A lower bound on compression of unknown alphabets

Nikola Jevtić[1] Alon Orlitsky[1,2] Narayana P. Santhanam[1]

[1]ECE and [2]CSE Departments

University of California, San Diego

{njevtic,alon,nsanthan}@ucsd.edu

October 20, 2004

Abstract

Many applications call for universal compression of strings over large, possibly infinite, alphabets. However, it has long been known that the resulting redundancy is infinite even for *i.i.d.* distributions. It was recently shown that the redudancy of the strings' *patterns*, which abstract the values of the symbols, retaining only their relative precedence, is sublinear in the blocklength n, hence the per-symbol redundancy diminishes to zero. In this paper we show that pattern redundancy is at least $(1.5 \log_2 e)\, n^{1/3}$ bits. To do so, we construct a generating function whose coefficients lower bound the redundancy, and use Hayman's saddle-point approximation technique to determine the coefficients' asymptotic behavior.

1 Introduction

Many applications require compression of data generated by an unknown distribution. For example, while data often needs to be compressed to accomodate bandwidth constraints in wireless communications, its distribution is rarely known.

A typical approach to this problem assumes that the underlying distribution, though unknown, belongs to a known collection $\mathcal{P}$ of possible distributions, for example, the set of *i.i.d.* or Markov distributions.

When the underlying distribution is known, sources can be compressed to essentially their entropy. Even with the distribution unknown, we attempt to compress the data with a *universal* code so that the number of bits used is not much larger than the entropy of the underlying distribution, no matter which one in $\mathcal{P}$ it may be. The minimum number of extra bits used by any universal code in the worst case is the *redundancy*, $\hat{R}(\mathcal{P})$ of the collection $\mathcal{P}$ of distributions.

Let $\mathcal{P}$ be a collection of distributions over a set $\mathcal{X}$. Shtarkov [1] showed that

$$\hat{R}(\mathcal{P}) = \log\left(\sum_{x\in\mathcal{X}} \max_{p\in\mathcal{P}} p(x)\right), \tag{1}$$

where throughout the paper, logarithms are taken to base 2. For the collection $\mathcal{I}_m^n$ of *i.i.d.* distributions over length-n strings from an alphabet of a fixed size m, a number of researchers have shown [2, 3, 4, 5, 6, 7, 8, 9] that as n increases

$$\hat{R}(\mathcal{I}_m^n) = \frac{m-1}{2}\log\frac{n}{2\pi} + \log\frac{\Gamma^m(\frac{1}{2})}{\Gamma(\frac{m}{2})} + o_m(1), \tag{2}$$

where Γ is the gamma function, and the $o_m(1)$ term diminishes with increasing n at a rate determined by m.

This redundancy grows logarithmically with the blocklength n, hence as n increases, the per-symbol redundancy $\hat{R}(\mathcal{I}_m^n)/n$ diminishes, implying that asymptotically, strings generated by an unknown distribution in $\mathcal{I}_m^n$ can be compressed essentially as well as when the underlying distribution is known.

However, in many applications, such as language modeling, text, and image compression, the alphabet size m is large, often comparable to the blocklength, and the redundancy calculated in (2) high.

In the limit, Kieffer [10] showed that universal compression of *i.i.d.* sequences over infinite alphabets requires infinite per-symbol redundancy, and specified the condition for a collection of distributions to have negligible per-symbol redundancy.

Motivated by (2) and by Kieffer's result, researchers have generally avoided direct compression of sequences over infinite or large alphabets. Therefore, several researchers have attempted to get around Kieffer's negative result.

One line of work, along the lines of [11, 12, 13, 14, 15] constructed compression algorithms for collections satisfying Kieffer's condition. For example, [11] considered the collection of *i.i.d.* distributions that assign non-increasing probabilities to positive integers.

A second approach [16, 17] does not restrict the collection of distributions, but separates the description of the sequence into two parts: a description of the symbols appearing in the string, and of the *pattern* they form. For example, the string

"abracadabra"

can be described by conveying the *pattern*, Ψ("abracadabra"),

"12314151231"

and the *dictionary*

index	1	2	3	4	5
letter	a	b	r	c	d

.

Together, the pattern and the dictionary specify that the string "abracadabra" consists of the first letter to appear (a), followed by the second letter to appear (b), then by the third to appear (r), the first that appeared (a again), the fourth (c), the first (a), etc.

In many applications [18, 19, 20, 21, 22, 23, 24], the description of patterns is more important than the dictionary. For example, in language modeling, the pattern reflects the structure of the language while the dictionary plays a less important part. Consequently, we concentrate on the redundancy of compressing the patterns.

Any distribution induces a distribution on patterns, assigning to a pattern $\overline{\psi}$ the probability

$$p(\overline{\psi}) \stackrel{\text{def}}{=} p\{\overline{x} : \Psi(\overline{x}) = \overline{\psi}\}$$

of all sequences whose pattern is $\overline{\psi}$.

Letting Ψ^n denote the set of all length-n patterns, Shtarkov's sum (1) implies that the *pattern redundancy* of $\mathcal{I}^n$, *i.e.,* the redundancy of the collection of distributions induced on patterns by $\mathcal{I}^n$ is

$$\hat{R}(\mathcal{I}^n_\Psi) = \log\left(\sum_{\overline{\psi}\in\Psi^n} \max_{p\in\mathcal{I}^n} p(\overline{\psi})\right).$$

It has been shown [25] that, irrespective of the alphabet size, patterns of *i.i.d.* distributed strings can be compressed with redundancy of at most

$$\hat{R}(\mathcal{I}^n_\Psi) \le \left(\pi\sqrt{\frac{2}{3}}\log e\right)\sqrt{n}$$

bits. Hence as the blocklength n grows, the redundancy of patterns increases sublinearly with n, and the per-symbol redundancy diminishes to zero, even for infinite alphabets.

In this paper we improve on a lower bound on $\hat{R}(\mathcal{I}^n_\Psi)$ presented in [25]. To do so, we lower bound the highest probability of a pattern $\overline{\psi}$ by the highest probability of any single *i.i.d.* string whose pattern is $\overline{\psi}$. We obtain,

$$\hat{R}(\mathcal{I}^n_\Psi) \ge \hat{R}^-(\mathcal{I}^n_\Psi) \stackrel{\text{def}}{=} \log\left(\sum_{\overline{\psi}\in\Psi^n}\prod_{\mu=1}^{n}\left(\frac{\mu}{n}\right)^{\mu\varphi_\mu(\overline{\psi})}\right),$$

where $\varphi_\mu(\overline{\psi})$ is the number of symbols appearing μ times in $\overline{\psi}$. $\hat{R}^-(\mathcal{I}^n_\Psi)$ is of mathematical interest of its own and its simple formulation allows for a precise evaluation of its growth order. In Theorem 10, we use Hayman's saddle point analysis on its generating function to show that

$$\hat{R}^-(\mathcal{I}^n_\Psi) = \left(\frac{3}{2}\log e\right)n^{\frac{1}{3}} - \frac{1}{3}\log n - \frac{2}{3}\log e - \frac{1}{2}\log 3 + o(1). \tag{3}$$

This bound is related to the bound for fixed m,

$$\hat{R}(\mathcal{I}^n_m) \ge \frac{m-1}{2}\log\frac{n}{2\pi} + \log\frac{\Gamma^m(\frac{1}{2})}{\Gamma(\frac{m}{2})} + o_m(1), \tag{4}$$

presented in Appendix 1 of [4]. However, it is not clear whether the latter bound holds when m grows with n (see discussion in Section 5.2 of [26]). If Bound (4) held for m growing as $\sqrt[3]{n}$, then it could be applied to obtain a lower bound on $\hat{R}(\mathcal{I}_\Psi^n)$, as described in [16]. However, this approach of [16] would yield only the (matching) leading coefficient of Bound (3), and it can be shown that if additional coefficients were calculated, they would not exceed those in Bound (3). We note that recently Shamir [27] showed that a weaker form of the lower bound in Corollary 11 applies to average-case redundancy.

