ENTROPY COMPUTATIONS VIA ANALYTIC DEPOISSONIZATION

February 4, 1998

Philippe Jacquet*
INRIA
Rocquencourt
78153 Le Chesnay Cedex
France
jacquet@blagny.inria.fr

Wojciech Szpankowski†
Department of Computer Science
Purdue University
W. Lafayette, IN 47907
U.S.A.
spa@cs.purdue.edu

Abstract

We investigate the basic question of information theory, namely, evaluation of Shannon entropy, and a more general Rényi entropy, for some discrete distributions (e.g., binomial, negative binomial, etc.). We aim at establishing analytic methods (i.e., those in which complex analysis plays a pivotal role) for such computations which often yield estimates of unparalleled precision. The main analytic tool used here is that of analytic poissonization and depoissonization. We illustrate our approach on the entropy evaluation of the binomial distribution, that is, we prove that for $Binomial(n,p)$ distribution the entropy h_n becomes $h_n \asymp \frac{1}{2}\ln n + \frac{1}{2} + \ln\sqrt{2\pi p(1-p)} + \sum_{k\geq 1} a_k n^{-k}$ where a_k are explicitly computable constants. Moreover, we shall argue that analytic methods (e.g., complex asymptotics such as Rice's method and singularity analysis, Mellin transforms, poissonization and depoissonization) can offer new tools for information theory, especially for studying second-order asymptotics (e.g., for code redundancy). In fact, in recent years there has been a resurgence of interest and a few successful applications of analytic methods to a variety of problems of information theory, therefore, we we coin the term *analytic information theory* for such investigations.

Key Words: Shannon entropy, Rényi entropy, discrete distributions, analytic poissonization and depoissonization, formal power series, binomial sums.

*This work was partly supported by the ESPRIT Basic Research Action No. 7141 (ALCOM II).
†The work of this author was supported by NSF Grants NCR-9206315 and NCR-9415491, and NATO Collaborative Grant CRG.950060.

1 Introduction

Information theory celebrates its 50-th birthday. Although it is a mature area of research by any standard, new challenges arise due to new applications and new theoretical developments. In the *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 by replacing first-order analysis by full asymptotic expansions and more accurate analysis. To accomplish this goal we explore problems of information theory by analytic methods, that is, those in which complex analysis plays a pivotal role. As argued by Andrew Odlyzko [26]: "*Analytic methods are extremely powerful and when they apply, they often yield estimates of unparalleled precision.*" Certainly, analytic methods were used before in information theory, notably in coding and signal processing. Our goal, however, is to bring to information theory new tools of the analysis of algorithms (initiated by D.E. Knuth [18] and further developed by Flajolet and his collaborators [9, 29, 33]) such as: generating functions, differential–functional equations, complex asymptotics, Mellin transforms, poissonization and depoissonization, etc. In recent years there has been a resurgence of interest and a few successful applications of analytic methods to a variety of problems on words (strings, sequences) that are at the core of information theory. Among others we mention here: analysis of Lempel-Ziv data compression schemes, (cf. [13, 14, 16, 21, 22, 23]), channel and source coding (cf. [19, 31, 32]); conflict resolution algorithms for multiaccess communications (cf. [6, 7, 25, 28, 30]); digital trees and their applications (cf. [12, 14, 15, 24, 29]); self-similar traffic with long-range dependence in computer networks (cf. [11]), and so forth.

Analytic number theory was born in 1896 when independently Hadamard and de la Vallée-Poussin proved the *prime number theory* through a sophisticated application of complex analysis methods. *Analytic* average–case analysis of algorithms started more than thirty years ago by Knuth who used systematically generating functions to predict the performance of computer algorithms. Flajolet, Odlyzko and Sedgewick launched *analytic* combinatorics by applying Mellin transform and the singularity analysis to combinatorial problems. In all of the above cases, theory of *analytic* functions was used to solve problems of diversified areas of science and engineering. In view of this, it is natural to coin the term *analytic information theory* to describe problems of information theory that are solved by analytic methods, that is, those in which complex analysis plays a pivotal role.

In this paper, we address a basic question of information theory, namely, a precise evaluation of Shannon entropy as well as a more general Rényi entropy [27] defined as

$$h_n(\omega) = \frac{1}{1-\omega} \ln\left(\sum_{k\geq 0} p_{n,k}^{\omega}\right), \tag{1}$$

where $p_{n,k}$ is a sequence of discrete distributions taking value k for the nth distribution. Of course, Shannon entropy h_n is the limiting case of Rényi entropy, namely: $h_n = \lim_{\omega\to 1} h_n(\omega)$. We provide here an analytic method to evaluate Shannon and Rényi entropy for a class of discrete distributions, however, we illustrate most of our analysis on the binomial distribution $p_{n,k} = \binom{n}{k} p^k (1-p)^{n-k}$. The following quite imprecise bound on the binomial distribution's entropy (for $p = 1/2$) $\frac{1}{2}\ln(\pi n/2) \leq h_n \leq \frac{1}{2}\ln(\pi e n/2)$ is often used in the literature (cf. [3]), but it is known (cf. [10]) that the upper bound is asymptotically accurate. We concluded that a more precise asymptotic approximation of the entropy for discrete distributions is quite desirable (cf. [10]). Thus, we present here various asymptotic approximations for Shannon and Rényi entropy. For example, for a class of discrete distributions $p_{n,k}$ that can be approximated by the normal distribution with variance $n\sigma^2 > 0$ when $n \to \infty$, we shall show that (cf. Theorem 1):

$$h_n(\omega) = \frac{\ln\left(\omega^{-\frac{1}{2}} (2\pi n\sigma^2)^{-\frac{\omega-1}{2}}\right)}{1-\omega} + o(1).$$

The above formula is only a *first-order* asymptotic approximation, and we have a more ambitious goal of deriving a full asymptotic expansion of $h_n(\omega)$ for a class of discrete distributions. For example, for the binomial distribution we shall prove that (cf. Theorem 2)

$$h_n \asymp \frac{1}{2}\ln n + \frac{1}{2} + \ln\sqrt{2\pi p(1-p)} + \sum_{k\geq 1} \frac{a_k}{n^k}$$

where a_k are explicitly computable constants (see also [17]).

This paper aims at presenting analytic tools to compute a full asymptotic expansion for entropy-like sums that often arise in information theory. Throughout, we shall use repeatedly analytic poissonization and depoissonization. Poissonization is an applied probability tool that replaces a deterministic input, say n, by a Poisson process N with mean $z = n$ and assuming z to be a complex variable. Here, we use *analytic poissonization* that maps a sequence g_n into the Poisson transform $\widetilde{G}(z) = \sum_{n\geq 0} g_n \frac{z^n}{n!} e^{-z}$ which is next *analyticly depoissonize* to recover asymptotics of the original sequence g_n. This turns out to be a good approach to some problems (of combinatorics, analysis of algorithms, applied probability, etc.) in which poissonization removes some "tricky statistical dependence" (cf. [2]). Recently, analytic poissonization and depoissonization were used in the analysis of algorithms (e.g., [18, 24, 15, 28]), performance evaluation of Lempel-Ziv schemes (cf. [21, 22, 23, 14]), and we use it here to estimate the entropy. In fact, we extend the poissonization/depoissonization method in two directions. First of all, in Theorem 4 we show how to estimate asymptotically the Poisson transform without computing it explicitly (the situation that occurs quite often in practice). Secondly, in Lemma 1 we prove that for sequences of a polynomial growth, the depoissonization holds automatically (so we do not need to verify sometimes cumbersome depoissonization conditions).

We believe the main contribution of this paper is of a methodological nature, not in terms of the obtained results. Therefore, our presentation is centered around the methodology of computations. In the next section, we propose a first order approximation for Rényi entropies of some discrete distributions using a simple asymptotic tool. Then, we use analytic poissonization/depoissonization method in order to obtain a full asymptotic expansion for Shannon entropy of the binomial distribution. In concluding remarks we show how to apply our tool to entropy evaluation of some other distributions (e.g., the negative binomial distribution) and entropy-like sums (e.g., the redundancy of predicting the $(n+1)$st symbol in a random sequence).

