ANALYTICAL DEPOISSONIZATION AND ITS APPLICATIONS TO COMBINATORICS AND ANALYSIS OF ALGORITHMS*

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. In combinatorics and analysis of algorithms often a Poisson version of a problem (called *Poisson model* or *poissonization*) is easier to solve than the original one, which is known as the *Bernoulli model*. Poissonization is a technique that replaces the original input by a Poisson process. More precisely, an analytical Poisson transform maps a sequence (e.g., characterizing the Bernoulli model) into a generating function of a complex variable. However, after poissonization one must *depoissonize* in order to translate the results of the Poisson model into the original (i.e., Bernoulli) model. We present here some analytical depoissonization results that fall into the following general scheme: *if the Poisson transform has an appropriate growth in the complex plane, then an asymptotic expansion of the sequence can be expressed in terms of the Poisson transform and its derivatives evaluated on the real line.* We illustrate our results on a few examples from combinatorics and the analysis of algorithms and data structures.

Résumé. En analyse combinatoire ou algorithmique, la version sous modèle de Poisson (ou dit "poissonisée") d'un problème est souvent plus facile à résoudre que le problème original, dit sous modèle de Bernoulli. La poissonisation est la technique qui consiste à remplacer un paramètre d'entrée entier fixé par une variable aléatoire de Poisson. Plus précisément, la transformation de Poisson analytique opère sur une séquence (spécifique au modèle Bernoulli) pour en faire une fonction génératrice à variable complexe. Inversement, après poissonisation on doit dépoissoniser afin de traduire les résultats obtenus sur le modèle de Poisson en des propriétés du modèle de Bernoulli original. Nous présentons quelques résultats de dépoissonissation analytique qui s'expriment dans le schéma suivant : *si la transformée de Poisson présente une croissance convenable dans le plan complexe, donc un développement asymptotique de la séquence originelle peut être obtenue au moyen de la transformée de Poisson et de ses dérivées évaluées sur l'axe réel.* Nous illustrons nos résultats par quelques exemples tirés de l'analyse combinatoire et algorithmique, notamment sur les structures de données.

1 Introduction

Consider the following scenario that often arises in combinatorial algorithms (e.g., sorting and hashing are prominent examples): n elements are placed randomly into m "bins". It is often called the "balls and urns" model. Questions arise such as how many urns are empty, how many balls are required to fill up all urns, etc. It is easy to see that the occupancy of urns are not independent (e.g., think all of balls falling into one urn, thus we *know* that all the remaining urns are empty). To overcome this difficulty an interesting probabilistic technique called *poissonization* was suggested (cf. [1, 8, 9, 17]).

*This work was partially supported by NATO Collaborative Grant CRG.950060.
*additional support by the Esprit Basic Research Action No. 7141 (Alcom II).
†Additionally support by NSF Grants NSF CCR-9201078, NCR-9206315 and NCR-9415491.

Namely, it is assumed that balls are "generated" by a Poisson process with mean n. Due to unique properties of Poisson process, the stream of balls are placed independently in every urn, thus overcoming the above mention difficulty. Observe, however, that poissonization has its own problems since one has to "extract" the original results from the Poisson model (i.e., depoissonize).

To the best of our knowledge, poissonization was introduced by Marek Kac [17] a half a century ago when investigating deviations between theoretical and empirical distributions. Gonnet and Munro [8] introduced the Poisson transform to the field of the analysis of algorithms. Recently, poissonization was further popularized in the context of analysis of algorithms and combinatorial structures by Aldous [1], Fill *et. al.* [3], Gonnet [7], Holst [9], Jacquet and Régnier [10, 11, 27], Jacquet and Szpankowski [12, 13], Poblete *et. al.* [24], Rais *et. al* [26] and others [3, 18, 16].

Before we spell out more succinctly depoissonization, we first describe another scenario. In the analysis of algorithms and/or enumeration of combinatorial structures, often the original problem – which we further call the Bernoulli model – is represented by a recurrence equation or by a functional/differential equation. For example, this situation arises in a large class of algorithms based on a splitting process and/or digital trees (cf. [3, 10, 11, 12, 13, 18, 19, 21, 26, 27]). Imbedding this splitting process into a Poisson process leads often to a more tractable functional/differential equations. We shall call such a technique *analytical poissonization*.

More formally, let g_n be a sequence describing a parameter of interest in the Bernoulli model of size n. The Poisson transform is: $\widetilde{G}(z) = \sum_{n\geq 0} g_n \frac{z^n}{n!}e^{-z}$, that is, the input n becomes a Poisson process with mean z, and z can be a complex number. Throughout, we assume that $\widetilde{G}(z)$ is an *entire* function. If $\widetilde{G}(z)$ is known, one can extract the coefficient $g_n = n![z^n]\widetilde{G}(z)e^z$. However, in most interesting situation $\widetilde{G}(z)$ satisfies a complicated functional/differential equation that is difficult to solve exactly. Nevertheless, one can find an asymptotic expansion of $\widetilde{G}(z)$ for $z \to \infty$ in a cone. Then, the main problem is how to find an asymptotic expansion of g_n from the asymptotics of $\widetilde{G}(z)$. We shall prove that under certain conditions on the growth of $\widetilde{G}(z)$ we have: $g_n = \widetilde{G}(n) +$ *lower order terms*. This is called *analytical depoissonization*, and our main goal is to provide easily applicable tools for depoissonization. In passing, we observe that probabilistic depoissonization faces the same kind of difficulties (cf. [1, 9]).

