n}}\right)$$

(ii) *For* $m = \alpha n + \ell(n)$, *where* α *is a positive constant and* $\ell(n) = o(n)$,

$$R^*_{n,m}(\mathcal{M}_0) \quad = \quad n\log B_\alpha + \ell(n)\log C_\alpha - \log\sqrt{A_\alpha} + O(\ell(n)^2/n)$$

where $C_\alpha := 0.5 + 0.5\sqrt{1+4/\alpha}$, $A_\alpha := C_\alpha + 2/\alpha$, $B_\alpha = \alpha C_\alpha^{\alpha+2} e^{-\frac{1}{C_\alpha}}$.

(iii) *For* $n = o(m)$

$$R^*_{n,m}(\mathcal{M}_0) = n\log\frac{m}{n} + \frac{3}{2}\frac{n^2}{m}\log e - \frac{3}{2}\frac{n}{m}\log e + O\left(\frac{1}{\sqrt{n}} + \frac{n^3}{m^2}\right).$$

Constrained Memoryless Sources

Consider the following generalized source $\widetilde{\mathcal{M}}_0$:

- alphabet is $\mathcal{A} \cup \mathcal{B}$, where $|\mathcal{A}| = m$ and $|\mathcal{B}| = M$;
- probabilities $p_1, \ldots, p_m$, of $\mathcal{A}$ are unknown, while the probabilities $q_1, \ldots, q_M$ of $\mathcal{B}$ are fixed;
- define $q = q_1 + \cdots + q_M$ and $p = 1 - q$.

Our goal is to compute the minimax redundancy $R_{n,m,M}(\widetilde{\mathcal{M}}_0)$. Recall

$$R_{n,m,M}(\widetilde{\mathcal{M}}_0) = \log\left(D_{n,m,M}\right) + O(1), \qquad D_{n,m,M} = 2^{d_{n,m,M}}.$$

Lemma 1.

$$D_{n,m,M} = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} D_{k,m_k}.$$

where $D_{n,m} = R_{n,m}(\mathcal{M}_0) + O(1)$ is the minimax redundancy over $\mathcal{A}$ as presented in Theorem 1.

Proof. It basically follows from

$$D_{n,m,M} = \sum_{x \in (\mathcal{A}\cup\mathcal{B})^n} \sup_P P(x) = \sum_{y \in \mathcal{A}^{n-i}} \sum_{z \in \mathcal{B}^i} \sup P_{n-i}(y) P_i(z)$$

Binomial Sums

Consider the following binomial sum

$$S_f(n) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} f(k)$$

In our case, $f(k) = D_{n,m,M} = 2^{d_{n,m,M}}$.

Case: $f(k) = O(n^a \log^b n) = o(e^{\sqrt{n}})$:

$$S_f(n) = f(np)(1 + O(1/n))$$

(cf. Jacquet & W.S. (1999), and Flajolet (1999)).

Case: $f(k) = (\alpha^k)$

$$S_f(n) \sim (p\alpha + 1 - p)^n.$$

Case: $f(k) = O(k^{k^\beta})$

$$S_f(n) \sim p^n f(n).$$

Main Results for the Constrained Model $\widetilde{\mathcal{M}}_0$

Theorem 2 (W.S and Weinberger, 2010). *Write m_n for m depending on n.*

(i) $m_n = o(n)$. *Assume:*
(a) $m(x) := m_x$ *and its derivatives are continuous functions.*
(b) $\Delta_n := m_{n+1} - m_n = O(m'(n))$, $m'(n) = O(m/n)$, $m''(n) = O(m/n^2)$.
If $m_n = o(\sqrt{n}/\log n)$, *then*

$$R^*_{n,m,M} = \frac{m_{np}-1}{2}\log\left(\frac{np}{m_{np}}\right) + \frac{m_{np}}{2}\log e - \frac{1}{2} + O\left(\frac{1}{\sqrt{m_n}} + \frac{m_n^2 \log^2 n}{n}\right).$$

Otherwise,

$$R^*_{n,m,M} = \frac{m_{np}}{2}\log\left(\frac{np}{m_{np}}\right) + \frac{m_{np}}{2}\log e + O\left(\log n + \frac{m_n^2}{n}\log^2\frac{n}{m_n}\right).$$

(ii) *Let* $m_n = \alpha n + \ell(n)$, *where* α *is a positive constant and* $\ell(n) = o(n)$.

$$R^*_{n,m,M} = n\log\left(B_\alpha p + 1 - p\right) - \log\sqrt{A_\alpha} + O\left(\ell(n) + \frac{1}{\sqrt{n}}\right),$$

where A_α *and* B_α *are defined in Theorem 1.*

(iii) *Let* $n = o(m_n)$ *and assume* m_k/k *is a nondecreasing sequence. Then,*

$$R^*_{n,m,M} = n\log\left(\frac{pm_n}{n}\right) + O\left(\frac{n^2}{m_n} + \frac{1}{\sqrt{n}}\right).$$

Outline Update

1. Source Coding: The Redundancy Rate Problem
2. Universal Memoryless Sources
3. Universal Renewal Sources

Renewal Sources (Virtual Large Alphabet)

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

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

Renewal Sources (Virtual Large Alphabet)

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

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

For a sequence

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

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

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

Renewal Sources (Virtual Large Alphabet)

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

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

For a sequence

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

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

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

Theorem 3 (Flajolet and W.S., 1998). *Consider the class of renewal processes. Then*

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

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

Maximal Minimax Redundancy

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

$$r_n = \sum_{k=0}^{n} r_{n,k}, \quad r_{n,k} = \sum_{\mathcal{I}(n,k)} \binom{k}{k_0 \ldots k_{n-1}} \left(\frac{k_0}{k}\right)^{k_0} \left(\frac{k_1}{k}\right)^{k_1} \cdots \left(\frac{k_{n-1}}{k}\right)^{k_{n-1}}$$

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

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

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

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

since $S(z,u) = \sum_{k,n} s_{n,k} u^k z^n = \prod_{i=1}^{\infty} \beta(z^i u)$.

Maximal Minimax Redundancy

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

$$r_n = \sum_{k=0}^{n} r_{n,k}, \quad r_{n,k} = \sum_{\mathcal{I}(n,k)} \binom{k}{k_0 \dots k_{n-1}} \left(\frac{k_0}{k}\right)^{k_0} \left(\frac{k_1}{k}\right)^{k_1} \cdots \left(\frac{k_{n-1}}{k}\right)^{k_{n-1}}$$

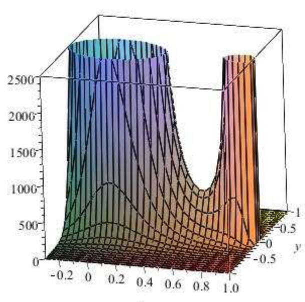
where $\mathcal{I}(n,k)$ is is the integer partition of n into k terms, i.e.,

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

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

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

since $S(z,u) = \sum_{k,n} s_{n,k} u^k z^n = \prod_{i=1}^{\infty} \beta(z^i u)$.



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

Theorem 4 (Flajolet and W.S., 1998). *We have the following asymptotics*

$$s_n \sim \exp\left(2\sqrt{cn} - \frac{7}{8}\log n + O(1)\right), \quad \log r_n = \frac{2}{\log 2}\sqrt{cn} - \frac{5}{8}\log n + \frac{1}{2}\log\log n + O(1).$$

That's IT