2 First Order Asymptotics

Let us start with a first-order approximation of Rényi entropy for a class of sequences of discrete distribution $p_{n,k}$ satisfying the following two assumptions:

(A1) Let $n\alpha$ and $n\sigma^2$ be the mean and the variance of the distribution $p_{n,k}$, where α and $\sigma^2 > 0$ are some constants. We postulate that for $k = n\alpha + x\sqrt{n}$ as $n \to \infty$

$$p_{n,k} = \frac{1}{\sqrt{2\pi n\sigma^2}} \exp\left(-\frac{k - n\alpha}{2n\sigma^2}\right)\left(1 + O\left(\frac{x^3}{\sqrt{n}}\right)\right) , \tag{2}$$

where $x = o(n^{\frac{1}{6}})$, that is, $p_{n,k}$ converges with rate $O(x^3 n^{-1/2})$ to the normal distribution with mean $n\alpha$ and variance $n\sigma^2$.

(A2) For any $\varepsilon > 0$ and complex ω such that $\Re(\omega) > 0$ the following holds:

$$\sum_{|k-n\alpha|>n^{\frac{1}{2}+\varepsilon}} p_{n,k}^{\omega} = o(n^{-(\omega-1)/2}) \tag{3}$$

as $n \to \infty$.

Under these assumptions, we can evaluate Rényi entropy of order ω. For any $0 < \varepsilon < \frac{1}{6}$,

$$
\begin{aligned}
\exp\left((1-\omega)h_n(\omega)\right) &= \sum_{|k-n\alpha|\le n^{\frac{1}{2}+\varepsilon}} p_{n,k}^{\omega} + o(n^{-(\omega-1)/2}) \\
&= \left(1+O\left(n^{-\frac{1}{2}+\varepsilon}\right)\right)\frac{1}{(2\pi n\sigma^2)^{\omega/2}}\int_{-n^{\frac{1}{2}+\varepsilon}}^{n^{\frac{1}{2}+\varepsilon}} e^{-\frac{y^2\omega}{2n\sigma^2}}dy + o(n^{-(\omega-1)/2}) \quad (4)\\
&= \left(1+O\left(n^{-\frac{1}{2}+\varepsilon}\right)\right)\frac{1}{(2\pi n\sigma^2)^{\omega/2}}\int_{-\infty}^{\infty} e^{-\frac{y^2\omega}{2n\sigma^2}}dy + o(n^{-(\omega-1)/2}) \quad (5)\\
&= \frac{1}{\sqrt{\omega}(2\pi n\sigma^2)^{\frac{\omega-1}{2}}}\left(1+O\left(n^{-\frac{1}{2}+\varepsilon}\right)\right) + o(n^{-(\omega-1)/2})
\end{aligned}
$$

where (4) follows directly from the Euler-Maclaurin formula (cf. [18]) and (5) is a consequence of the following well known approximation (cf. [26])

$$
\int_{\theta}^{\infty} x^k e^{-tx^2}dx = O\left(e^{-\frac{1}{2}t\theta^2}\right)
$$

for any k, where θ is a positive number. In order to recover asymptotically Shannon entropy h_n we either repeat the above derivations or apply Cauchy's formula

$$
h_n = \frac{1}{2\pi i}\oint\frac{h_n(\omega)}{1-\omega}d\omega
$$

to the above expression, where the integration is on any loop encircling $\omega = 1$.

In fact, the latter derivation shows the power of analytic methods and provides an interesting insights. Since $h_n(\omega)$ is an analytic function for $\omega \ne 1$, its asymptotic around $\omega = 1$ can be extended to the asymptotic for $\omega = 1$ leading to an asymptotic evaluation of Shannon entropy. Indeed, it follows directly from the Cauchy formula and the above estimate of $h_n(\omega)$ for $\omega \ne 1$, namely

$$
h_n = \frac{1}{2i\pi}\oint(\frac{\ln\left(\omega^{-\frac{1}{2}}(2\pi n\sigma^2)^{-\frac{\omega-1}{2}}\right)}{(1-\omega)^2}d\omega + \frac{1}{2i\pi}\oint\frac{o(1)}{1-\omega}d\omega \ .
$$

In summary, we just proved:

Theorem 1 *Under assumptions (A1) and (A2) the Rényi entropy of order ω attains the following first order asymptotic*

$$
h_n(\omega) = \frac{\ln\left(\omega^{-\frac{1}{2}}(2\pi n\sigma^2)^{-\frac{\omega-1}{2}}\right)}{1-\omega} + o(1) \quad (6)
$$

In particular, Shannon entropy becomes (as $\omega \to 1$)

$$
h_n = \frac{1}{2}\ln n + \frac{1}{2}\left(1+\ln(2\pi\sigma^2)\right) + o(1) \quad (7)
$$

as $n \to \infty$.

Let us illustrate the above result on two distributions, namely:

- *Binomial distribution* defined for $k = 0, 1, \ldots, n$

$$p_{n,k} = \binom{n}{k} p^k (1-p)^{n-k}$$

  with mean= np and variance= npq. Throughout the paper we write $q := 1 - p$. Observe that assumptions (A1)–(A2) hold since a random variable having a binomial distribution is a sum of n i.i.d. Bernoulli distributed random variables.

- *Negative binomial distribution* defined for for $n \geq 1$ and $k \geq n$ as

$$p_{n,k} = \binom{k}{n} p^{n+1} q^{k-n}$$

  with mean= n/p and variance= nq/p^2 (and, of course, $q = 1 - p$). This distribution also satisfies (A1)–(A2) since a random variables representing such a distribution is a sum of n i.i.d. geometrically distributed random variables.

Corollary 1 *The Rényi entropy of order ω for the binomial distribution and the negative binomial distribution becomes for any $\varepsilon > 0$, respectively,*

$$\begin{aligned}
h_n(\omega) &= \frac{\ln\left(\omega^{-\frac{1}{2}}(2\pi npq)^{-\frac{\omega-1}{2}}\right)}{1-\omega} + O(n^{-\omega/2+\varepsilon}) \ , \\
h_n(\omega) &= \frac{\ln\left(\omega^{-\frac{1}{2}}(2\pi nqp^{-2})^{-\frac{\omega-1}{2}}\right)}{1-\omega} + O(n^{-\omega/2+\varepsilon}) \ ,
\end{aligned}$$

while Shannon entropy becomes, respectively,

$$\begin{aligned}
h_n &= \frac{1}{2}\ln n + \frac{1}{2}\left(1 + \ln(2\pi pq)\right) + O\left(n^{-\frac{1}{2}+\varepsilon}\right) \ , \\
h_n &= \frac{1}{2}\ln n + \frac{1}{2}\left(1 + \ln(2\pi qp^{-2})\right) + O\left(n^{-\frac{1}{2}+\varepsilon}\right)
\end{aligned}$$

as $n \to \infty$.

3 Full Asymptotic Expansions

In the previous section, we derived a first order asymptotic expansion for the Rényi entropy of order ω as well as for Shannon entropy. We used a very simple asymptotic technique (cf. [26]): Indeed, after observing that all terms of the sum involved in the Rényi entropy are positive, they are concentrated around certain maximum value, and decay exponentially fast when away from this maximum value, we estimated the sum by the Gaussian integral around the maximum value. This gives us the leading term of the asymptotic expansion, but an extension to a full asymptotic expansion is quite troublesome by this technique, if possible at all.

In this section, we propose a new technique that will lead to a full asymptotic expansion of entropy-like sums (cf. Sections 4 and 5 for more detailed statements). This method is based on an application of our recent results from [15] concerning analytic depoissonization. In this paper, we present a full asymptotic expansion of Shannon entropy for the binomial distribution, that is, we analyze

$$h_n = -\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \ln\left(\binom{n}{k} p^k q^{n-k}\right)$$

for large n. We summarize our findings in the following theorem that is proved in Section 3.