There is not much literature on analytical depoissonization, and most published results are tailored to specific problems. It was initiated by Jacquet and Régnier in the analysis of limiting distributions of depth and size in tries [10, 11, 27]. They introduced the basic idea of the analytical depoissonization on an (implicit) solution of a specific non–linear functional equation: *i.e.* the asymptotic analysis of a sequence of Cauchy integrals on the complex plan with saddle point-like estimates. Jacquet and Szpankowski [12, 13], and Rais *et al.* [26] extended this result in order to obtain a first general (simple) version of the depoissonization lemma. The results of this paper are in some sense an outcome of our depoissonization result from [13] where we deal with a nonlinear multiplicative differential/functional equation. However, all depoissonization results in this paper are new. Analytical depoissonization was also used in [3, 16, 18, 24], and we summarize (and improve) some of these results here. An exhaustive list of depoissonization results can be found in [14].

2 Depoissonization Results

Let Y_n be a characteristic of the Bernoulli model that depends on n (e.g., number of balls), and let Y_N be the corresponding quantity in the Poisson model where N is a random variable distributed as Poisson with parameter z, that is, $Pr\{N = k\} = e^{-z}z^k/k!$. Then

$$\widetilde{Y}(z) := EY_N = \sum_{n\geq 0} E\left(Y_N \mid N = n\right) e^{-z}\frac{z^n}{n!} = \sum_{n\geq 0} EY_n e^{-z}\frac{z^n}{n!} \tag{1}$$

where the last equality follows from Poisson process properties. Throughout, we make the following assumption; *The Poisson transform $\widetilde{Y}(z)$ is an entire function of a complex variable z (no singularities of $\widetilde{Y}(z)$ in the whole complex plane).*

Equation (1) can be viewed as a definition of the *Poisson transform* (cf. [8]) with *z complex.* In general, we write $\widetilde{G}(z)$ or $\mathcal{P}(g_n; z)$ for the Poisson transform of a single-index sequence g_n, that is,

$$\widetilde{G}(z) = \mathcal{P}(g_n; z) \stackrel{def}{=} \sum_{n\geq 0} g_n \frac{z^n}{n!}e^{-z} .$$

However, poissonization and depoissonization is also useful for studying limiting distributions. For example, we may define a double-indexed sequence $g_{n,k} = Pr\{X_n = k\}$ (where X_n is a random variable), and then the Poisson transform becomes

$$\widetilde{G}(z,u) = \mathcal{P}(g_{n,k}; z) \stackrel{def}{=} \sum_{n=0}^{\infty} \frac{z^n}{n!}e^{-z} \sum_{k=0}^{\infty} g_{n,k}u^k .$$

As discussed above, our goal is to infer asymptotics of the Bernoulli model from the asymptotic behavior of the Poisson generating function for $z \to \infty$. However, in most cases the behavior of $\widetilde{G}(z,u)$ as $z \to \infty$ depends on the region in the complex plane in which $z \to \infty$. Usually, we must restrict the asymptotics of $\widetilde{G}(z,u)$ to a cone around the positive axis of $\Re(z)$.

Definition 1. *A linear $\mathcal{S}_\theta$ cone is defined as $\mathcal{S}_\theta = \{z : \ |\arg z| \leq \theta \quad |\theta| < \pi/2\}$* for real θ. □

The depoissonization is based on Cauchy's formula which becomes in our case:

$$g_n = \frac{n!}{2\pi i}\oint \frac{\widetilde{G}(z)e^z}{z^{n+1}}dz = \frac{n!}{n^n 2\pi i}\int_{-\pi}^{\pi} \widetilde{G}(ne^{it}) \exp\left(ne^{it}\right) e^{-nit} dt . \tag{2}$$

We evaluate this integral inside and outside a cone $\mathcal{S}_\theta$. During such a derivation, we often restore to the following results that are of their own interest:

Lemma 1 (N. G. Bruijn [2]) *The following identities are true:*

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^k e^{-tx^2}dx = \begin{cases} 0 & k = 1,3,5,\ldots \\ \frac{t^{-1/2-k/2}k!}{(k/2)!2^{k+1/2}} & k = 2,4,6,\ldots \end{cases} \tag{3}$$

and

$$\int_{\theta}^{\infty} x^k e^{-tx^2}dx = O\left(e^{-\frac{1}{2}t\theta^2}\right) \tag{4}$$

where θ is a real number.

Lemma 2 (Cauchy's Estimate) *Suppose that f is an analytic function in a domain containing the disc $B(z_0, r) = \{z : |z - z_0| \le r\}$, for some $z_0 \in \mathbb{C}$ and $r > 0$. Suppose further that $|f(z)| \le M$ when $z \in B(z_0, r)$. Then,*

$$|f^{\langle k \rangle}(z_0)| \le k! M r^{-k}, \qquad k \ge 0.$$

$f^{\langle k \rangle}(z)$ being the kth derivative of $f(z)$.