An interesting property of patterns arises also in connection with Good Turing estimators. Applications considered here, like language modeling, text compression etc, typically involve distributions over a large alphabet, with most letters in the alphabet having insignificant probabilities. The maximum likelihood estimates of the distribution from a given data sample are known to be unreliable for these applications, and of the several alternatives proposed, the Good Turing estimate and its modifications are known to perform well. In contrast to the maximum likelihood estimate which explains the count of each symbol, these estimates look at the number of symbols appearing once, twice, and so on. But *i.i.d.* sources on patterns are defined by this statistic, and it can be shown, *e.g.,* [25], that the *i.i.d.* distribution assigning the highest probability to a pattern is the best *i.i.d.* distribution to explain this statistic. For a detailed discussion along this angle, see [28, 29].

2 Patterns and their redundancy

We formally define patterns and discuss their compression.

Let $\mathcal{A}$ be any alphabet. For $\overline{x} = x_1 \dots x_n \stackrel{\text{def}}{=} x_1^n \in \mathcal{A}^n$,

$$\mathcal{A}(\overline{x}) \stackrel{\text{def}}{=} \{x_1, \dots, x_n\}$$

is the set of symbols appearing in $\overline{x}$. The *index* of $x \in \mathcal{A}(\overline{x})$ is

$$\imath_{\overline{x}}(x) \stackrel{\text{def}}{=} \min\{|\mathcal{A}(x_1^i)| : 1 \le i \le n \text{ and } x_i = x\},$$

one more than the number of distinct symbols preceding x's first appearance in $\overline{x}$. The *pattern* of $\overline{x}$ is the concatenation

$$\Psi(\overline{x}) \stackrel{\text{def}}{=} \imath_{\overline{x}}(x_1)\imath_{\overline{x}}(x_2)\dots\imath_{\overline{x}}(x_n),$$

of all indices. For example, if $\overline{x} =$ "*abracadabra*", $\imath_{\overline{x}}(a) = 1$, $\imath_{\overline{x}}(b) = 2$, $\imath_{\overline{x}}(r) = 3$, $\imath_{\overline{x}}(c) = 4$, and $\imath_{\overline{x}}(d) = 5$, hence

$$\Psi(abracadabra) = 12314151231.$$

Let

$$\Psi(\mathcal{A}^n) = \{\Psi(\overline{x}) : \overline{x} \in \mathcal{A}^n\}$$

denote the set of patterns of all strings in $\mathcal{A}^n$. For example, if $\mathcal{A}$ consists of two elements, then $\Psi(\mathcal{A}) = \{1\}$, $\Psi(\mathcal{A}^2) = \{11, 12\}$, $\Psi(\mathcal{A}^3) = \{111, 112, 121, 122\}$, etc. Let

$$\Psi^n \stackrel{\text{def}}{=} \bigcup_{\mathcal{A}} \Psi(\mathcal{A}^n)$$

be the set of all length-n patterns, and let

$$\Psi^* \stackrel{\text{def}}{=} \bigcup_{n=0}^{\infty} \Psi^n$$

be the set of all patterns. For example,

$$\begin{aligned}
\Psi^0 &= \{\lambda\}, \\
\Psi^1 &= \{1\}, \\
\Psi^2 &= \{11, 12\}, \\
\Psi^3 &= \{111, 112, 121, 122, 123\},
\end{aligned}$$

where λ is the empty string, and so on.

It is easy to see that a string $\overline{\psi}$ is a pattern if and only if the first occurance of any $i \in \mathbb{Z}^+$ in $\overline{\psi}$ precedes that of $i+1$. For example, 1, 12, and 1213 are patterns, while 21, 1312 are not.

Every probability distribution p over $\mathcal{A}^*$, the collection of all strings of symbols from $\mathcal{A}$ induces a distribution p_Ψ over patterns on Ψ^*, where

$$p_\Psi(\overline{\psi}) \stackrel{\text{def}}{=} p(\{\overline{x} \in \mathcal{A}^* : \Psi(\overline{x}) = \overline{\psi}\}),$$

is the probability that a string generated according to p has pattern $\overline{\psi}$. When pattern probabilities $p_\Psi(\overline{\psi})$ are evaluated, the subscript Ψ can be inferred, and is hence omitted. For example, let p be a uniform distribution over $\{a, b\}^2$. Then p induces on Ψ^2 the distribution

$$\begin{aligned}
p(11) &= p(\{aa, bb\}) = \frac{1}{2}, \\
p(12) &= p(\{ab, ba\}) = \frac{1}{2}.
\end{aligned}$$

For a collection $\mathcal{P}$ of distributions over $\mathcal{A}^*$ let

$$\mathcal{P}_\Psi \stackrel{\text{def}}{=} \{p_\Psi : p \in \mathcal{P}\}$$

denote the collection of distributions over Ψ^* induced by probability distributions in $\mathcal{P}$.

The *pattern redundancy* of $\mathcal{P}$, namely the worst case redundancy of universally coding patterns generated according to an unknown distribution in $\mathcal{P}_\Psi$, is

$$\hat{R}(\mathcal{P}_\Psi) = \min_q \max_{p \in \mathcal{P}_\Psi} \max_{\psi \in \Psi^*} \log \frac{p(\psi)}{q(\psi)},$$

where q is any distribution on Ψ^*. Clearly, the pattern redundancy of all $\mathcal{P}$ is non-negative.

We will be mainly interested in pattern redundancy of the collection $\mathcal{I}^n$ of all *i.i.d.* distributions, that is, we compare any code q's probabilities to the *maximum i.i.d. probabilities* of patterns, $\hat{p}(\psi) \stackrel{\text{def}}{=} \max_{p \in \mathcal{I}^n_\Psi} p(\psi)$.

We illustrate the computation of maximum probabilities for a few simple patterns. Observe that since any distribution p has $p(1) = 1$, we have $\hat{p}(1) = 1$. Since any distribution p concentrated on a single element has $p(1 \ldots 1) = 1$ for any number of 1's, we obtain $\hat{p}(1 \ldots 1) = 1$, and, since any continuous distribution p has $p(12 \ldots n) = 1$, we derive $\hat{p}(12 \ldots n) = 1$.