Theorem 2 *Shannon entropy, h_n, of the binomial distribution possesses the following asymptotic expansion*

$$\begin{aligned} h_n &\asymp \frac{1}{2}\ln n + \frac{1}{2} + \frac{1}{2}\ln(2\pi pq) \\ &+ \sum_{m=1}^{\infty} \frac{1}{n^{2m-1}} \left(\frac{B_{2m}(p^{1-2m} + q^{1-2m} - 1)}{2m(2m-1)} + o_m \right) + \sum_{m=1}^{\infty} \frac{e_m}{n^{2m}} \end{aligned} \tag{8}$$

where B_k are the Bernoulli numbers (cf. (28)), $q = 1 - p$, and

$$\begin{aligned} o_m &= \sum_{i=1}^{\infty} \left[(-1)^{2m+i}(2m+i-2)! \left\{ p^{1-2m} \left(c_{i,2m+i}(p) + \frac{1}{2}(4m+2i-3)c_{i,2m+i-1}(p) \right) \right.\right. \\ &+ \left. q^{1-2m} \left(c_{i,2m+i}(q) + \frac{1}{2}(4m+2i-3)c_{i,2m+i-1}(q) \right) \right\} \\ &+ \left. \sum_{d=0}^{m-1} \frac{(-1)^{2d+i} B_{2(m-d)}(2m+i-2)!}{(2(m-d))!} \left(p^{1-2m} c_{i,2d+i}(p) + q^{1-2m} c_{i,2d+i}(q) \right) \right], \end{aligned}$$

$$\begin{aligned} e_m &= \sum_{i=1}^{\infty} \left[(-1)^{2m+i-1}(2m+i-2)! \left\{ p^{2m} \left(c_{i,2m+i+1}(p) + \frac{1}{2}(4m+2i-1)c_{i,2m+i}(p) \right) \right.\right. \\ &+ \left. q^{2m} \left(c_{i,2m+i+1}(q) + \frac{1}{2}(4m+2i-1)c_{i,2m+i}(q) \right) \right\} \\ &+ \left. \sum_{d=0}^{m-1} \frac{(-1)^{2d+i+1} B_{2(m-d)}(2m+i-1)!}{(2(m-d))!} \left(p^{2m} c_{i,2d+i}(p) + q^{1-2m} c_{i,2d+i}(q) \right) \right]. \end{aligned}$$

In the above, $c_{i,k}(\gamma)$ are Taylor's coefficients of the following function

$$\exp\left(x \ln\left(1+\lambda(e^y-1)\right)-\lambda xy\right)=\sum_{i=1}^{\infty}\sum_{k=2i}^{\infty} c_{i,k}(\lambda)x^i y^k, \tag{9}$$

where in the above λ is either equal to p or to q.

3.1 Starting the Derivations

Observe that the formula on h_n can be re-written as follows:

$$\begin{aligned} h_n &= -\ln(n!)-n(p\ln p+q\ln q) \\ &+ \sum_{k=0}^{n}\ln(k!)\binom{n}{k}p^k q^{n-k}+\sum_{k=0}^{n}\ln((n-k)!)\binom{n}{k}p^k q^{n-k} \ . \end{aligned} \tag{10}$$

We must evaluate the last two sums in order to obtain precise asymptotics. Let

$$g_n := \sum_{k=0}^{n}\ln(k!)\binom{n}{k}p^k q^{n-k}+\sum_{k=0}^{n}\ln((n-k)!)\binom{n}{k}p^k q^{n-k} \ , \tag{11}$$

and define the Poisson transform of g_n as follows (cf. [15])

$$\widetilde{G}(z)=\sum_{n=0}^{\infty} g_n\frac{z^n}{n!}e^{-z}$$

for z complex. We also use the following additional notation for the above Poisson transform $\mathcal{P}(g_n;z):=\widetilde{G}(z)$.

We observe that "usually" the Poisson transform has a simpler form than the original sequence (as in the case of Fourier or Laplace transforms). Indeed, after multiplying (11) by $z^n/n!$ and summing up we quickly arrive at the following simple formula

$$\widetilde{G}(z)=\widetilde{F}(zp)+\widetilde{F}(zq) \tag{12}$$

where

$$\widetilde{F}(z)=\sum_{n=0}^{\infty}\ln(n!)\frac{z^n}{n!}e^{-z} \ . \tag{13}$$

This can be compared with (10).

We now want to recover g_n from its Poisson transform (12). This is called depoissonization, and we denote this inverse operation as $\mathcal{P}^{-1}(\widetilde{G}(z);n):=g_n$. Thus, from the analysis presented so far we conclude that:

$$h_n=-\ln(n!)-n(p\ln p+q\ln q)+\mathcal{P}^{-1}\left(\widetilde{F}(zp)+\widetilde{F}(zq);n\right) \ . \tag{14}$$

Before we complete the evaluation of h_n, we discuss below some depoissonization results from [15] as well as their generalizations.

3.2 Going into Complex Analysis

Let us consider a general situation. Assume a sequence, say g_n, is given, and its Poisson transform is $\widetilde{G}(z)$. We want to know simple conditions iunder which $g_n \sim \widetilde{G}(n)$ holds for $n \to \infty$. This is likely to happen since $\widetilde{G}(n) = \mathbf{E}[g_N]$ where N is a Poisson random variable with mean n, and $\mathbf{E}[\cdot]$ denotes the expectation operator. Indeed, since N is concentrated around its mean, one may expect that $\mathbf{E}[g_N] \sim g_n$. We actually would like to know a *full asymptotic expansion* of g_n in terms of $\widetilde{G}(n)$ and its derivatives.

The above problem was recently addressed in our paper [15] and we cite below one result from it.

Theorem 3 (Jacquet and Szpankowski [15]) *Consider a linear cone* $\mathcal{S}_\theta = \{z : \ |\arg(z)| \leq \theta, \ \ \theta < \pi/2\}$. *Let the following two conditions hold for some numbers* A, B, $R > 0$ *and* $\alpha < 1$ *and* β*:*
(I) *For* $z \in \mathcal{S}_\theta$

$$|z| > R \quad \Rightarrow \quad |\widetilde{G}(z)| \leq B|z|^\beta \Psi(|z|) \ , \tag{15}$$

where $\Psi(x)$ *is a slowly varying function , that is, such that for fixed* t $\lim_{x\to\infty} \frac{\Psi(tx)}{\Psi(x)} = 1$ *(e.g.,* $\Psi(x) = \ln^d x$ *for some* $d > 0$*);*
(O) *For* z *outside the cone* $\mathcal{S}_\theta$*, that is, for* $z \notin \mathcal{S}_\theta$

$$|z| > R \quad \Rightarrow \quad |\widetilde{G}(z)e^z| \leq A\exp(\alpha|z|) \tag{16}$$

Then, for every nonnegative integer m

$$\begin{aligned} g_n &= \sum_{i=0}^{m}\sum_{j=0}^{i+m} b_{ij} n^i \widetilde{G}^{\langle j\rangle}(n) + O(n^{\beta-m-1}\Psi(n)) \\ &= \widetilde{G}(n) + \sum_{k=1}^{m}\sum_{i=1}^{k} b_{i,k+i} n^i \widetilde{G}^{\langle k+i\rangle}(n) + O(n^{\beta-m-1}\Psi(n)) \end{aligned} \tag{17}$$

where $\widetilde{G}^{\langle j\rangle}(n)$ *is the* j*th derivative of* $\widetilde{G}(z)$ *at* $z = n$*, and* b_{ij} *are the coefficients of* $\exp(x\ln(1+y) - xy)$ *at* $x^i y^j$*, that is:*

$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} b_{ij} x^i y^j = \exp(x\ln(1+y) - xy) \ . \tag{18}$$

such that $b_{ij} = 0$ *for* $j < 2i$. *In fact,* $b_{ij} = s_2(j,i)/j!$ *where* $s_2(n,k)$ *are the associated Stirling numbers of the first kind* (cf. Comtet [4], pp. 295).

Before we proceed, let us explain how one can derive formally (18) (see [15] for a rigorous derivation of the error term). Observe that Taylor's expansion of $\widetilde{G}(z)$ around $z = n$ becomes

$$\widetilde{G}(z) = \sum_{k=0}^{\infty} \widetilde{G}^{\langle k \rangle}(n) \frac{(z-n)^k}{k!} \ .$$

Let us now derive a formula on the coefficients b_{ij} without worrying about the series convergence and the error term. For this we can apply the theory of formal power series viewed as purely algebraic objects [34] (as opposed to their analytic theory needed to derive the error term and the radius of convergence). Since we want to identify the coefficient at n^i and $\widetilde{G}^{\langle k \rangle}(n)$ we formally define $x^i = n^i$ and $y^k := \widetilde{G}^{\langle k \rangle}(n)$, rearrange the series, and extract the coefficient at $x^i y^k$. We proceed as above to obtain

$$\widetilde{G}(z) = \sum_{k \geq 0} \frac{(z-n)^n}{k!} y^k = \exp(y(z-n)) \qquad \text{where} \qquad y^k = \widetilde{G}^{\langle k \rangle}(n) \ .$$

Thus,

$$\begin{aligned} g_n &= n![z^n](e^z \widetilde{G}(z)) = n![z^n](e^{-yn} e^{(1+y)z}) = (1+y)^n e^{-ny} \\ &= \exp(n \ln(1+y) - ny) \qquad \text{where} \qquad y^k = \widetilde{G}^{\langle k \rangle}(n) \end{aligned}$$

which is exactly (17) written in a more compact form. We shall use this formal series approach when seeking coefficients of power series.