Lemma 3 *Let $\theta_0 < \pi/2$ and $\xi > 0$ be such that for all $z \in \mathcal{S}_{\theta_0}$:*

$$|z| > \xi \qquad \Longrightarrow \qquad |G(z)| \le B|z|^\beta \tag{5}$$

for some real β, and a positive constant $B > 0$. Then, for all $\theta < \theta_0$ there exists B' and $\xi' > \xi$ such that for all positive integers k the following holds

$$|z| > \xi' \qquad \Longrightarrow \qquad |G^{\langle k \rangle}(z)| \le k!(B')^k |z|^{\beta - k} \ . \tag{6}$$

Proof. We place a circle of radius $\Omega|z|$ for some $\Omega > 0$ at the closest corner of the cone $\mathcal{S}_{\theta_0}$ just after $\Re(z) = \xi$. Then, by Cauchy's formula we have $G^{\langle k \rangle}(z) = \frac{k!}{2\pi} \oint \frac{G(\omega) d\omega}{(\omega - z)^{k+1}}$. Setting $\omega = z + \Omega|z|e^{i\theta}$, and using (5) together with Lemma 2 we immediately prove (6). Another proof can be found in [13] for more general cones. ■

Now, we are in a position to formulate our main depoissonization result. We should point out that the theorem below is extended in many directions in our journal version [14].

Theorem 1 *Let $\widetilde{G}(z)$ be the Poisson transform of a sequence g_n that is assumed to be an entire function of z. We postulate that the following two conditions simultaneously hold*[1] *for some numbers $A, B, \xi > 0$, β, $\alpha < 1$, and a slowly varying function Ψ:*

(I) *For $z \in \mathcal{S}_\theta$*

$$|z| > \xi \quad \Rightarrow \quad |\widetilde{G}(z)| \le B|z|^\beta \Psi(|z|) \ , \tag{7}$$

(O) *For $z \notin \mathcal{S}_\theta$*

$$|z| > \xi \quad \Rightarrow \quad |\widetilde{G}(z)e^z| \le A \exp(\alpha|z|) \ . \tag{8}$$

Then, for large n

$$g_n = \widetilde{G}(n) + O(n^{\beta - 1}\Psi(n)) \ . \tag{9}$$

More precisely,

$$g_n = \widetilde{G}(n) - \tfrac{1}{2} n \widetilde{G}''(n) + O\left(n^{\beta - 2}\Psi(n)\right) \ . \tag{10}$$

The "Big-Oh" terms in (9) and (10) are uniform for any family of entire functions $\widetilde{G}(z)$ that satisfy the conditions with the same a, β, θ, c_1, c_2, z_0 and Ψ.

---

[1] *Recently, we have shown in [15] that if g_n has a polynomial growth, then the second condition (O) below is automatically satisfied.*

Proof: The proof relies on the Cauchy formula (2). By Stirling's approximation $n! = n^n e^{-n}\sqrt{2\pi n}(1+O(1/n))$, thus (2) becomes

$$\begin{aligned} g_n &= (1+O(1/n))\sqrt{\frac{n}{2\pi}}\int_{-\pi}^{\pi}\widetilde{G}(ne^{it})\exp\left(n\left(e^{it}-1-it\right)\right)dt \\ &= (1+O(1/n))(I_n+E_n), \end{aligned}$$

where

$$\begin{aligned} I_n &= \sqrt{\frac{n}{2\pi}}\int_{-\theta}^{\theta}\widetilde{G}(ne^{it})\exp\left(n\left(e^{it}-1-it\right)\right)dt \\ E_n &= \sqrt{\frac{n}{2\pi}}\int_{|t|\in[\theta,\pi]}\widetilde{G}(ne^{it})\exp\left(n\left(e^{it}-1-it\right)\right)dt \\ &= \frac{n^n e^{-n}\sqrt{2\pi n}}{n!2\pi i}\int_{|t|\in[\theta,\pi]}\frac{\widetilde{G}(z)e^z}{z^{n+1}}dz\ . \end{aligned}$$

We estimate the above two integrals. We begin with the latter. Observe that by condition (O) (cf. (8)), we obtain $|E_n|\le A'\sqrt{2\pi n}e^{-(1-\alpha)n}$ where A' depends only on A and ξ, thus E_n exponentially decays to zero.

Having E_n "under hat", we can turn our attention to integral I_n which is more intricate to handle. First of all, we substitute $t=t'/\sqrt{n}$ to get

$$I_n = \frac{1}{\sqrt{2\pi}}\int_{-\theta\sqrt{n}}^{\theta\sqrt{n}}\widetilde{G}(ne^{it/\sqrt{n}})\exp\left(n\left(e^{it/\sqrt{n}}-1-it/\sqrt{n}\right)\right)dt\ . \qquad (11)$$