In general it is difficult to determine the maximum probability of a pattern. For example, some work [30] is needed to show that $\hat{p}(112) = \frac{1}{4}$. Since it is difficult to obtain the maximum probability of patterns, it is difficult to compute the pattern redundancy of $\mathcal{I}^n$ exactly. In [25], an upper bound was obtained for the redundancy of patterns, showing that the per-symbol pattern redundancy of $\mathcal{I}^n$ diminishes to zero with increasing blocklengths. However, we prove here that the pattern redundancy of $\mathcal{I}^n$ is not less than $\mathcal{O}(n^{\frac{1}{3}})$.

3 The generating function

As mentioned earlier, it is difficult to obtain the maximum probability of patterns. Instead, we lower bound these probabilities of patterns, and use Shtarkov's sum to derive a lower bound on redundancy.

Let

$$\Psi_p^{-1}(\overline{\psi}) = \{\overline{x} \in \mathcal{A}^* : \Psi(\overline{x}) = \overline{\psi} \text{ and } p(\overline{x}) > 0\}$$

be the *support* of a pattern $\overline{\psi}$ with respect to a distribution p. For every $\overline{\psi} \in \Psi^n$,

$$\max_{p \in \mathcal{I}^n_\Psi} p(\overline{\psi}) = \max_{p \in \mathcal{I}^n} p\big(\Psi_p^{-1}(\overline{\psi})\big) \geq \max_{p \in \mathcal{I}^n} \max_{\overline{x} \in \Psi_p^{-1}(\overline{\psi})} p(\overline{x}).$$

Let the number of symbols occuring μ times in $\overline{\psi}$ be φ_μ. Standard maximum-likelihood arguments imply that

$$\max_{p \in \mathcal{I}^n} \max_{\overline{x} \in \Psi_p^{-1}(\overline{\psi})} p(\overline{x}) = \prod_{\mu=1}^{n} \left(\frac{\mu}{n}\right)^{\mu\varphi_\mu},$$

hence

$$\max_{p \in \mathcal{I}^n_\Psi} p(\overline{\psi}) \geq \prod_{\mu=1}^{n} \left(\frac{\mu}{n}\right)^{\mu\varphi_\mu}. \tag{5}$$

Let

$$\Phi^n = \{(\varphi_1, \ldots, \varphi_n) : \varphi_i \geq 0, \sum_{\mu=1}^{n} \mu\varphi_\mu = n\},$$

and

$$\Psi_{\overline{\varphi}} = \{\overline{\psi} : \varphi_\mu \text{ symbols appear } \mu \text{ times in pattern } \overline{\psi}\}.$$

Incorporating (5) into Shtarkov's sum (1), we obtain

$$\begin{aligned}
\hat{R}(\mathcal{I}_\Psi^n) &= \log\left(\sum_{\overline{\varphi}\in\Phi^n}\sum_{\overline{\psi}\in\Psi_{\overline{\varphi}}}\max_{p\in\mathcal{I}_\Psi^n} p(\overline{\psi})\right)\\
&\geq \log\left(\sum_{\overline{\varphi}\in\Phi^n}\sum_{\overline{\psi}\in\Psi_{\overline{\varphi}}}\prod_{\mu=1}^{n}\left(\frac{\mu}{n}\right)^{\mu\varphi_\mu}\right)\\
&= \log\left(\sum_{\overline{\varphi}\in\Phi^n}\frac{n!}{\prod_{\mu=1}^{n}(\mu!)^{\varphi_\mu}\varphi_\mu!}\prod_{\mu=1}^{n}\left(\frac{\mu}{n}\right)^{\mu\varphi_\mu}\right)\\
&\stackrel{\text{def}}{=} \log g(n). \qquad (6)
\end{aligned}$$

Direct computation of $g(n)$ appears to be difficult. Instead, we evaluate a *generating function* of $g(n)$,

$$G(z) \stackrel{\text{def}}{=} \sum_{n=0}^{\infty} g(n)\frac{n^n}{n!}z^n, \qquad (7)$$

from which the asymptotics of $g(n)$ can be obtained using *Hayman's* analysis [31].

To express the generating function $G(z)$ in a more explicit form, observe that

$$\begin{aligned}
G(z) &= \sum_{n=0}^{\infty}\sum_{(\varphi_1,\ldots,\varphi_n)\in\Phi^n}\prod_{\mu\geq 1}\left(\frac{\mu^\mu z^\mu}{\mu!}\right)^{\varphi_\mu}\frac{1}{\varphi_\mu!}\\
&= \sum_{(\varphi_1,\ldots,\varphi_n)\in\Phi^n}\prod_{\mu\geq 1}\left(\frac{\mu^\mu z^\mu}{\mu!}\right)^{\varphi_\mu}\frac{1}{\varphi_\mu!}\\
&= \prod_{\mu\geq 1}\sum_{(\varphi_1,\ldots,\varphi_n)\in\Phi^n}\left(\frac{\mu^\mu z^\mu}{\mu!}\right)^{\varphi_\mu}\frac{1}{\varphi_\mu!},
\end{aligned}$$

thus yielding

$$G(z) = \exp\left(\sum_{k=1}^{\infty}\frac{k^k z^k}{k!}\right). \qquad (8)$$

4 Hayman's analysis

In the last section, we lower bounded $\hat{R}(\mathcal{I}_\Psi^n)$ in terms of the coefficients of a generating function $G(z)$. Hayman [31] developed a technique to compute the asymptotics of the coefficients of power series that satisfy certain properties, which, as shown later, $G(z)$ also satisfies. In this section we describe Hayman's analysis. We follow the terminology used in [32].

Theorem 1. [**Hayman**] For

$$f(z) = \sum_{n=0}^{\infty} a_n z^n,$$

let

$$a(z) \stackrel{\text{def}}{=} \frac{d \log f(z)}{d \log z} \quad \text{and} \quad b(z) \stackrel{\text{def}}{=} \frac{d^2 \log f(z)}{d(\log z)^2} = z a'(z), \tag{9}$$

and let the *saddle point* r_n be the solution of

$$a(r_n) = n.$$

If for some real R_1, the following three conditions hold:

Nonnegativity: $\exists R_0 < R_1$ such that for $R_0 < x < R_1$,

$$f(x) \geq 0;$$

Fast growth: As $x \to R_1 - 0$, namely, x approaches R_1 from below, $b(x) \to \infty$;

Basic split: $\exists \phi(x) > 0$, called the *basic split* such that

Local approximation: for $|\theta| \leq \phi(x)$, uniformly in θ as $x \to R_1$

$$f(xe^{i\theta}) \sim f(x) \exp(ia(x)\theta - \frac{\theta^2}{2} b(x));$$

Fast taper: for $\phi(x) < |\theta| < \pi$, uniformly in θ as $x \to R_1$

$$f(xe^{i\theta}) \sim \frac{o(f(x))}{\sqrt{b(x)}};$$

then,

$$a_n \sim \frac{f(r_n)}{r_n^n \sqrt{2\pi b(r_n)}}. \qquad \square$$

Hayman's analysis can also be viewed as a special case of the class of *saddle point approximations*. It exploits the fact that for functions satisfying the conditions of Theorem 1, the value of Cauchy's integral $\int_C f(z)/z^{n+1}$ around a contour C through the saddle point r_n is captured by a short arc around r_n. For more details on the saddle point approximation and related results, see [31, 33, 32].