We cannot apply directly Theorem 3 since in our case $\widetilde{G}(z) = \widetilde{F}(zp) + \widetilde{F}(zq)$ and we do *not* have an explicit formula for $\widetilde{F}(z)$. Observe, however, that $\widetilde{F}(z) = \mathcal{P}(\ln(n!); z)$ and we know asymptotic expansion for $f_n = \ln(n!)$. Even more, we can analytically continued f_n to an analytic function $f(z) = \ln \Gamma(z+1)$ such that $f_n = f(n)$. Observe also that to apply (17) we need derivatives of $\widetilde{F}(z)$ at $z = np$ and $z = nq$.

Motivated by the above considerations, let us consider a general setting. We derive below an expansion of the Poisson transform $\widetilde{G}(z) = \mathcal{P}(g_n; z)$ as a function of the derivatives of $g(z)$ where $g(z)$ is an analytic continuation of $g_n := g(n)$. In other words, we shall try to find such coefficients, say a_{ij}, that

$$\widetilde{G}(z) = \sum_{i,j \geq 0} a_{ij} z^i g^{\langle j \rangle}(z),$$

and then identify the error term.

Our plan now is as follows: In the theorem below we identify the coefficients a_{ij} as Taylor's expansion of a certain explicitly given function. Then, computing $\widetilde{G}^{\langle k \rangle}(n)$ is a matter of a simple algebra. Hence, by (17) of Theorem 3 we can recover asymptotically g_n

provided we can verify conditions (I) and (O). But, usually condition (I) will follow directly from the form of $g(z)$ (which is known explicitly in our case), and one only needs to verify (O). We deal with condition (O) in Lemma 1 below.

We start with finding an asymptotic expression on $\widetilde{G}(z)$ knowing only asymptotic expansion for $g(z)$. This is presented in the lemma below.

Theorem 4 *Let $g(z)$ be an analytic continuation of a sequence g_n whose Poisson transform is $\widetilde{G}(z)$, and such that $g(z) = O(z^\beta)$ in a linear cone, where β is a constant. Then, for every nonnegative integer m and complex $w = \lambda z$ for some constant λ (more generally: $w = \Theta(z)$), as $z \to \infty$*

$$\begin{aligned} \widetilde{G}(z) &= \sum_{i=0}^{m} \sum_{j=0}^{i+m} a_{ij}(\lambda) z^i g^{\langle j \rangle}(w) + O(z^{\beta-m-1}) \\ &= g(w) + \sum_{k=1}^{m} \sum_{i=1}^{k} a_{i,k+i}(\lambda) z^i g^{\langle k+i \rangle}(w) + O(z^{\beta-m-1}) \end{aligned} \tag{19}$$

where $a_{ij}(\lambda)$ are coefficients of $\exp(x(e^y - 1) - wy)$ *at $x^i y^j$. More precisely: for $\lambda = 1$*

$$\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} a_{ij} x^i y^j = \exp(x(e^y - 1) - xy) \ . \tag{20}$$

and $a_{ij} = 0$ for $j < 2i$. In fact, $a_{ij} = S_2(j,i)/j!$ where $S_2(n,k)$ are the 2-associated Stirling numbers of the second kind (cf. Comtet [4], pp. 222).

Remark. The above can be formally re-written as

$$\widetilde{G}(z) = \exp\left(z(e^y - 1) - wy\right) \qquad \text{where} \qquad y^j = g^{\langle j \rangle}(w) \ . \tag{21}$$

In addition,

$$\widetilde{G}^{\langle k \rangle}(z) = (e^y - 1)^k \exp\left(z(e^y - 1) - wy\right) \qquad \text{where} \qquad y^j = g^{\langle j \rangle}(w) \ . \tag{22}$$

for $w = \lambda z$. □

Proof. We first identify the coefficients at $z^i g^{\langle j \rangle}(w)$ without worrying about the convergence and the error term. Thus, we apply the formal series approach that leads to the following

$$\begin{aligned} \widetilde{G}(z) &= \sum_{n=0}^{\infty} g(n) \frac{z^n}{n!} e^{-z} = \sum_{n=0}^{\infty} \frac{z^n}{n!} e^{-z} \sum_{j=0}^{\infty} g^{\langle j \rangle}(w) \frac{(n-w)^j}{j!} \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} a_{ij}(\lambda) z^i g^{\langle j \rangle}(w) \end{aligned}$$

where in the second equality we formally applied Taylor's series expansion of $g(n)$ around w, while the last line is a consequence of a re-arrangement of coefficients (i.e., expanding e^z and $(n-w)^j$ into the Taylor series). In order to identify the coefficient $a_{ij}(\lambda)$, we use a formal series of $e^{y(n-w)}$ with respect to n and substitute $y^j = g^{\langle j\rangle}(w)$. We obtain

$$\begin{aligned}\widetilde{G}(z) &= \sum_{n=0}^{\infty} \frac{z^n}{n!} e^{-z} \sum_{j=0}^{\infty} y^j \frac{(n-w)^j}{j!} & y^j = g^{\langle j\rangle}(w) \\ &= \sum_{n=0}^{\infty} \frac{z^n}{n!} e^{-z} e^{y(n-w)} = \exp\left(z(e^y-1) - wy\right) & y^j = g^{\langle j\rangle}(w)\ .\end{aligned}$$

Thus, we found an explicit series representation for the coefficients $a_{ij}(\lambda)$, but in order to recover the error term $O(z^{\beta-m-1})$, we must enter the realm of analysis. Unlike the depoissonization discussed in Theorem 3, this is quite easy. Indeed, in [14, 15] we proved that *if a complex function $g(z) = O(z^\beta)$ in a linear cone, then the jth derivative $g^{\langle j\rangle}(z)$ in the cone can be bounded as $g^{\langle j\rangle}(z) = O(z^{\beta-j})$*. From the above, we conclude that the error term is

$$\sum_{j=0}^{2m+1} a_{m+1,j}(\lambda) z^m g^{\langle j\rangle}(w) = O(z^{\beta-m-1})$$

as $z \to \infty$, since $w = \lambda z$.

Finally, we prove that $a_{ij} = 0$ for $j < 2i$. Let $f(x,y) = \exp\left(x(e^y-1) - xy\right)$. Observe that $f(xy^{-2}, y)$ is analytic at $x = y = 0$, hence its Laurent series expansion possesses only terms like $x^i y^{j-2i}$ with *nonnegative* powers leading to $j \geq 2i$ for nonzero coefficients a_{ij}, as desired. This proves the theorem. ■

Remark. To visualize expansion (19) and (21) we illustrate them in the following two examples:

$$\begin{aligned} g_n &= \widetilde{G}(n) - \frac{1}{2} n \widetilde{G}^{\langle 2\rangle}(n) + \frac{1}{3} n \widetilde{G}^{\langle 3\rangle}(n) + \frac{1}{8} n^2 \widetilde{G}^{\langle 4\rangle}(n) - \frac{1}{4} n \widetilde{G}^{\langle 4\rangle}(n) - \frac{1}{6} n^2 \widetilde{G}^{\langle 5\rangle}(n) - \\ &\quad - \frac{1}{48} n^3 \widetilde{G}^{\langle 6\rangle}(n) + \frac{1}{5} n \widetilde{G}^{\langle 5\rangle}(n) + \frac{13}{72} n^2 \widetilde{G}^{\langle 6\rangle}(n) + \frac{1}{24} n^3 \widetilde{G}^{\langle 7\rangle}(n) + \frac{1}{384} n^4 \widetilde{G}^{\langle 8\rangle}(n) - \\ &\quad - \frac{1}{6} n \widetilde{G}^{\langle 6\rangle}(n) - \frac{11}{60} n^2 \widetilde{G}^{\langle 7\rangle}(n) - \frac{17}{288} n^3 \widetilde{G}^{\langle 8\rangle}(n) - \frac{1}{144} n^4 \widetilde{G}^{\langle 9\rangle}(n) - \frac{1}{3840} n^5 \widetilde{G}^{\langle 10\rangle}(n) \\ &\quad + O(n^{\beta-6})\ , \end{aligned}$$