Let $h_n(t)=\exp\left(n\left(e^{it/\sqrt{n}}-1-it/\sqrt{n}\right)\right)$. We need to estimate $h_n(t)$ in the interval $t\in[-\theta\sqrt{n},\theta\sqrt{n}]$, and find the Taylor expansion of it in a smaller interval, say for $t\in[-\log n,\log n]$. The latter restriction is necessary since $\frac{t}{\sqrt{n}}=O(1)$ for $t\in[-\theta\sqrt{n},\theta\sqrt{n}]$. Thus, we can split the integral I_n into two terms I'_n and I''_n, that is, $I_n=I'_n+I''_n$ where

$$\begin{aligned} I'_n &= \frac{1}{\sqrt{2\pi}}\int_{-\log n}^{\log n}\widetilde{G}(ne^{it/\sqrt{n}})\exp\left(n\left(e^{it/\sqrt{n}}-1-it/\sqrt{n}\right)\right)dt\ , \\ I''_n &= \frac{1}{\sqrt{2\pi}}\int_{t\in[-\theta\sqrt{n},-\log n]}\widetilde{G}(ne^{it/\sqrt{n}})\exp\left(n\left(e^{it/\sqrt{n}}-1-it/\sqrt{n}\right)\right)dt + \\ &+ \frac{1}{\sqrt{2\pi}}\int_{t\in[\log n,\theta\sqrt{n}]}\widetilde{G}(ne^{it/\sqrt{n}})\exp\left(n\left(e^{it/\sqrt{n}}-1-it/\sqrt{n}\right)\right)dt\ . \end{aligned}$$

To estimate the second integral I''_n we observe that $|h_n(t)|\le e^{-\mu t^2}$ for $t\in[-\theta\sqrt{n},\theta\sqrt{n}]$, where μ is a constant (in fact, $\mu=1/\pi^2$ suffices). Thus, by Lemma 1 and condition (I) we immediately show that $I''_n=O(n^\beta e^{-\mu\log^2 n/2})$ which decays faster than any polynomial.

Now, we estimate I'_n. Observe first that for $t\in[-\log n,\log n]$ we can write

$$\begin{aligned} h_n(t) &= \exp\left(n\left(e^{it/\sqrt{n}}-1-it/\sqrt{n}\right)\right) \\ &= e^{-t^2/2}\left(1-\frac{it^3}{6\sqrt{n}}+\frac{t^4}{24n}-\frac{it^6}{72n}+O\left(\frac{\log^9 n}{n\sqrt{n}}\right)\right) \end{aligned}$$

Furthermore, using (I) and Lemma 3 for $|z| > (1+\Omega)\xi$ and $z \in \mathcal{S}_{\theta'}$ for $\theta' < \theta$ we have $|\widetilde{G}'(z)| \le B_1|z|^{\beta-1}$ and $|\widetilde{G}''(z)| \le B_2|z|^{\beta-2}$, for some constants B_1 and B_2. Thus, we can expand $G(ne^{it/\sqrt{n}})$ around $t=0$ as $\widetilde{G}(ne^{it/\sqrt{n}}) = \widetilde{G}(n) + it\sqrt{n}\widetilde{G}'(n) + \Delta_n(t)t^2$ where $|\Delta_n(t)| \le (B_1+B_2)n^{\beta-1}$. In summary, the integral I'_n becomes

$$
\begin{aligned}
I'_n &= \frac{1}{\sqrt{2\pi}}\int_{-\log n}^{\log n} e^{-t^2/2}\left(\widetilde{G}(n) + \widetilde{G}'(n)it\sqrt{n}\right)\left(1 - \frac{it^3}{6\sqrt{n}} + \frac{t^4}{24n} - \frac{it^6}{72n}\right)dt + \\
&\quad + \frac{1}{\sqrt{2\pi}}\int_{-\log n}^{\log n} e^{-t^2/2}\Delta_n(t)t^2 h_n(t)dt + \\
&\quad + \frac{1}{\sqrt{2\pi}}\int_{-\log n}^{\log n} e^{-t^2/2}\left(\widetilde{G}(n) + \widetilde{G}'(n)it\sqrt{n}\right) O\left(\frac{\log^9 n}{n\sqrt{n}}\right) dt \ .
\end{aligned}
$$

To complete the proof we must estimate the above three integrals. From Lemmas 1 and 3 we assess that the first integral is equal to $G(n) + O(n^{\beta-1})$. Thus it suffices to show that the last two integrals are $O(n^{\beta-1})$. Indeed, using an estimate on $\Delta_n(t)$ we observe that the absolute value of second integral is smaller than $(B_1+B_2)n^{\beta-1}$. Finally, the last integral is $O(n^{\beta-\frac{3}{2}}\log^9 n)$. All together the error becomes $O(n^{\beta-1})$. Theorem 1 is proved. Our result (10) can be proved ina similar manner but one needs to use few more expansion terms in the above derivations. ■

As mentioned before, poissonization and depoissonization can be also used to study limiting distributions. In such a case, we should investigate double-indexed sequence $g_{n,k}$ (e.g., for a random variable X_n one can define $g_{n,k} = Pr\{X_n = k\}$) through its generating function $G_n(u) = \sum_{k\ge 0} g_{n,k}u^k$ and its Poisson transform $\widetilde{G}(z,u) = \sum_{n\ge 0} G_n(u)\frac{z^n}{n!}e^{-z}$. If u belongs to a compact set, then our previous depoissonization results can be directly applied. For example, we have the following two corollaries that are quite useful in some analysis (cf. [3, 16, 26]).