For the generating function G defined in Equation (8), the functions $a(z)$ and $b(z)$ of Equation (9) are

$$a(z) = \sum_{k=1}^{\infty} \frac{k^{k+1} z^k}{k!} \quad \text{and} \quad b(z) = \sum_{k=1}^{\infty} \frac{k^{k+2} z^k}{k!}. \tag{10}$$

We pick $R_1 = \frac{1}{e}$. The first two conditions are clearly satisfied for $G(z)$. For $\phi(x) = (1-ex)^{\frac{6}{5}}$, we show in Theorem 6, that the local approximation for G holds, and in Theorem 7 that $|G(z)|$ does drop rapidly for $|\theta| \geq \phi(x)$.

5 Preliminaries

We outline some results that will be extensively used in this paper.

Observe that we can expand $G(xe^{i\theta})$ in θ as

$$G(xe^{i\theta}) = G(x)\exp\left(\sum_{l=1}^{\infty}\frac{(i\theta)^l}{l!}\frac{d^l\log G(z)}{d(\log z)^l}\bigg|_{z=x}\right) = G(x)\exp\left(\sum_{l=1}^{\infty}\frac{(i\theta)^l}{l!}\sum_{k=1}^{\infty}\frac{k^{k+l}x^k}{k!}\bigg|_{z=x}\right).$$

We first check for convergence of each of the summations over k. Indeed,

Lemma 2. For any l, $\sum_{k=1}^{\infty}\frac{k^{k+l}x^k}{k!}$ converges for $x < \frac{1}{e}$.
Proof By the Cauchy ratio test, *e.g.,* [34]. □

Therefore, in order to evaluate the n'th coefficient in the Taylor series, Hayman's theorem approximates the value of $G(z)$ in the complex integration over the circle $|z| = x$ by a correction over the value $G(x)$ for points on the circle near the positive real line, and by a term much smaller than $G(x)$ for points on the circle away from the positive real line.

Intuitively speaking it follows that at the basic split ϕ, the contribution of higher order terms is negligible and that the contribution of the second coefficient is large enough to satisfy fast taper. We choose a ϕ based on these criteria, and then prove that our choice indeed works.

We also use *Feller's bounds* [35] on Stirling's approximation for all $n \geq 1$,

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \leq n! \leq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n}}, \tag{11}$$

extensively in the paper. Further, we shall denote by C positive constants that are, in particular, independent of x, θ and l.

6 Locating the basic split

We locate the basic split ϕ for

$$G(xe^{i\theta}) = G(x)\exp\left(\sum_{l=1}^{\infty}\frac{(i\theta)^l}{l!}\sum_{k=1}^{\infty}\frac{k^{k+l}x^k}{k!}\right).$$

To do so, we estimate the magnitude of the coefficients of θ, and ensure that at our choice of the ϕ, the second term is unbounded, and the contribution of any term beyond the second is negligible. In Theorems 6 and 7, we show that this choice works.

We upper bound the magnitude of the coefficients of θ as follows.

Lemma 3. For integers $l \geq 2$ and $x < \frac{1}{e}$,

$$\sum_{k=1}^{\infty}\frac{k^{k+l}x^k}{k!} \leq \frac{\sqrt{(2l)!}}{2^l(1-ex)^{l+\frac{1}{2}}}.$$

Proof From Feller's bounds (11),

$$\sum_{k=1}^{\infty} \frac{k^{k+l} x^k}{k!} \leq \frac{1}{\sqrt{2\pi}} \sum_{k=1}^{\infty} \frac{k^l x^k e^k}{\sqrt{k}}.$$

Squaring the right side,

$$\begin{aligned}
\left(\sum_{k=1}^{\infty} \frac{k^l x^k e^k}{\sqrt{k}} \right)^2 &= \sum_{k=2}^{\infty} x^k e^k \sum_{m=1}^{k-1} ((k-m) \cdot m)^{l-\frac{1}{2}} \\
&\leq \sum_{k=2}^{\infty} x^k e^k (k-1) \left(\frac{k}{2} \cdot \frac{k}{2} \right)^{l-\frac{1}{2}} \\
&\leq \frac{1}{2^{2l-1}} \sum_{k=0}^{\infty} k^{2l} x^k e^k \\
&\leq \frac{(2l)!}{2^{2l-1}} \sum_{k=0}^{\infty} \binom{k+2l}{2l} x^k e^k \\
&= \frac{(2l)!}{2^{2l-1} \cdot (1-ex)^{2l+1}}.
\end{aligned}$$

Taking the positive square root proves the lemma. □

We lower bound the magnitude of the coefficient of θ^2 as follows.

Lemma 4. For $\frac{5}{6e} < x < \frac{1}{e}$,

$$\sum_{k=1}^{\infty} \frac{k^{k+2} x^k}{k!} \geq \frac{C}{(1-ex)^{\frac{5}{2}}}.$$

Proof From Feller's bounds (11),

$$\sum_{k=1}^{\infty} \frac{k^{k+2} x^k}{k!} \geq C_1 \sum_{k=1}^{\infty} \frac{k^2 x^k e^k}{\sqrt{k}}.$$

Squaring the right side,

$$\begin{aligned}
\left(\sum_{k=1}^{\infty} \frac{k^2 x^k e^k}{\sqrt{k}}\right)^2 &= \sum_{k=2}^{\infty} x^k e^k \sum_{m=1}^{k-1} ((k-m)\cdot m)^{\frac{3}{2}} \\
&\geq \sum_{k=2}^{\infty} x^k e^k \sum_{m=\lfloor \frac{k}{4} \rfloor}^{\lceil \frac{3k}{4} \rceil} ((k-m)\cdot m)^{\frac{3}{2}} \\
&\geq \sum_{k=2}^{\infty} x^k e^k \cdot \frac{k}{2} \left(\frac{3k}{4}\cdot\frac{k}{4}\right)^{\frac{3}{2}} \\
&= C_2 \sum_{k=2}^{\infty} k^4 x^k e^k \\
&\geq C_2 \cdot 4! \cdot \sum_{k=2}^{\infty} \binom{k}{4} x^k e^k \\
&= C_2 \cdot 4! \cdot (xe)^4 \cdot \sum_{k=4}^{\infty} \binom{k}{4} x^{k-4} e^{k-4} \\
&\geq \frac{C_3}{(1-ex)^5}.
\end{aligned}$$

In the last step we observed that $\frac{5}{6} < xe < 1$, and thus included it in the constant. Taking the positive square root proves the lemma. □

The following Lemma locates the basic split.