and

$$\begin{aligned} \widetilde{G}(z) &= g(z) + \frac{1}{2} z\, g^{\langle 2\rangle}(z) + \frac{1}{6} z\, g^{\langle 3\rangle}(z) + \left(\frac{1}{24} z + \frac{1}{8} z^2\right) g^{\langle 4\rangle}(z) + \left(\frac{1}{120} z + \frac{1}{12} z^2\right) g^{\langle 5\rangle}(z) \\ &+ \left(\frac{1}{720} z + \frac{5}{144} z^2 + \frac{1}{48} z^3\right) g^{\langle 6\rangle}(z) + \left(\frac{1}{5040} z + \frac{1}{90} z^2 + \frac{1}{48} z^3\right) g^{\langle 7\rangle}(z) \end{aligned}$$

$$
\begin{aligned}
&+ \quad \left(\frac{1}{40320}\, z + \frac{17}{5760}\, z^2 + \frac{7}{576}\, z^3 + \frac{1}{384}\, z^4\right) g^{\langle 8\rangle}(z) \\
&+ \quad \left(\frac{1}{362880}\, z + \frac{41}{60480}\, z^2 + \frac{137}{25920}\, z^3 + \frac{1}{288}\, z^4\right) g^{\langle 9\rangle}(z) + \mathrm{O}(\, z^{\beta-6}\,)
\end{aligned}
$$

for $m=6$. The above expansion is obtained from (21) after setting $w=z$. $\square$

As mentioned above, to recover the original g_n we must apply Theorem 3 which requires the verification of conditions (I) and (O). We shall take care of condition (O) (and in fact of condition (I), too) in the lemma below which can be considered our main new contribution to the analytic depoissonization, and should find many of applications in other problems of information theory.

Lemma 1 *Let $g(z)$ be an analytic continuation of a sequence g_n whose Poisson transform is $\widetilde{G}(z)$, and such that $g(z) = O(z^\beta)$ in a linear cone. Then, for some θ_0 and for all linear cones $\mathcal{S}_\theta$ ($\theta < \theta_0$), there exists $\alpha < 1$ and $A > 0$ such that*

$$
z \notin \mathcal{S}_\theta \qquad \Rightarrow \qquad |\widetilde{G}(z)e^z| \le Ae^{\alpha|z|} \ .
$$

Proof: Let $\mathcal{S}_{\theta_0}$ be the linear cone for which the polynomial bound over $g(z)$ holds. Let also $g^*(s)$ be the Laplace transform of the function $g(x)$ of a real variable x:

$$
g^*(s) = \int_0^\infty g(x)e^{-sx}dx
$$

defined for $\Re(s) > 0$. It is well known that (cf. [5]) the inverse Laplace transform of $g^*(s)$ exists in $\Re(s) > 0$ and one can write

$$
g(x) = \frac{1}{2\pi i}\int_{\varepsilon - i\infty}^{\varepsilon + i\infty} g^*(s)e^{sx}ds
$$

with $\varepsilon > 0$. In addition, $g^*(s)$ is absolutely integrable on the line of the integration.

Observe now that the exponential generating function $G(z) = \widetilde{G}(z)e^z$ of $g(n)$ can be represented in terms of $g^*(s)$ as follows

$$
\begin{aligned}
G(z) \quad &= \; = \sum_{n=0}^{\infty} g(n)\frac{z^n}{n!} = \sum_{n=0}^{\infty} \frac{1}{2\pi i}\int_{\varepsilon - i\infty}^{\varepsilon + i\infty} g^*(s)\exp(ns)\frac{z^n}{n!}ds \\
&= \; \frac{1}{2\pi i}\int_{\varepsilon - i\infty}^{\varepsilon + i\infty} g^*(s)\sum_{n=0}^{\infty}\exp(ns)\frac{z^n}{n!}ds = \frac{1}{2\pi i}\int_{\varepsilon - i\infty}^{\varepsilon + i\infty} g^*(s)\exp\left(ze^s\right)ds
\end{aligned}
$$

where the interchange of the integral and the summation is justifiable since both converge in their domains of definition. Thus, the Poisson transform $\widetilde{G}(z)$ becomes

$$
\widetilde{G}(z)e^z = \frac{1}{2\pi i}\int_{\varepsilon - i\infty}^{\varepsilon + i\infty} g^*(s)e^{e^s z}ds \ . \tag{23}
$$

To take advantage of the above formula, we need an extension of the Laplace transform of a real variable to a complex variable. Let $g(z)$ be an analytical continuation of $g(x)$ in a cone $\mathcal{S}_\theta$ where $\theta < \theta_0$. In Appendix A, we show that the inverse Laplace $g^*(s)$ of the function $g(z)$ (of complex z) exists in a *bigger* cone $\mathcal{S}_{\theta+\pi/2}$ for all $\theta < \theta_0$ *provided* $g(z) = O(z^\beta)$ in the cone $\mathcal{S}_{\theta_0}$. More precisely:

$$g(z) = \frac{1}{2\pi i} \int_{\mathcal{L}_\varepsilon} g^*(s) e^{sz} ds$$

where $\mathcal{L}_\varepsilon$ is a piece-linear curve that parallels the boundary of the cone $\mathcal{S}_{\theta+\pi/2}$ at distance ε. That is, in $s = (x, y)$ coordinates, the curve $\mathcal{L}_\varepsilon$ can be described as

$$y = -\mathrm{sign}(y)(x - \varepsilon)\tan(\pi/2 - \theta) = -\mathrm{sign}(y)(x - \varepsilon)\cot\theta$$

where $\mathrm{sign}(y)$ is the sign function (i.e., equal to 1 when $y \geq 0$ and -1 otherwise). We also define $\mathcal{L}_0$ which is a curve obtained from $\mathcal{L}_\varepsilon$ as $\varepsilon \to 0$, that is, having the description $y = -\mathrm{sign}(y)x\cot\theta$. In view of the above, we can write (23) as

$$\widetilde{G}(z)e^z = \frac{1}{2\pi i} \int_{\mathcal{L}_\varepsilon} g^*(s) e^{e^s z} ds \ . \tag{24}$$

We can now upper bound $\widetilde{G}(z)e^z$ as follows

$$|\widetilde{G}(z)e^z| \leq \frac{1}{2\pi} \int_{\mathcal{L}_\varepsilon} |g^*(s)| \exp(\Re(e^s z)) ds \leq \frac{1}{2\pi} \int_{\mathcal{L}_\varepsilon} |g^*(s)| \exp\left(\Re(e^s e^{i\theta})|z|\right) ds \tag{25}$$

since for $z \notin \mathcal{S}_\theta$ we have $\cos(\arg z) \leq \cos\theta$ for $|\theta| \leq \pi/2$. To complete the proof, we show that $\Re(e^s e^{i\theta}) < \alpha$ for some $\alpha < 1$ and all $s \in \mathcal{L}_\varepsilon$. Provided it is true, we obtain immediately $|\widetilde{G}(z)e^z| \leq Ae^{\alpha|z|}$ for $\alpha < 1$ where $A = \frac{1}{2\pi}\int_{\mathcal{L}_\varepsilon} |g^*(s)| ds$.