Corollary 1 *Suppose* $\widetilde{G}_k(z) = \sum_{n=0}^{\infty} g_{n,k}\frac{z^n}{n!}e^{-z}$*, for* k *belonging to some set* K*, are entire functions of* z*. Let for some constants* A*,* B*,* $R > 0$*,* β *and* $\alpha < 1$ *the following two conditions holds uniformly in* $k \in K$*:*

(I) *For* $z \in \mathcal{S}_\theta$

$$
|z| > R \quad \Rightarrow \quad |\widetilde{G}_k(z)| \le B|z|^\beta \ , \tag{12}
$$

(O) *For* $z \notin \mathcal{S}_\theta$

$$
|z| > R \quad \Rightarrow \quad |\widetilde{G}_k(z)e^z| \le A\exp(\alpha|z|) \ . \tag{13}
$$

Then, uniformly in $k \in K$

$$
g_{n,k} = \widetilde{G}_k(n) + O(n^{\beta-1}) \ . \tag{14}
$$

for large n*, and the error estimate does not depend on* K*.*

The next corollary is a direct consequence of Corollary 1 and it already found several applications in the analysis of algorithms 9cf. [26]). Indeed, setting $\widetilde{G}_u(z) = \widetilde{G}(z,u)$ in the above corollary we immediately obtain:

Corollary 2 *Let $\widetilde{G}(z,u)$ satisfy hypothesis of Theorem 1, i.e. for some numbers $\theta < \pi/2$, A, B, $\xi > 0$, β, and $\alpha < 1$ (I) and (O) hold for all u in a set $\mathcal{U}$. In particular, for $u \in \mathcal{U}$ condition (I) and (O) hold. Then,*

$$G_n(u) = \widetilde{G}(n,u) + O(n^{\beta-1})$$

uniformly in $u \in \mathcal{U}$.

3 Applications

In this section we discuss three applications of our depoissonization results: generalized digital search trees, leader election algorithm, and generalized probabilistic counting.

3.1 Generalized Digital Search Trees

Let us consider generalized digital search trees in which a node can hold up to b strings [6, 19, 20, 21] built from m independent strings over a binary alphabet $\Sigma = \{0,1\}$ with "0" is generated with probability p while "1" with probability $q = 1-p$. Let us investigate the total path length L_m which is the sum of depths of all nodes. The Poisson transform $\widetilde{X}(z)$ of EL_m satisfies the following functional-differential equation (cf. [20])

$$(1+\frac{\partial}{\partial z})^b \widetilde{X}(z) = z + \widetilde{X}(pz) + \widetilde{X}(qz) \tag{15}$$

where $(1+\frac{\partial}{\partial z})^b f(z) =^{def} \sum_{i=0}^{b} \binom{b}{i} \frac{\partial^i f(z)}{\partial z^i}$. To assess EL_m for large m one must depoissonize $\widetilde{X}(z)$. As in [20], we can use Mellin transform to prove that for large z inside a cone $\mathcal{S}_\theta$ we have $\widetilde{X}(z) = O(z \log z)$ which suffices to prove condition (I) in Theorem 1.

Thus, to apply Theorem 1, we need only to verify condition (O) outside the cone $\mathcal{S}_\theta$. Let $X(z) = \widetilde{X}(z)e^z$. Then, functional equation (15) transforms into $X^{(b)}(z) = X(zp)e^{zq} + X(zq)e^{zp} + ze^z$ where $X^{(b)}(z)$ denotes the bth derivative of $X(z)$. Observe that the above equation can be alternatively represented as

$$X(z) = \underbrace{\int_0^z \int_0^{w_2} \cdots \int_0^{w_b}}_{b\ times} (X(w_1 p))e^{w_1 q} + X(w_1 q)e^{w_1 p} + w_1 e^{w_1})\, dw_1 dw_2 \cdots dw_b \tag{16}$$

We must prove $|X(z)| \le e^{\alpha|z|}$ for $z \notin \mathcal{S}_\theta$ for $\alpha < 1$. We use induction over the so called *increasing domains* defined as follows (cf. [13, 21]): For a positive integer k and a constant $\xi > 0$ one denotes

$$\mathcal{F}_k = \{z = \rho e^{i\vartheta} :\ \rho \in [\xi\delta, \xi\nu^{-k}],\ \ 0 \le \vartheta < 2\pi\}$$

where $\max\{p,q\} \le \nu < 1$ and $\delta \le \min\{p,q\}$. The point to observe is that if $z \in \mathcal{F}_{k+1} - \mathcal{F}_k$ then $zp, zq \in \mathcal{F}_k$.

To carry out the induction, we first define $\bar{\mathcal{F}}_k = \mathcal{F}_k \cap \bar{S}_\theta$ where $\bar{S}_\theta$ denotes points in the complex plane outside $\mathcal{S}_\theta$. Since $X(z)$ is bounded for $z \in \bar{\mathcal{F}}_1$, thus the initial step of induction holds. Let us now assume that for $z \in \bar{\mathcal{F}}_k$ we have $|X(z)| \le e^{\alpha|z|}$ for some k and $\alpha < 1$. We intend to prove that $|X(z)| \le e^{\alpha|z|}$ for $z \in \bar{\mathcal{F}}_{k+1}$. Indeed, let $z \in \bar{\mathcal{F}}_{k+1}$. If also $z \in \mathcal{F}_k$, then the proof is completed. So let us now assume that $z \in \bar{\mathcal{F}}_{k+1} - \bar{\mathcal{F}}_k$. Since then $zp, zq \in \bar{\mathcal{F}}_k$, we can use our induction hypothesis together with the integral equation (16) to obtain the following estimate for $|z| > \xi$ where ξ is large enough

$$|X(z)| \le |z|^b \left(e^{|z|(p\alpha + q\cos\theta)} + e^{|z|(q\alpha + p\cos\theta)} + e^{|z|\cos\theta} \right) .$$