Lemma 5. $\exists \phi(x)$ so that

$$\lim_{x\to\frac{1}{e}} \phi(x)^2 \sum_{k=1}^{\infty} \frac{k^{k+2}x^k}{k!} \to \infty$$

and simultaneously for $l \geq 3$,

$$\lim_{x\to\frac{1}{e}} \phi(x)^l \sum_{k=1}^{\infty} \frac{k^{k+l}x^k}{k!} = 0.$$

Proof Take $\phi(x) = (1-ex)^\alpha$ with $\frac{7}{6} < \alpha < \frac{5}{4}$. From Lemma 4,

$$\lim_{x\to\frac{1}{e}} \phi(x)^2 \sum_{k=1}^{\infty} \frac{k^{k+2}x^k}{k!} \geq \lim_{x\to\frac{1}{e}} \frac{C\cdot\phi(x)^2}{(1-ex)^{\frac{5}{2}}} \to \infty$$

because $\alpha < \frac{5}{4}$, and from Lemma 3,

$$\lim_{x\to\frac{1}{e}} \phi(x)^l \sum_{k=1}^{\infty} \frac{k^{k+l}x^k}{k!} \leq \lim_{x\to\frac{1}{e}} \frac{\phi(x)^l\sqrt{(2l)!}}{2^l(1-ex)^{l+\frac{1}{2}}} = 0$$

because $\alpha > \max_{l\geq 3}(1+\frac{1}{2l}) = \frac{7}{6}$. Therefore all $\phi(x) = (1-ex)^\alpha$ with $\frac{7}{6} < \alpha < \frac{5}{4}$ satisfy the lemma. In particular we will be using $\phi(x) = (1-ex)^{\frac{6}{5}}$. □

7 Local approximation

We show that all points on the circle $|z| = x$ with argument $|\theta| \leq \phi(x) = (1-ex)^{\frac{6}{5}}$ can be approximated by a small correction over the value on the positive real line.

Theorem 6. Let $\phi(x) = (1-ex)^{\frac{6}{5}}$. Uniformly in θ, for $0 \leq |\theta| \leq \phi(x)$,

$$G(xe^{i\theta}) \sim G(x) \exp\left(i\theta a(x) - \frac{\theta^2}{2} b(x)\right),$$

i.e., for $0 \leq |\theta| \leq \phi(x)$, $\forall\ \epsilon > 0$, $\exists\ \delta(\epsilon)$ such that if $0 < |x - \frac{1}{e}| < \delta$,

$$\left| \frac{G(xe^{i\theta})}{G(x)\exp\left(i\theta a(x) - \frac{\theta^2}{2} b(x)\right)} - 1 \right| < \epsilon.$$

Proof Observe that

$$\begin{aligned} G(xe^{i\theta}) &= \exp\left(\sum_{k=1}^{\infty} \frac{k^k x^k e^{ik\theta}}{k!}\right) \\ &= \exp\left(\sum_{k=1}^{\infty} \frac{k^k x^k}{k!} \sum_{l=0}^{\infty} \frac{(ik\theta)^l}{l!}\right) \\ &= \exp\left(\sum_{l=0}^{\infty} \frac{(i\theta)^l}{l!} \sum_{k=1}^{\infty} \frac{k^{k+l} x^k}{k!}\right). \end{aligned}$$

The rearrangement can be done for all $x < \frac{1}{e}$, as the original series is absolutely convergent for $x < \frac{1}{e}$. Split the term in the exponent as,

$$\begin{aligned} \sum_{l=0}^{\infty} \frac{(i\theta)^l}{l!} \sum_{k=1}^{\infty} \frac{k^{k+l} x^k}{k!} &= \sum_{k=1}^{\infty} \frac{k^k x^k}{k!} + i\theta \sum_{k=1}^{\infty} \frac{k^{k+1} x^k}{k!} - \frac{\theta^2}{2} \sum_{k=1}^{\infty} \frac{k^{k+2} x^k}{k!} + \sum_{l=3}^{\infty} \frac{(i\theta)^l}{l!} \sum_{k=1}^{\infty} \frac{k^{k+l} x^k}{k!} \\ &= \log(G(x)) + i\theta a(x) - \frac{\theta^2}{2} b(x) + \sum_{l=3}^{\infty} \frac{(i\theta)^l}{l!} \sum_{k=1}^{\infty} \frac{k^{k+l} x^k}{k!}. \end{aligned}$$

Observing that if $|t| < \epsilon \leq 1$, $|e^t - 1| \leq \left|e^{|t|} - 1\right| < (e-1)|t| < (e-1)\epsilon$, an equivalent statement for the local approximation would be that for $|\theta| < \phi(x)$, given $\epsilon > 0$, $\exists\ \delta(\epsilon)$ such that for $|x - \frac{1}{e}| < \delta$,

$$\left| \sum_{l=3}^{\infty} \frac{(i\theta)^l}{l!} \sum_{k=1}^{\infty} \frac{k^{k+l} x^k}{k!} \right| < \epsilon.$$

To reduce the above expression note that each term, $\frac{(i\theta)^l}{l!} \sum_{k=1}^{\infty} \frac{k^{k+l} x^k}{k!}$, approaches 0 as

$x \to \frac{1}{e}$, and that the summation converges by Cauchy's root test *e.g.,* [36]. Therefore,

$$\begin{aligned}
\left|\sum_{l=3}^{\infty}\frac{(i\theta)^l}{l!}\sum_{k=1}^{\infty}\frac{k^{k+l}x^k}{k!}\right| &\overset{(a)}{\leq} \sum_{l=3}^{\infty}\frac{|\theta|^l}{l!}\sum_{k=1}^{\infty}\frac{k^{k+l}x^k}{k!} \\
&\overset{(b)}{\leq} \sum_{l=3}^{\infty}\frac{\phi(x)^l}{l!}\frac{\sqrt{(2l)!}}{2^l(1-ex)^{l+\frac{1}{2}}} \\
&= \sum_{l=3}^{\infty}(1-ex)^{\frac{l}{5}-\frac{1}{2}}\frac{\sqrt{(2l)!}}{2^l l!} \\
&= \left(\sum_{m=0}^{\infty}\frac{(1-ex)^{\frac{m}{5}}\sqrt{(2m+6)!}}{2^{m+3}(m+3)!}\right)(1-ex)^{\frac{1}{10}}. \qquad (12)
\end{aligned}$$

(a) is the mod-sum inequality and (b) from Lemma 3. Observe that the coefficient of $(1-ex)^{\frac{1}{10}}$ converges when $ex < 1$. To see this, observe that each term is finite when $x \to \frac{1}{e}$, and use Cauchy's root test for the convergence of the series. Since expression (12) can be made smaller than ϵ by taking x close enough to $\frac{1}{e}$, the theorem follows. □

8 Fast taper

We prove that our choice, $\phi(x) = (1-ex)^{\frac{6}{5}}$ from Lemma 5 is indeed a basic split.