We concentrate now on showing that $\Re(e^s e^{i\theta}) < \alpha < 1$. We study the image $\mathcal{I}_\varepsilon$ of $\mathcal{L}_\varepsilon$ over the function e^s which is plotted in Figure 1. (In fact, in Figure 1 we assume $\theta = \pi/6$, and $\mathcal{I}_\varepsilon$ with $\varepsilon = 0.1$ has the following parametric description: $(\exp(-t/2+0.1)\cos(\pm t),\ \exp(-t/2+0.1)\sin(\pm t))$.) When $\varepsilon \to 0$, this image tends to the image $\mathcal{I}_0$ of $\mathcal{L}_0$. Observe that $\mathcal{I}_0$ is contained in the unit disk, and the only common point of $\mathcal{I}_0$ and the unit circle is at $s = 0$. Now, we rewrite (24) as follows:

$$\widetilde{G}(z)e^z = \frac{1}{2\pi i} \int_{\mathcal{L}_\varepsilon} g^*(s) \exp[e^s e^{i\arg(z)}|z|] ds \ .$$

The above expression is equivalent to applying to the image $\mathcal{I}_\varepsilon$ of $\mathcal{L}_\varepsilon$ a rotation by the angle $\arg(z)$. Let us now fix an arbitrary $0 < \theta < \theta_0$, and consider for a moment only the image $\mathcal{I}_0$ of $\mathcal{L}_0$. Observe that if one rotates the image $\mathcal{I}_0$ by a non zero argument, then the new image has the real part strictly less than 1. Of course, this is the case for all rotations of argument

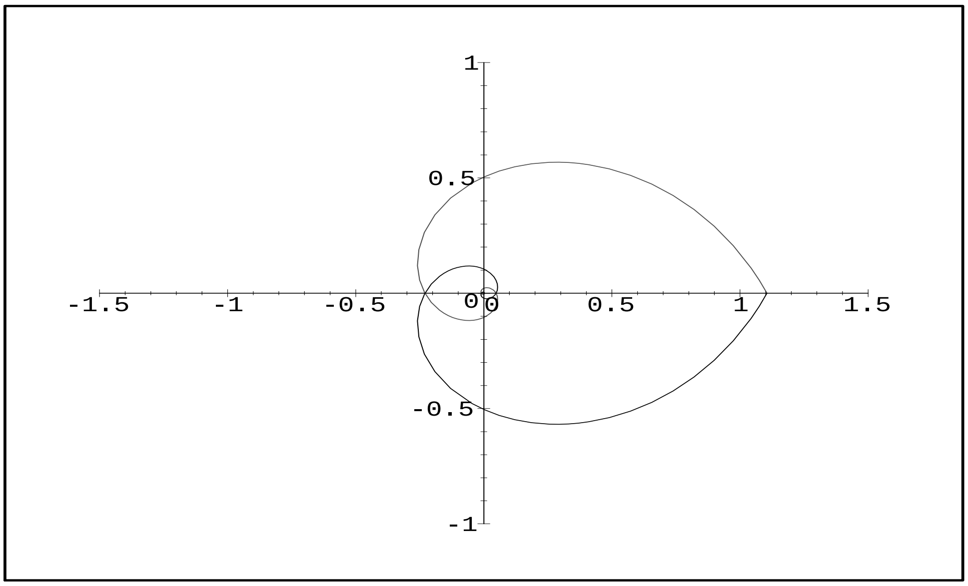

Figure 1: Image of $\mathcal{L}_\varepsilon$ with $\varepsilon = 0.1$ by e^s.

strictly greater than θ mod 2π. In fact, the real part will be smaller than $1 - O(\theta^2)$. Finally, considering the image $\mathcal{I}_\varepsilon$ of $\mathcal{L}_\varepsilon$, we easily can tune up ε so that the real part of the rotated images of $\mathcal{L}_\varepsilon$ remains smaller than some α such that $1 - O(\theta^2) < \alpha < 1$ for all arguments greater than θ. Thus, $\Re(e^s e^{i\theta}) < \alpha < 1$, which completes the proof. ■

3.3 Finishing the Derivations

Finally, we return to our original problem and compute the entropy expressed in (14) as

$$h_n = -\ln(n!) - n(p \ln p + q \ln q) + \mathcal{P}^{-1}\left(\widetilde{F}(zp) + \widetilde{F}(zq); n\right)$$

where $\widetilde{F}(\lambda z) = \mathcal{P}(f_n; \lambda z)$ with λ either equal to p or q. In the above,

$$\ln(n!) = f_n = f(n) = \ln \Gamma(n+1)$$

where $\Gamma(z)$ is the Euler gamma function. Observe that we do *not* have an explicit formula for $\widetilde{F}(z)$, so we must use Theorem 4 to obtain an asymptotic expansion of $\widetilde{F}(\lambda z)$. This is needed to depoissonize $\widetilde{F}(\lambda z)$ by Theorem 3. Let $g(\lambda n) = \mathcal{P}^{-1}\left(\widetilde{F}(\lambda z); n\right)$. Then

$$h_n = -\ln \Gamma(n+1) - n(p \ln p + q \ln q) + g(pn) + g(qn),$$

and we must find asymptotic expansion of $g(\lambda n)$.

Let us now concentrate on depoissonizing $\widetilde{F}(\lambda z)$, that is, finding asymptotic expansion of $g(\lambda n)$. From Stirling's formula we easily verify that $\widetilde{F}(z) = O(z \ln z)$ in a cone around real axis, thus condition (I) of Theorem 3 holds. But, since also $f(z) = O(z \ln z)$, then by Lemma 1 condition (O) is automatically verified. Therefore, (17) of Theorem 3 implies

$$g(\lambda n) = \exp(n \ln(1+\lambda u) - n\lambda u) \qquad \text{where} \qquad u^j = \widetilde{F}^{\langle j \rangle}(\lambda n) \ .$$

The problem now reduces to the estimation of $\widetilde{F}^{\langle j \rangle}(\lambda n)$. We, of course, shall use Theorem 4. By (22) we have

$$\widetilde{F}^{\langle k \rangle}(\lambda n) = (e^y - 1)^k \exp\left(n\lambda(e^y - 1) - n\lambda y\right) \qquad \text{where} \qquad y^j = f^{\langle j \rangle}(\lambda n) \ .$$

Putting everything together, we finally obtain

$$\begin{aligned} g(\lambda n) &= \exp(n \ln(1+\lambda u) - n\lambda u) && \text{where} \quad u^j = \widetilde{F}^{\langle j \rangle}(\lambda n) \\ &= \sum_{i,j \ge 0} b_{ij} n^i \lambda^j (e^y - 1)^j \exp\left(n\lambda(e^y - 1) - n\lambda y\right) && \text{where} \quad y^j = f^{\langle j \rangle}(\lambda n) \\ &= \exp\left(n \ln(1 + \lambda(e^y - 1)) - n\lambda y\right) && \text{where} \quad y^j = f^{\langle j \rangle}(\lambda n). \quad (26) \end{aligned}$$

The reader should identify in (26) the function defined in (9) of Theorem 2, that is,

$$\exp\left(x \log\left(1 + \lambda(e^y - 1)\right) - \lambda x y\right) = \sum_{i=0}^{\infty} \sum_{k=2i}^{\infty} c_{i,k}(\lambda) x^i y^k .$$

The first few terms of the above series are as follows:

$$\begin{aligned} \exp(x \log(1 &+ \lambda(e^y - 1)) - \lambda x y) = 1 + \frac{1}{2}\lambda(1-\lambda)xy^2 + \frac{1}{6}\lambda(1 - 3\lambda + 2\lambda^2)xy^3 \\ &+ \frac{1}{24}(\lambda - 7\lambda^2 + 12\lambda^3 - 6\lambda^4)xy^4 + \frac{1}{8}(\lambda^2 - 2\lambda^3 + \lambda^4)x^2y^4 + O(y^5) \end{aligned}$$

Thus, the entropy h_n in terms of the function $f(n)$ can be expressed as

$$\begin{aligned} h_n &= -f(n) + f(np) + f(nq) - n(p \ln p + q \ln q) \qquad (27) \\ &+ \sum_{i=1}^{\infty} \sum_{k=2i}^{\infty} n^i \left(c_{i,k}(p) p^i f^{\langle k \rangle}(np) + c_{i,k}(q) q^i f^{\langle k \rangle}(nq) \right) . \end{aligned}$$

To complete our derivation, we need asymptotic expansions of

$$\begin{aligned} f(z) &= \ln \Gamma(z+1) = \ln z + \ln \Gamma(z), \\ f^{\langle k \rangle}(z) &= \frac{\mathrm{d}^k}{\mathrm{d}z^k} \Gamma(z+1) = \Psi^{(k-1)}(z+1) = \Psi^{(k-1)}(z) + \frac{(-1)^{k-1} k!}{z^k}, \end{aligned}$$

where $\Psi^{(k)}(z)$ is the kth derivative of the psi function (i.e., $\Psi(z) = \Gamma'(z)/\Gamma(z)$). The last equality of the above follows from the well–known property of the psi function, namely

$\Psi(z+1) = \Psi(z) + 1/z$ (cf. [1]). Using asymptotic series expansions of the gamma function and the psi function we obtain (cf. [1])

$$\begin{aligned}
f(z) &= \left(z+\frac{1}{2}\right)\ln z - z\frac{1}{2} + \sum_{m=1}^{\infty}\frac{B_{2m}}{2m(2m-1)z^{2m-1}} \\
f^{\langle k\rangle}(z) &= \Psi^{(k-1)}(z+1) = \frac{(-1)^k(k-2)!}{z^{k-1}} + \frac{(-1)^{k-1}(k-1)!(2k-1)}{2z^k} \qquad \text{for } \ k \geq 2 \\
&\quad + (-1)^k \sum_{l=1}^{\infty}\frac{B_{2l}(2l+k-2)!}{(2l)!\ z^{2l+k-1}}
\end{aligned}$$

where B_n is the nth Bernoulli number defined as

$$\sum_{n=0}^{\infty} B_n \frac{z^n}{n!} = \frac{z}{e^z - 1}. \tag{28}$$

Finally, substituting the above in (27), we obtain (8) of Theorem 2.