Let us now define $\alpha > \cos\theta$ such that the following three inequalities are simultaneously fulfilled

$$|z|^b e^{|z|(p\alpha + q\cos\theta)} \le \frac{1}{3} e^{\alpha|z|} \quad , \quad |z|^b e^{|z|(q\alpha + p\cos\theta)} \le \frac{1}{3} e^{\alpha|z|} \quad , \quad |z|^{b+1} e^{|z|\cos\theta} \le \frac{1}{3} e^{\alpha|z|} .$$

Then, for $z \in \bar{\mathcal{F}}_{k+1}$ we have $|X(z)| \le e^{\alpha|z|}$, as needed to verify condition (O) of the depoissonization lemma. This example is discussed in depth in [20].

3.2 Leader Election Algorithm

The following elimination process (cf. [3, 25]) has several applications, such as the election of a leader in a distributed or parallel system. A group of n people (users) wishes to identify a leader by tossing fair coins. All n people who throw heads are losers and do not participate any more in the election, but those who throw tails are candidate leaders and should flip their coins again. The process is repeated among candidate leaders until one leader is identified. If at any stage all remaining candidate leaders throw heads, the tosses are considered inconclusive and they all participate again as candidate leaders in the next round of coin tossing.

It is easy to notice that this election process can be represented by a special trie which was named by Prodinger [25] an *incomplete trie* since only one side of this tree is developed. Let D_n be the number of tosses until a leader is elected, that is, it is the path length in the incomplete trie from the root to the furthest terminal node on the left (assuming we move to the left when we toss a tail). If $G_n(u) = Eu^{D_n}$ denotes its probability generating function and $\widetilde{G}(z,u)$ its Poisson generating function, then the following functional equation can be derived (cf. [3, 25])

$$\widetilde{G}(z,u) = u(1 + e^{-z/2})\widetilde{G}\left(\frac{z}{2}, u\right) + e^{-z}[(1+z)(1-u) - ue^{z/2}] . \tag{17}$$

Our goal is to compute the limiting distribution of D_n, that is, $Pr\{D_n \le k\}$ for large n. Define $\widetilde{G}_k(z) = \sum_{n=0}^{\infty} Pr\{D_n \le k\} \frac{z^n}{n!} e^{-z}$. Fill, Mahmoud and Szpankowski [3] used Mellin transform approach to show that

$$\widetilde{G}_k(z) = e^{-z} + \left(1 - e^{-z}\right) \frac{z/2^k}{\exp(z/2^k) - 1} . \tag{18}$$

One needs to depoissonize this equation, and we apply Corollary 1. Observe that for $z \in \mathcal{S}_\theta$ we have $|\widetilde{G}_k(z)| \le 1 + 2\sec\theta = O(1)$, while outside the cone $\mathcal{S}_\theta$ we obtain

$|\widetilde{G}_k(z)e^z| \le \sec\theta e^{|z|\cos\theta}$. Thus, Corollary 1 can be used with $\alpha = \cos\theta$, and this leads to the following result

$$Pr\{D_n \le k\} = \frac{2^{\{\log_2 n\}-k}}{\exp\left(2^{\alpha(n)-k}\right) - 1} + O\left(\frac{1}{n}\right) \tag{19}$$

uniformly in all integers k, where $\{\log_2 n\} = \log_2 n - \lfloor \log_2 n \rfloor$. This improves the error bound from [3].

An interesting extension was recently considered by Janson and Szpankowski [16] who analyzed an asymmetric leader election process in which p is the probability of survival at each round ($p \neq 1/2$). The Poisson transform $\widetilde{G}(z,u)$ satisfies the following functional equation

$$\widetilde{G}(z,u) = u\widetilde{G}(pz,u) + u\widetilde{G}(qz,u)e^{-pz} + (1-u)ze^{-z} .$$

Asymptotic solution of this equation is much more intricate. We observe that the Poisson transform $\widetilde{G}_k(z)$ defined as above fulfills

$$\begin{aligned} \widetilde{G}_0(z) &= ze^{-z} \\ \widetilde{G}_{k+1}(z) &= \widetilde{G}_k(pz) + e^{-pz}\widetilde{G}_k(qz) , \qquad k \ge 1 \end{aligned}$$

This yields the following solution as proved in [16]

$$\widetilde{G}_k(z) = p^k z \int_0^{p^{-k}} e^{-p^k z t} d\mu(t) \tag{20}$$

where $\mu(t)$ is defined as follows: Partition the positive real axis into an infinite sequence of *consecutive* intervals $I_0, I_1, \ldots$ such that I_k has length $(q/p)^{s(k)}$, where $s(k)$ is the number of 1's in the binary expansion of k. Thus $I_0 = [0,1], I_1 = [1, 1+q/p], \ldots$. Note that the total length of the first 2^m intervals $(I_0, \ldots, I_{2^m-1})$ is p^{-m}, and that these 2^m intervals are obtained by repeated subdivisions of $[0, p^{-m}]$, each time dividing each interval in the proportions $p : q$. Given these intervals, define μ by putting a point mass $|I_k|$ at the right endpoint of I_k, for each $k = 0, 1, \ldots$