Theorem 7. Let $\phi(x) = (1-ex)^{\frac{6}{5}}$. Uniformly in θ as $x \to \frac{1}{e}$

$$\left|G(xe^{i\theta})\right| \sim \frac{o(G(x))}{\sqrt{b(x)}} \qquad \forall\, \theta : 0 < \phi(x) \leq |\theta| < \pi$$

i.e., for $\phi(x) \leq |\theta| < \pi$, $\forall\, \epsilon > 0$, $\exists\, \delta(\epsilon)$ such that if $|x - \frac{1}{e}| < \delta$,

$$\left|\frac{G^2\left(xe^{i\theta}\right)b(x)}{G^2(x)}\right| = \left|\frac{\sum_{k=1}^{\infty}\frac{k^{k+2}}{k!}x^k}{\exp\left(4\sum_{k=1}^{\infty}\frac{k^k}{k!}x^k\sin^2\left(\frac{k\theta}{2}\right)\right)}\right| < \epsilon$$

Proof We first upper bound $b(x)$ using Lemma 3.
We bound the denominator separately in the regions $(1-ex)^{\frac{6}{5}} \leq |\theta| \leq (1-ex)^{\frac{1}{8}}$ and $(1-ex)^{\frac{1}{8}} \leq |\theta| \leq \pi$. The bound for the second region will apply uniformly in any range lower bounded by $(1-ex)^{\alpha}$ with $\alpha < \frac{1}{4}$, in particular, we choose $\frac{1}{8}$.

We first consider the second region. Let $\phi'(x) = (1-ex)^{\frac{1}{8}}$. In the sum

$$\sum_{k=1}^{\infty}\frac{k^k}{k!}x^k\sin^2\left(\frac{k\theta}{2}\right),$$

reject all terms for which $\left|\frac{k\theta}{2}\right|$ is less than $\frac{1}{4}\phi'(x)$ or between $\pi \pm \frac{1}{4}\phi'(x)$. The sequence $\left|\frac{1}{2}\theta\right|$, $|\theta| \ldots \left|\frac{k}{2}\theta\right| \ldots$ will never have 2 consecutive terms $< \frac{1}{4}\phi'(x)$ or between $\pi \pm \frac{1}{4}\phi'(x)$ because $\phi'(x) \leq |\theta| \leq \pi$. Consequently, for any M consecutive terms, after this rejection process, we will have at least $\lfloor \frac{M}{2} \rfloor$ terms remaining. Lower bounding all remaining $\sin^2\left(\frac{k\theta}{2}\right)$ by

$\sin^2\left(\frac{\phi'(x)}{4}\right)$ allows us to factor the $\sin^2(\frac{\phi'(x)}{4})$ term out of the summation. Call the sum of the remaining terms *residual* summation. The terms $\frac{k^k}{k!}x^k$ decrease monotonically with k for $x \leq \frac{1}{e}$. So the lower bound for any residual summation is, using Lemma 8,

$$\sin^2\left(\frac{\phi'}{4}\right) \sum_{\substack{k=2 \\ k \text{ even}}}^{\infty} \frac{k^k}{k!}x^k \geq C\frac{\sin^2\left(\frac{\phi'(x)}{4}\right)}{\sqrt{1-ex}}$$

Define $v = \frac{1}{\sqrt[4]{1-ex}}$. Combining all that has been proved so far

$$\left|\frac{\sum_{k=1}^{\infty}\frac{k^{k+2}}{k!}x^k}{\exp\left(4\sum_{k=1}^{\infty}\frac{k^k}{k!}x^k\sin^2\left(\frac{k\theta}{2}\right)\right)}\right| \leq \frac{Cv^{10}}{e^{cv}}$$

which can be made smaller than any $\epsilon > 0$, for all $|\theta| \geq (1-ex)^{\frac{1}{8}}$ by choosing $|x-\frac{1}{e}| < \delta_1(\epsilon)$. Note that the $\frac{\sin^2(\phi'(x))}{\sqrt[4]{1-ex}} \to \infty$ for $\beta < \frac{1}{8}$, and equals 1 for $\beta = \frac{1}{8}$.

To tackle the remaining region, *i.e.,* $(1-ex)^{\frac{6}{5}} \leq |\theta| \leq (1-ex)^{\frac{1}{8}}$, we use the following inequality for $\frac{k\theta}{2} \leq \frac{\pi}{2}$,

$$\left|\sin\left(\frac{k\theta}{2}\right)\right| \geq \frac{2}{\pi}\frac{k\theta}{2}.$$

In this region, we will have $\frac{\pi}{\theta}$ terms for which the inequality holds with both sides being positive.

We write $\theta = (1-ex)^{\alpha}$. Therefore $\frac{1}{8} \leq \alpha \leq \frac{6}{5}$. Squaring and substituting the above inequality into the left side of the Theorem,

$$\left|\frac{\sum_{k=1}^{\infty}\frac{k^{k+2}}{k!}x^k}{\exp\left(4\sum_{k=1}^{\infty}\frac{k^k}{k!}x^k\sin^2\left(\frac{k\theta}{2}\right)\right)}\right| \leq \left|\frac{\sum_{k=1}^{\infty}\frac{k^{k+2}}{k!}x^k}{\exp\left(\frac{16}{\pi^2}\sum_{k=1}^{\frac{\pi}{\theta}}\frac{k^k}{k!}x^k\frac{k^2\theta^2}{4}\right)}\right|.$$

We lower bound $(1-ex)^{2\alpha}\sum_{k=1}^{\frac{\pi}{(1-ex)^{\alpha}}}\frac{k^{k+2}x^k}{k!}$ using Lemma 9,

$$(1-ex)^{2\alpha}\sum_{k=1}^{\frac{\pi}{(1-ex)^{\alpha}}}\frac{k^{k+2}x^k}{k!} \geq \begin{cases} \frac{C}{(1-ex)^{\frac{\alpha}{2}}} & \frac{1}{8} \leq \alpha \leq 1 \\ \frac{C}{(1-ex)^{\frac{5}{2}-2\alpha}} & 1 \leq \alpha \leq \frac{6}{5}. \end{cases}$$

Define $v = \frac{1}{\sqrt[16]{1-ex}}$. We conclude

$$\left|\frac{\sum_{k=1}^{\infty}\frac{k^{k+2}}{k!}x^k}{\exp\left(4\sum_{k=1}^{\infty}\frac{k^k}{k!}x^k\sin^2\left(\frac{k\theta}{2}\right)\right)}\right| \leq \frac{Cv^{40}}{e^{cv}}$$

which, for $\phi(x) \leq |\theta| \leq \sqrt[8]{1-ex}$, can be made smaller than $\epsilon > 0$, by taking $|x-\frac{1}{e}| \leq \delta_2(\epsilon)$. Picking $\delta = \min(\delta_1, \delta_2) = \delta_2$ concludes the proof for all $(1-ex)^{\frac{6}{5}} \leq \phi(x) \leq \pi$. □

We prove Lemma 8 and Lemma 9 used in Theorem 7.