4 More Applications

In the previous section, we presented a detailed derivation of a full asymptotic expansion for Shannon entropy of the binomial distribution. In fact, we did much more since we provided a method based on analytic poissonization and depoissonization that allows to derive asymptotic expansions for entropy-like sums. To convene this fact, we discuss below two other examples, namely, the entropy evaluation of the negative binomial distribution and the redundancy of "add-β" rule (cf. [20]).

4.1 The Entropy of the Negative Binomial Distribution

Let us consider the evaluation of the entropy of the the negative binomial distribution. One must deal with a sum of the following form

$$g_n = \sum_{k\geq n}\binom{k}{n}p^n q^{k-n} f_n \qquad n \geq 0$$

where f_n is a sequence (e.g., $f_n = \ln(n!)$), and $q = 1-p$. Our goal is to find an equivalence of (26) for the negative binomial distribution.

Let us define the exponential generating function of f_n as $F(z) = \sum_{n\geq 0} f_n z^n/n!$, and let $\widetilde{F}(z) = F(z)e^{-z}$ be its Poisson transform. Let also

$$G_n(v) = \sum_{k\geq n}\binom{k}{n}p^n q^{k-n} f_n \frac{v^k}{k!}\ .$$

Observe that

$$g_n = \int_0^\infty G_n(v)e^{-v}dv \ , \tag{29}$$

since $\int_0^\infty x^k e^{-x}dx = k!$. Finally, we obtain

$$\begin{aligned}\sum_{n=0}^{\infty} G_n(v)z^n &= \sum_{k=0}^{\infty}\frac{f_k}{k!}\sum_{n=0}^{k}\binom{k}{n}(pvz)^n(qv)^{k-n} \\ &= F(pzv+qv) = \widetilde{F}(pzv+qv)e^{pzv+qv} \ .\end{aligned}$$

The last expression can be used to extract $G_n(v)$ (through a depoissonization of $\widetilde{F}(pzv+qv)$), and ultimately g_n by (29). Indeed, to accomplish this goal, we proceed formally as follows: Note first that in our formal series approach

$$\begin{aligned}\widetilde{F}(z)e^z &= \sum_{n=0}^{\infty}\frac{z^n}{n!}f(n) = \sum_{n=0}^{\infty}\frac{z^n}{n!}\sum_{j=0}^{\infty}\frac{(n-w/p)^j}{j!}f^{\langle j\rangle}(w/p) \\ &= \sum_{n=0}^{\infty}\frac{z^n}{n!}\sum_{j=0}^{\infty}\frac{(n-w/p)^j}{j!}y^j = \exp\left(ze^y - y\frac{w}{p}\right) \qquad \text{where} \quad y^j = f^{\langle j\rangle}(w/p)\end{aligned}$$

provided $f(z)$ is an analytic continuation of f_n. Hence

$$G_n(v) = [z^n]\exp\left((pzv+qv)e^y - y\frac{w}{p}\right) = \frac{(e^y pv)^n}{n!}e^{qve^y - yw/p}$$

where $[z^n]f(z)$ is the coefficient of $f(z)$ at z^n. Finally,

$$\begin{aligned}g_n &= \int_0^\infty \frac{(e^y pv)^n}{n!}e^{qve^y-yn/p}e^{-v}dv \\ &= \left(\frac{pe^{qy/p}}{1-qe^y}\right)^n \qquad\qquad \text{where} \qquad y^j = f^{\langle j\rangle}(n/p) \ .\end{aligned} \tag{30}$$

In order to use (30) — which is the corresponding expansion to (26) — we only need to compute derivatives of $f(z)$ at $z = n/p$. In fact, expanding (30) into a formal Taylor's series we obtain

$$\begin{aligned}g_n &= f(n/p) + \frac{1}{2}n\left(\frac{q}{p}+\frac{q^2}{p^2}\right)f^{\langle 2\rangle}(n/p) \\ &+ \left(n\left(\frac{q}{6p}+\frac{q^2}{2p^2}+\frac{q^3}{3p^3}\right) - \frac{1}{2}n^2\left(\frac{q^2}{p^2}+\frac{q^3}{p^3}+\frac{q^2}{p^3}\right)\right)f^{\langle 3\rangle}(n/p) + O(n^3 f^{\langle 4\rangle}(n/p) \ .\end{aligned}$$

The above leads to a full asymptotic expansion as long as the function $f(z)$ has a polynomial growth, as we assumed throughout this paper.

4.2 The Redundancy of the "Add-β" Rule

During the revision of our paper, Krichevskiy's analysis [20] of the redundancy of the "add-β" rule for predicting the $(n+1)$st symbol was published. The following function is studied in [20]:

$$F(n,\beta,p) = \beta + np\ln(np) - np\sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}\ln(k+\beta),$$

where $0 < p < 1$ and β is a constant. The author of [20] calls this function "rather complicated", and approximates it by a simpler function (i.e., a Poisson approximation of the binomial distribution). Using our approach, we can easily provide a full asymptotic expansion of it. Indeed, let

$$g_n = \sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}\ln(k+\beta).$$

Then, the Poisson transform $\widetilde{G}(z) = \mathcal{P}(g_n; z)$ of g_n is $\widetilde{G}(z) = \widetilde{H}(zp)$ where

$$\widetilde{H}(z) = \sum_{n\geq 0}\ln(n+\beta)\frac{z^n}{n!}e^{-z}.$$

Let $f(z) = \ln(z+\beta)$. As in Section 3, we immediately obtain

$$\begin{aligned} g_n &= \sum_{i\geq 0}\sum_{j\geq 2i} c_{ij}(p)n^i f^{\langle j\rangle}(np) \\ &= f(np) + \sum_{i\geq 1}\sum_{j\geq 2i} c_{ij}(p)n^i f^{\langle j\rangle}(np) \\ &= \ln(np+\beta) - \frac{q}{2np} + \sum_{k\geq 2}\frac{a_k(p,\beta)}{n^k}, \end{aligned}$$

where $c_{ij}(p)$ are defined in (9), and $a_k(p,\beta)$ are explicitly computable constants (in terms of $c_{ij}(p)$). In view of this, we finally obtain the following full asymptotic expansion for Krichevskiy's function

$$F(n,\beta,p) = \frac{q}{2} - \sum_{k\geq 1}\frac{1}{n^k}\left(pa_{k+1}(p,\beta) + \frac{(-1)^k\beta^{k+1}}{(k+1)p^k}\right) \tag{31}$$

where, as always, $q = 1-p$.

5 Concluding Remarks

In this concluding remarks, we repeat the claim that the method proposed above may indeed lead to a full asymptotic expansion for a large spectrum of entropy-like sums. More precisely:

Analytic poissonization and depoissonization methodology gives an unified framework for determining a full asymptotic expansion of the following sum

$$\sum_{k_1+\cdots+k_m=n} \binom{n}{k_1 \ \cdots \ k_m} p_1^{k_1} \cdots p_m^{k_n} g_1(n) \cdots g_m(n) \tag{32}$$

where $g_i(z)$ are analytical functions defined in a cone around the real axis with polynomial growth (e.g. $g_1(z) = \ln^3 \Gamma(z+1)$ or $g_2(z) = z^2$), and $p_1+\cdots+p_m = 1$.

Indeed, via poissonization the above sum becomes $\widetilde{G}_1(p_1 z) \cdots \widetilde{G}_m(p_m z)$ where $\widetilde{G}_i(z)$ is the Poisson transform of the sequence $g_i(n)$. As long as $g_i(n)$ have analytic continuations of polynomial growths, we can estimate the Poisson transforms $\widetilde{G}_i(z)$ by Theorem 4, and then use the analytic depoissonization (cf. Theorem 3 and Lemma 1) to obtain the asymptotic expansion up to any desired order of accuracy.