Finally, observing that verification of conditions (I) and (O) of Corollary 1 follows the same route as above, and after applying Corollary 1, we can prove:

$$Pr\{D_n \le k\} = F(p^k n) + O(1/n),$$

where

$$F(x) = x\int_0^\infty e^{-xt} d\mu(t) = \int_0^\infty e^{-t} d\mu_x(t).$$

Observe that, as in the symmetric case, the limiting distribution of D_n does not exist. But, the following can be claimed: When $k = \lfloor \log_P n \rfloor + \kappa$ where κ is an integer, then for large n uniformly over κ

$$Pr\{D_n \le \lfloor \log_P n \rfloor + \kappa\} = p^{\kappa - \{\log_P n\}} \int_0^\infty e^{-tp^{\kappa-\{\log_P n\}}} d\mu(t) + O\left(\frac{1}{n}\right),$$

where $\{\log_P n\} = \log_P n - \lfloor \log_P n \rfloor$, where $P = 1/p$.

3.3 Generalized Probabilistic Counting

In some applications (cf. [4]) one needs to estimate quickly the cardinality of a large set. A probabilistic counting is a solution. To estimate the cardinality n of a set (with replications) every element of the set is hashed into a binary string of size m. The bitwise OR-composition of modified hashed strings is used to build the so called the `bitmap` string and to obtain the estimate R_n of the cardinality n of a set. We consider a generalized probabilistic counting as in Kirschenhofer, Prodinger and Szpankowski [18] which can be describes as follows: Let us consider an empty `bitmap` string. Assume that n objects randomly insert a 1 at any position of the `bitmap`, and the probability of hitting the $j \geq 0$ position is equal to 2^{-j-1}. In terms of probabilistic counting, this means that the probability of the occurrence of the pattern like $0^j 1 \cdots$ is equal to 2^{-j-1} since 0 and 1 are equally likely (0^j denotes j consecutive zeros). We count the number of hits in any position of the `bitmap`, but we count the number of hits only up to some value $d+1$, where d is a given parameter. The parameter of interest is the length $R_{n,d}$ of the longest run of $d+1$'s in the front of the `bitmap`.

Let $G_n(u) = Eu^{R_{n,d}}$, and let $\widetilde{G}(z,u) = Eu^{R_{n,d}}$ be its Poisson generating function. As in [18] we observe that

$$\widetilde{G}(z,u) = uf_d(z/2)\widetilde{G}(z/2,u) + (u-1)(f_d(z/2)-1) ,$$

where $f_d(z) = 1 - e_d(z)e^{-z}$ and $e_d(z) = 1 + \frac{z^1}{1!} + \cdots + \frac{z^d}{d!}$ is the truncated exponential function.

Our goal is to estimate the limiting distribution through depoissonization. First of all, we consider $\widetilde{H}(z,u) = \widetilde{G}(z,u) - 1$ instead of $\widetilde{G}(z,u)$ which satisfies the following functional equation:

$$\widetilde{H}(z,u) = uf_d(z/2)\widetilde{H}(z/2,u) + (u-1)f_d(z/2) . \tag{21}$$

This is a linear functional equation, and by consecutive iterations we obtain

$$\frac{1-\widetilde{G}(z,u)}{1-u} = \sum_{k=0}^{\infty} u^k \frac{\varphi(z2^{-k-1})}{\varphi(z)} ,$$

which finally implies

$$\widetilde{G}_k(z) = [u^k]\left(\frac{\widetilde{H}(z,u)}{1-u}\right) = e^{-z}\sum_{N=0}^{\infty} Pr\{R_{n,d} > k\}\frac{z^N}{N!} = \frac{\varphi(z2^{-k-1})}{\varphi(z)} . \tag{22}$$

To evaluate the limiting distribution we apply Corollary 1 to $\widetilde{G}_k(z)$. We must verify conditions (I) and (O). The condition (O) outside the cone $\mathcal{S}_\theta$ where $\theta < \pi/2$ is easy to check (cf. [18]), and it follows directly from the functional equation. The difficulty is to prove a bound on $\widetilde{G}_k(z)$ inside the cone $\mathcal{S}_\theta$. We use a *complex* Mellin transform as discussed in [5, 14] to establish the following bound:

Lemma 4 *For all $z \in \mathcal{S}_\theta$ with $z \neq 0$ and $|\theta| < \pi/2$ we have uniformly*

$$|\varphi(z)| < A \tag{23}$$

where A is a constant. Also, for large z we have $\varphi(z) = 1 + O(z^{-M})$ for any $M > 0$.