Lemma 8. For $\frac{5}{6e} < x < \frac{1}{e}$,

$$\sum_{\substack{k=2 \\ k \text{ even}}}^{\infty} \frac{k^k}{k!} x^k \geq \frac{C}{\sqrt{1-ex}}.$$

Proof From Feller's bounds (11),

$$\sum_{\substack{k=2 \\ k \text{ even}}}^{\infty} \frac{k^k}{k!} x^k \geq \frac{e^{-\frac{1}{12}}}{\sqrt{2\pi}} \sum_{\substack{k=2 \\ k \text{ even}}}^{\infty} \frac{1}{\sqrt{k}} (ex)^k.$$

To lower bound the sum on the right observe that,

$$\begin{aligned}
\left(\sum_{\substack{k=2 \\ k \text{ even}}}^{\infty} \frac{(ex)^k}{\sqrt{k}} \right)^2 &= \sum_{\substack{k=4 \\ k \text{ even}}}^{\infty} (ex)^k \sum_{\substack{l=2 \\ l \text{ even}}}^{k-2} \frac{1}{\sqrt{l(k-l)}} \\
&\geq \sum_{\substack{k=4 \\ k \text{ even}}}^{\infty} (ex)^k \frac{k}{4} \cdot \frac{2}{k} \\
&= \frac{(ex)^4}{2(1-ex)(1+ex)} \\
&\geq \frac{C}{1-ex}.
\end{aligned}$$

By observing that $\frac{5}{6} < ex < 1$ we incorporate ex and $1+ex$ into a constant. Taking the positive square root proves the lemma. □

Lemma 9. For $x > \frac{5}{6e}$,

$$(1-ex)^{2\alpha} \sum_{k=1}^{\frac{2}{(1-ex)^\alpha}} \frac{k^{k+2} x^k}{k!} \geq \begin{cases} \frac{C}{(1-ex)^{\frac{\alpha}{2}}} & \alpha \leq 1 \\ \frac{C}{(1-ex)^{\frac{5}{2}-2\alpha}} & \alpha \geq 1. \end{cases}$$

Proof For any m, from Feller Bounds (11),

$$\sum_{k=1}^{m} \frac{k^{k+2} x^k}{k!} \geq C_1 \sum_{k=1}^{m} \frac{k^2 x^k e^k}{\sqrt{k}}.$$

We first show that if $k < \frac{3ex}{2(1-ex)}$, the k'th term is less than the $k+1'$th term in the above summation.

To see that observe that the ratio of the $k+1'$th to the kth term is

$$\left(1 + \frac{1}{k}\right)^{3/2} xe$$

and that

$$\left(1 + \frac{1}{k}\right)^{3/2} xe \geq \left(1 + \frac{3}{2k}\right) xe \geq 1.$$

where the second inequality holds if $k < \frac{3ex}{2(1-ex)}$. Since $ex > \frac{5}{6}$, the terms before $k = \frac{5}{4(1-ex)}$ in the summation are nondecreasing.

For $\alpha \leq 1$, observe that

$$\frac{1}{(1-ex)^\alpha} \leq \frac{1}{(1-ex)},$$

so that

$$\begin{aligned}
\sum_{k=1}^{\frac{2}{(1-ex)^\alpha}} k^{3/2} x^k e^k &\geq \sum_{\frac{1}{(1-ex)^\alpha}}^{\frac{5}{4(1-ex)^\alpha}} k^{3/2} x^k e^k \\
&\overset{(a)}{\geq} \frac{1}{(1-ex)^{\frac{3}{2}\alpha}} \cdot (xe)^{\frac{1}{(1-ex)^\alpha}} \cdot \frac{1}{4(1-ex)^\alpha} \\
&\overset{(b)}{\geq} \frac{1}{4(1-ex)^{\frac{5}{2}\alpha}} \cdot \frac{1}{4},
\end{aligned}$$

where (a) follows by replacing all terms of the summation with the first term in the summation and (b) because for all $\frac{1}{2} \leq y < 1$,

$$y^{\frac{1}{(1-y)^\alpha}} \geq y^{\frac{1}{1-y}} \geq \left(\frac{1}{2}\right)^2.$$

We complete the proof for $\alpha > 1$ by using the lemma for $\alpha = 1$, which we just proved. For $\alpha > 1$, observe that

$$\frac{1}{(1-ex)^\alpha} > \frac{1}{(1-ex)}.$$

Using the inequality,

$$(1-ex)^2 \sum_{k=1}^{\frac{1}{(1-ex)}} \frac{k^{k+2} x^k}{k!} \geq \frac{C}{(1-ex)^{\frac{1}{2}}},$$

observe that

$$\begin{aligned}
(1-ex)^{2\alpha} \sum_{k=1}^{\frac{2}{(1-ex)^\alpha}} \frac{k^{k+2} x^k}{k!} &\geq (1-ex)^{2\alpha} \sum_{k=1}^{\frac{2}{(1-ex)}} \frac{k^{k+2} x^k}{k!} \\
&\geq \frac{C}{(1-ex)^{\frac{5}{2}-2\alpha}}.
\end{aligned}$$

□

9 Evaluation of coefficients

Using Hayman's analysis, we evaluate the lower bound on $\hat{R}(\mathcal{I}_\Psi^n)$, namely $\frac{n!}{n^n}$ times the n'th coefficient of the expansion of $G(z)$.

Theorem 10.

$$\hat{R}^-(\mathcal{I}_\Psi^n) = \left(\frac{3}{2}\log e\right) n^{\frac{1}{3}} - \frac{1}{3}\log n - \frac{2}{3}\log e - \frac{1}{2}\log 3 + o(1).$$

Proof From (6), (7) and (8), we have that

$$\sum_{n=0}^{\infty} 2^{\hat{R}^-(\mathcal{I}_\Psi^n)} \frac{n^n}{n!} z^n = G(z) = \exp\left(\sum_{k=1}^{\infty} \frac{k^k z^k}{k!}\right).$$

From the observations following (10) and Theorems 6 and 7, we conclude that $G(z)$ satisfies the conditions of Theorem 1.

To use (9), we need to evaluate the function $a(z)$ shown in (10) to be

$$\sum_{k=1}^{\infty} \frac{k^{k+1} z^k}{k!}.$$

We do so using the related "tree function" [32]

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

which satisfies [32] the equation

$$T(z) = z e^{T(z)}. \tag{13}$$

Therefore,

$$\sum_{k=1}^{\infty} \frac{k^k z^k}{k!} = \frac{T(z)}{1 - T(z)}. \tag{14}$$

By differentiating Equations (13) and (14) and using the absolute convergence of the series, we obtain

$$a(z) = \frac{T(z)}{(1 - T(z))^3}, \text{ and } b(z) = \frac{2T(z)^2 + T(z)}{(1 - T(z))^5}.$$

At $z = \frac{1}{e}$, we have the following singular expansion [32],

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

Consequently, it can be verified that

$$a(z) = \frac{1}{(2(1 - ez))^{\frac{3}{2}}} + \mathcal{O}\left(\frac{1}{\sqrt{1 - ez}}\right),$$

and the solution to $a(r_n) = n$ is

$$r_n = \frac{1}{e}\left(1 - \frac{1}{2n^{\frac{2}{3}}}\right) + \mathcal{O}\left(\frac{1}{n^{\frac{4}{3}}}\right).$$

The n'th coefficient of the $G(z)$ therefore equals

$$\frac{G(r_n)}{r_n^n \sqrt{2\pi b(r_n)}}(1 + o(1)). \tag{15}$$

We evaluate the terms to be

$$\begin{aligned} G(r_n) &= \exp(n^{\frac{1}{3}} - \frac{2}{3} + \mathcal{O}(n^{-\frac{1}{3}})), \\ r_n^n &= \exp(-n - \frac{1}{2}n^{\frac{1}{3}} + \mathcal{O}(n^{-\frac{1}{3}}))(1 + \mathcal{O}(n^{-\frac{2}{3}})), \text{ and} \\ b(r_n) &= 3n^{\frac{5}{3}} + \mathcal{O}(n), \end{aligned}$$

and use them to evaluate $\hat{R}^-(\mathcal{I}_\Psi^n)$,

$$\hat{R}^-(\mathcal{I}_\Psi^n) = \left(\frac{3}{2}\log e\right)n^{\frac{1}{3}} - \frac{1}{3}\log n - \frac{2}{3}\log e - \frac{1}{2}\log 3 + o(1). \qquad \square$$

We note that this is the highest accuracy of the asymptotic expansion allowed by the Hayman's theorem, limited by the form of the Equation (15).