The foundation of the presented methodology is clearly rooted in complex analysis while the problems solved are at the heart of information theory. This paper is a simple and straightforward example of how analytic function theory are applicable to problems of information theory leading to analytic information theory.

APPENDIX A: Laplace Transform of a Complex Variable

We want to extend the Laplace transform of a function $g(x)$ of real x to a function $g(z)$ of complex z where $g(z)$ is an analytic continuation of $g(x)$ to a cone $\mathcal{S}_{\theta_0}$ for $\theta_0 < \pi/2$. We expected to find such an extension in books (cf. [5] Chap. 11) but since we fail to provide a definite reference, we include here a brief discussion.

First of all, if $g(z)$ is an analytic continuation of $g(x)$ in a cone $\mathcal{S}_{\theta_0}$, then we can write $z = xe^{i\theta}$. When $-\theta_0 \le \theta \le \theta_0$, then $z = xe^{i\theta}$ sweeps the whole cone as $0 \le x < \infty$. But, one can easily find the Laplace transform of $g(xe^{i\theta})$. Indeed, for $\Re(s) > 0$ we have:

$$e^{-i\theta} g^*(e^{-i\theta} s) = \int_0^\infty g(xe^{i\theta}) e^{-sx} dx.$$

Secondly, we should observe that the plane of convergence (i.e., $\mathcal{S}_{\pi/2}$ for $g^*(s)$ of a function $g(x)$ of real x) extends to a bigger cone, namely: $\mathcal{S}_{\pi/2+\theta}$ for any $\theta < \theta_0$ *provided* $g(z)$ has a *subexponential* growth (e.g., polynomial growth) in the cone $\mathcal{S}_{\theta_0}$. Indeed, let $g(z) = O(e^{\delta z})$ for any $\delta > 0$. Then, clearly the Laplace transform $g^*(s)$ is defined for all s satisfying $\Re\left(se^{i\theta}\right) > 0$, that is, for $-\pi/2 < \arg(s) + \theta < \pi/2$ where $-\theta_0 \le \theta \le \theta_0$. Thus, the Laplace transform $g^*(s)$ is defined the cone $\mathcal{S}_{\theta_0+\pi/2}$.

Finally, since $g^*(s)$ is analytic, then by Cauchy's theorem we can write the inverse Laplace transform as follows:

$$g(x) = \frac{1}{2\pi i} \int_{\mathcal{L}} g^*(s) e^{sx} ds \ ,$$

where $\mathcal{L}$ is any curve inside $\mathcal{S}_{\theta_0+\pi/2}$.

ACKNOWLEDGEMENT

We thank C. Knessl and H-K. Hwang for comments concerning this paper.

References

[1] M. Abramowitz, and I. Stegun, *Handbook of Mathematical Functions*, Dover, New York 1964.

[2] Aldous, D., *Probability Approximations via the Poisson Clumping Heuristic*, Springer Verlag, New York 1989.

[3] S. Chang and E. Weldon, Coding for T-user Multiple-Access Channels, *IEEE Trans. Information Theory*, 25, 684–691, 1979.

[4] L. Comtet, *Advanced Combinatorics*, D. Reidel Publishing Company, Dordrecht, 1974.

[5] G. Doetsch, *Handbuch der Laplace Transformation,* Vol. 1–3, Birkhäuser-Verlag, Basel 1955.

[6] M. Drmota and U. Schmid, The Analysis of the expected Successful Operation Time of Slotted ALOHA, *IEEE Trans. Information Theory*, 39, 1567–1577, 1993

[7] G. Fayolle, P. Flajolet, M. Hofri and P. Jacquet, Analysis of Stack Algorithm for Random Multiple-Access Communication, *IEEE Trans. Information Theory*, 31, 244–254, 1985.

[8] P. Flajolet and A. Odlyzko, Singularity Analysis of Generating Functions, *SIAM J. Disc. Methods*, 3, 216-240, 1990.

[9] P. Flajolet, and R. Sedgewick, *Analytical Combinatorics*, in preparation; see also INRIA TR-1888 1993, TR-2026 1993 and TR-2376 1994.

[10] B. Hughes and A. B. Cooper, Near Optimal Multiuser Codes for Binary Adder Channel, *IEEE Trans. Information Theory*, 42, 387–398, 1996

[11] P. Jacquet, Analytical Information Theory in Service of Queueing with Aggregated Up/Down Arrivals, *Proc. 35th Allerton Conference*, Monticello 1997.

[12] P. Jacquet and W. Szpankowski, Analysis of Digital Tries with Markovian Dependency, *IEEE Transactions on Information Theory*, 37, 1470-1475, 1991.

[13] P. Jacquet and W. Szpankowski, Autocorrelation on Words and Its Applications. Analysis of Suffix Trees by String-Ruler Approach, *J. Combin. Theory Ser. A*, 66, 237-269, 1994.

[14] P. Jacquet and W. Szpankowski, Asymptotic Behavior of the Lempel-Ziv Parsing Scheme and Digital Search Trees, *Theoretical Computer Science*, 144, 161-197, 1995.

[15] P. Jacquet and W. Szpankowski, Analytical Depoissonization and its Applications, *Theoretical Computer Science*, 199, 1998

[16] T. Kawabata, Exact Analysis of the Lempel-Ziv Algorithm for I.I.D. Sources, *IEEE Trans. Information Theory*, 39, 698-702, 1993.

[17] C. Knessl, Integral Representations and Asymptotic Expansions for Shannon and Renyi Entropies, preprint 1997.

[18] D.E. Knuth, *The Art of Programming*, Vol. I–III, Addison Wesley, 1973 and 1981.

[19] C. Krattenhaler and P. Slater, Asymptotic Redundancy for Universal Quantum Coding, preprint.

[20] R. Krichevskiy, Laplace's Law of Succession and Universal Coding, *IEEE Trans. Information Theory*, 44, 296–303, 1998.

[21] G. Louchard and W. Szpankowski, Average Profile and Limiting Distribution for a Phrase Size in the Lempel-Ziv Parsing Algorithm, *IEEE Trans. Information Theory*, 41, 478-488, 1995.

[22] G. Louchard and W. Szpankowski, On the Average Redundancy Rate of the Lempel-Ziv Code, *IEEE Trans. Information Theory*, 43, 2–8, 1997.

[23] G. Louchard, W. Szpankowski, and J. Tang, Average Profile for the Generalized Digital Search Trees and the Generalized Lempel-Ziv Algorithm, *SIAM J. Computing*, to appear.

[24] H. Mahmoud, *Evolution of Random Search Trees*, John Wiley & Sons, New York 1992.

[25] P. Mathys, P. Flajolet, Q-ary Collision Resolution Algorithms in Random-Access Systems with Free and Blocked Channel Access, *IEEE Trans. Information Theory*, 31, 217–243, 1985.

[26] A. Odlyzko, Asymptotic Enumeration, in *Handbook of Combinatorics*, Vol. II, (Eds. R. Graham, M. Götschel and L. Lovász), Elsevier Science, 1063-1229, 1995.

[27] A., Renyi, On Measures of Entropy and Information. *Proc. 4th Berkeley Symp. on Math. Statist. and Prob.*, 547-561, 1961.

[28] M. Régnier and P. Jacquet, New Results on the Size of Tries, *IEEE Trans. Information Theory*, 35, 203-205 (1989).

[29] R. Sedgewick and P. Flajolet, *Analysis of Algorithms*, Addison Wesley, Reading, Mass. 1995.

[30] W. Szpankowski, On a Recurrence Equation Arising in the Analysis of Conflict Resolution Algorithms, *Stochastic Models*, 3, 89-114, 1987.

[31] W. Szpankowski, On Asymptotics of Certain Sums Arising in Coding Theory, *IEEE Trans. Information Theory*, 41, 2087–2090, 1995.

[32] W. Szpankowski, On Asymptotics of Certain Recurrences Arising in Universal Coding, *Problems of Information Transmissions*, 1998 (to appear).

[33] W. Szpankowski, Techniques of the Average Case Analysis of Algorithms, in *Handbook on Algorithms and Theory of Computation* (Ed. M. Atallah), CRC 1998.

[34] H. Wilf, *generatingfunctionology*, Academic Press, Boston 1990.