Proof. First of all, we observe that for large z the series in $\varphi(z)$ converges exponentially fast, and one immediately proves that $\varphi(z) = 1 + O(z^{-M})$ for any $M > 0$. Thus, the only difficult part is to prove (23) for z close to zero, which we consider next. For simplicity of presentation we only assume $d = 0$. We use complex Mellin transform. Let $\ell(z) = \log \varphi(z)$. Observe that (for $d = 0$) $\ell(z) = \log(1 - e^{-z}) + \ell(2z)$. Hence, the Mellin $\ell^*(s)$ (cf. [5]) becomes $\ell^*(s) = -\frac{\Gamma(s)\zeta(s+1)}{(1-2^{-s})}$. There is a triple pole at $s = 0$ which gives $\ell(z) = -\frac{1}{2}(\log^2 z + \log z) + O(1)$ for $z \to 0$. Thus $\varphi(z) = \exp(\ell(z)) = \exp(-1/2 \log^2 z - 1/2 \log z + O(1))$. To show uniform bound around $z = 0$, let $z = \rho e^{i\theta}$. Then,

$$|\varphi(z)| = \exp(\Re(\ell(z)) = \exp(-0.5(\log^2 \rho - \theta^2 + \log \rho)) \ .$$

Observe that for $\rho < \rho_c$ where $\log^2 \rho_c - \pi^2/4 + \log \rho_c > 0$ the function $\varphi(z)$ is bound for ρ small, and this completes the proof. ■

Having the above result, we easily show that for large $z \in S_\theta$

$$\frac{|\varphi(z2^{-k-1})|}{|\varphi(z)|} \leq A \ .$$

which completes the proof of the depoissonization inequality (I). Thus, by Corollary 1 we prove that $Pr\{R_{n,d} > k\} = \frac{\varphi(n2^{-k-1})}{\varphi(n)} + O(1/n)$, or more precisely for any integer m

$$Pr\{R_{n,d} \leq \log_2 n + m - 1\} = 1 - \varphi\left(2^{-m-\{\log_2 n\}}\right) + O(n^{-1}) \ .$$

where $\{\log_2 n\} = \log_2 n - \lfloor \log_2 n \rfloor$. This is a result from [18] but with much improved error bound.

References

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

[2] N. G. de Bruijn, *Asymptotic Methods in Analysis*, Dover, 1962.

[3] J. Fill. H. Mahmoud, and W. Szpankowski, On the Distribution for the Duration of a Randomized Leader Election Algorithm, *Annals of Applied Probability*, 6, 1260–1283, 1996.

[4] P. Flajolet and G.N. Martin, Probabilistic Counting Algorithms for Data Base Applications, *J. Comp. Syst. Sci.*, 31, 182-209, 1985.

[5] P. Flajolet, X. Gourdon, P. Dumas, Mellin Transforms and Asymptotics: Harmonic Sums, *Theoretical Computer Science*, 144, 3-58, 1995.

[6] P. Flajolet and B. Richmond, Generalized Digital Trees and Their Difference-Differential Equations, *Random Structures & Algorithms*, 3, 305-320 (1992).

[7] G.H. Gonnet Expected Length of the Longest Probe Sequence in Hash Code Searching *J. of the ACM*, 28, 289-304, 1981.

[8] G. Gonnet and J. Munro, The Analysis of Linear Probing Sort by the Use of a New Mathematical Transform, *Journal of Algorithms*, 5, 451-470, 1984.

[9] L. Holst, On Birthday, Collector's, Occupancy and Other Classical Urn Problems, *International Statistical Review*, 54, 15-27, 1986.

[10] P. Jacquet and M. Régnier, Limiting Distributions for Trie Parameters, *Lecture Notes in Computer Science*, 214, 196-210, 1986.

[11] P. Jacquet and M. Régnier, Normal Limiting Distribution of the Size of Tries, *Proc. Performance'87*, 209-223, North Holland, Amsterdam 1987

[12] P. Jacquet and W. Szpankowski, Ultimate Characterization of the Burst Response of an Interval Searching Algorithm: A Study of a Functional Equation, *SIAM J. Comp.*, 18, 777-791, 1989.

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

[14] P. Jacquet and W. Szpankowski, Analytical Depoissonization and Its Applications, *Theoretical Computer Science*, to appear.

[15] P. Jacquet and W. Szpankowski, Towards Analytical Information Theory: Entropy Computations, Purdue University, CSD-TR-96-085, 1996.

[16] S. Janson and W. Szpankowski, Analysis of an Asymmetric Leader Election Algorithm, Purdue University CSD-TR-96-049, 1996.

[17] M. Kac, On the Deviations between Theoretical and Empirical Distributions, *Proc. Natl. Acad. Sci. USA*, 35, 252-257, 1949.

[18] P. Kirschenhofer, H. Prodinger and W. Szpankowski, Analysis of a Splitting Process Arising in Probabilistic Counting and Other Related Algorithms, *Random Structures & Algorithms*, 9, 379–402, 1996.

[19] D. Knuth, *The Art of Computer Programming. Sorting and Searching,* Vo. 3., Addison-Wesley 1973.

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

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

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

[23] P. V., Poblete, Approximating Functions by Their Poisson Transform, *Information Processing Letters*, 23, 127-130, 1986.

[24] P. Poblete, A. Viola and J.I. Munro, The Diagonal Poisson Transform and its Application to the Analysis of a Hashing Scheme, *Random Structures & Algorithms*, 10, 1997.

[25] H. Prodinger, How To Select a Loser. *Discrete Math.*, 120, 149–159, 1993.

[26] B. Rais, P. Jacquet and W. Szpankowski, A Limiting Distribution for the Depth in PATRICIA Tries, *SIAM J. Discrete Mathematics*, 6, 197-213 (1993).

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