Corollary 11.

$$\hat{R}(\mathcal{I}_\Psi^n) \geq \left(\frac{3}{2}\log e\right)n^{\frac{1}{3}} - \frac{1}{3}\log n - \frac{2}{3}\log e - \frac{1}{2}\log 3 + o(1). \qquad \square$$

Acknowledgements

We thank Wojciech Szpankowski for sharing his intuition behind saddle point approximations.

References

[1] Y.M. Shtarkov. Universal sequential coding of single messages. *Problems of Information Transmission*, 23(3):3—17, 1987.

[2] R.E. Krichevsky and V.K. Trofimov. The preformance of universal coding. *IEEE Transactions on Information Theory*, 27(2):199—207, March 1981.

[3] T.M. Cover. Universal portfolios. *Mathematical Finance*, 1(1):1—29, January 1991.

[4] Y.M. Shtarkov, T.J. Tjalkens, and F.M.J. Willems. Multialphabet universal coding of memoryless sources. *Problems of Information Transmission*, 31(2):114—127, 1995.

[5] T.M. Cover and E. Ordentlich. Universal portfolios with side information. *IEEE Transactions on Information Theory*, 42(2):348—363, March 1996.

[6] J. Rissanen. Fisher information and stochastic complexity. *IEEE Transactions on Information Theory*, 42(1):40—47, January 1996.

[7] W. Szpankowski. On asymptotics of certain recurrences arising in universal coding. *Problems of Information Transmission*, 34(2):142—146, 1998.

[8] Q. Xie and A.R. Barron. Asymptotic minimax regret for data compression, gambling and prediction. *IEEE Transactions on Information Theory*, 46(2):431—445, March 2000.

[9] M. Drmota and W. Szpankowski. The precise minimax redundancy. In *Proceedings of IEEE Symposium on Information Theory*, 2002.

[10] J.C. Kieffer. A unified approach to weak universal source coding. *IEEE Transactions on Information Theory*, 24(6):674—682, November 1978.

[11] P. Elias. Universal codeword sets and representations of integers. *IEEE Transactions on Information Theory*, 21(2):194—203, March 1975.

[12] L. Györfi, I. Pali, and E.C. Van der Meulen. On universal noiseless source coding for infinite source alphabets. *European Transactions on Telecommunications and Related Technologies*, 4:125—132, 1993.

[13] D.P. Foster, R.A. Stine, and A.J. Wyner. Universal codes for finite sequences of integers drawn from a monotone distribution. *IEEE Transactions on Information Theory*, 48(6):1713—1720, June 2002.

[14] T. Uyematsu and F. Kanaya. Asymptotic optimality of two variations of Lempel-Ziv codes for sources with countably infinite alphabet. In *Proceedings of IEEE Symposium on Information Theory*, 2002.

[15] D. He and E Yang. On the universality of grammar-based codes for sources with countably infinite alphabets. In *Proceedings of IEEE Symposium on Information Theory*, 2003.

[16] J. Åberg, Y.M. Shtarkov, and B.J.M. Smeets. Multialphabet coding with separate alphabet description. In *Proceedings of Compression and Complexity of Sequences*, 1997.

[17] N. Jevtić, A. Orlitsky, and N.P. Santhanam. Universal compression of unknown alphabets. In *Proceedings of IEEE Symposium on Information Theory*, 2002.

[18] K.W. Church and W.A. Gale. Probability scoring for spelling correction. *Statistics and Computing*, 1:93—103, 1991.

[19] W.A. Gale, K.W. Church, and D. Yarowsky. A method for disambiguating word senses. *Computers and Humanities*, 26:415—419, 1993.

[20] S.F. Chen and J. Goodman. An empirical study of smoothing techniques for language modeling. In *Proceedings of the Thirty-Fourth Annual Meeting of the Association for Computational Linguistics*, pages 310—318, San Francisco, 1996. Morgan Kaufmann Publishers.

[21] K. Yamanishi. A decision-theoretic extension of stochastic complexity and its application to learning. *IEEE Transactions on Information Theory*, 44(4):1424—1439, July 1998.

[22] V.G. Vovk. A game of prediction with expert advice. *Journal of Computer and System Sciences*, 56(2):153—173, 1998.

[23] N. Cesa-Bianchi and G. Lugosi. Minimax regret under log loss for general classes of experts. In *Proceedings of the Twelfth Annual Conference on Computational Learning Theory*, pages 12—18, 1999.

[24] F. Song and W.B. Croft. A general language model for information retrieval (poster abstract). In *Research and Development in Information Retrieval*, pages 279—280, 1999.

[25] A. Orlitsky, N.P. Santhanam, and J. Zhang. Universal compression of memoryless sources over unknown alphabets. *IEEE Transactions on Information Theory*, 50(7):1469—1481, July 2004.

[26] A. Orlitsky and N.P. Santhanam. Speaking of infinity. *IEEE Transactions on Information Theory*, 50(10):2215—2230, October 2004.

[27] G. Shamir. Universal lossless compression with unknown alphabets—the average case. Submitted for publication, IEEE Transactions on Information Theory, 2003.

[28] A. Orlitsky, N.P. Santhanam, and J. Zhang. Always Good Turing: Asymptotically optimal probability estimation. In *Proceedings of the 44th Annual Symposium on Foundations of Computer Science*, October 2003.

[29] A. Orlitsky, N.P. Santhanam, and J. Zhang. Always Good Turing: Asymptotically optimal probability estimation. *Science*, 302(5644):427—431, October 17 2003.

[30] A. Orlitsky and K. Viswanathan. One-way communication and error-correcting codes. *IEEE Transactions on Information Theory*, 49(7):1781—1788, July 2003.

[31] W.K. Hayman. A generalization of Stirling's formula. *Journal für die reine und angewandte Mathematik*, 196:67—95, 1956.

[32] W. Szpankowski. *Average Case Analysis of Algorithms on Sequences*. Wiley, 2001.

[33] P. Flajolet and R. Sedgewick. Average case analysis of algorithms: Saddle point asymptotics. Technical Report 2376, INRIA, 1994.

[34] E.W. Weisstein. "Ratio Test." From Mathworld—A Wolfram Web Resource. `http://mathworld.wolfram.com/RatioTest.html`.

[35] W. Feller. *An introduction to probability theory*. Wiley, 1968.

[36] E.W. Weisstein. "Root Test." From Mathworld—A Wolfram Web Resource. http://mathworld.wolfram.com/RootTest.html.