

On the asymptotic behavior of some Algorithms

Philippe Robert

► **To cite this version:**

Philippe Robert. On the asymptotic behavior of some Algorithms. [Research Report] RR-5421, INRIA. 2005, pp.19. inria-00070585

HAL Id: inria-00070585

<https://hal.inria.fr/inria-00070585>

Submitted on 19 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

On the asymptotic behavior of some Algorithms

Philippe Robert

N° 5421

Deembre 2004

_____ Thème COM _____



*R*apport
de recherche



On the asymptotic behavior of some Algorithms

Philippe Robert

Thème COM — Systèmes communicants
Projet Rap

Rapport de recherche n° 5421 — Deembre 2004 — 19 pages

Abstract: A simple approach is presented to study the asymptotic behavior of some algorithms with an underlying tree structure. It is shown that some asymptotic oscillating behaviors can be precisely analyzed without resorting to complex analysis techniques as it is usually done in this context. A new explicit representation of periodic functions involved is obtained at the same time.

Key-words: Splitting Algorithms. Asymptotic Oscillating Behavior. Poisson Transform. Mellin Transform. Ethernet. Tree Communication Protocol

Sur l'analyse asymptotique de certains algorithmes

Résumé : Le comportement asymptotique plusieurs suites réelles oscillantes est analysée. Cette approche donne une approche unifiée de toute une classe résultats.

Mots-clés : Algorithmes diviser pour régner. Suites oscillantes

1 Introduction

Algorithms with an underlying tree structure are quite common in computer science, they are used to sort, store and search... See Cormen *et al.* [2] and Knuth [15]. Splitting algorithms are examples of such algorithms, they can be described as follows

SPLITTING ALGORITHM $\mathcal{S}(n)$

- TERMINATION CONDITION.
For some subset F of \mathbb{N} , if $n \in F$
→ STOP.
 - TREE STRUCTURE.
Randomly divide n into n_1, \dots, n_d , with $n_1 + \dots + n_d = n$.
→ APPLY $\mathcal{S}(n_1), \mathcal{S}(n_2), \dots, \mathcal{S}(n_d)$.
-

To the algorithm $\mathcal{S}(n)$ is attached a *cost* R_n which can be the number of steps required to terminate for example. This cost is assumed to be *additive* that is, with the above notations, the relation

$$R_n = 1 + R_{n_1} + R_{n_2} + \dots + R_{n_d}, \quad (1)$$

holds. Note that R_n is a *random variable* since the algorithm divides randomly into subgroups. This quantity can be thought as the cost of processing n items. If $\mathbb{E}(R_n)$ is its expected value, $\mathbb{E}(R_n)/n$ is the average processing time of one item among n . From a probabilistic point of view, it is natural to expect that the sequence (R_n) satisfies a kind of law of large numbers, i.e. that $(\mathbb{E}(R_n)/n)$ converges to some quantity α . The constant α would be, in some sense, the asymptotic average processing time of an item. In the context of a communication network, R_n is the total transmission time of n initial messages, $1/\alpha$ would be the asymptotic throughput of the protocol.

Curiously, this law of large numbers does not always hold. In some situations, the sequence $(\mathbb{E}(R_n)/n)$ does not converge at all and, moreover, exhibits an oscillating behavior. Up to now, these phenomena have been mostly analyzed by using sophisticated complex analysis techniques (via functional transforms) by Knuth [15], Flajolet *et al.* [9] and many others. See Mahmoud [17] for a comprehensive treatment of this approach. For alternative methods using real analysis on related problems, see the nice paper by Delange [5] and also Pippenger [20].

In this paper, a direct, simple, approach is proposed to study these uncommon laws of large numbers. At the same time, it sheds a new light on the oscillating phenomena involved. First, the classical approach, i.e. with complex analysis, is briefly recalled.

Analytic approach: An excursion inside the complex plane

A classical method to derive asymptotics of a sequence (r_n) related to a splitting algorithm consists in taking successive transforms:

$$(r_n) \xrightarrow{\text{Poisson}} r(x) = \sum_{n \geq 0} r_n \frac{x^n}{n!} e^{-x} \xrightarrow{\text{Mellin}} r^*(s) = \int_0^{+\infty} r(x) x^{s-1} dx. \quad (2)$$

The Poisson transform step may be, sometimes, skipped:

$$(r_n) \xrightarrow{\text{Mellin}} r^*(s) = \sum_{n \geq 1} r_n \frac{1}{n^s}. \quad (3)$$

In this way, Equation (1) can be translated, via a Poisson transform, into a functional equation of the form

$$r(x) = r(\beta x) + h(x), \quad x \geq 0, \quad (4)$$

where $0 < \beta < 1$ and h is some fixed function. Provided that an iteration scheme is valid, the problem is then to get an asymptotic expansion of the series

$$r(x) = r(0) + \sum_{n \geq 0} h(\beta^n x)$$

as x goes to infinity. This is usually done by taking the Mellin transform of the function $x \rightarrow r(x)$.

Mellin Transform

The Mellin transform r^* is defined in a vertical strip of \mathbb{C} and, under some growth conditions, the asymptotic behavior of r_n [resp. $r(x)$], as n [resp. x] gets large, can be expressed by using the poles of r^* on the right of the strip (provided that some growth conditions of the Mellin transform are satisfied). See Flajolet *et al.* [9]. When the Poisson transform is used, the next step is to justify that the behavior of $r(x)$ as x gets large is similar to the behavior of r_n for n large.

Poisson Transform

If $(N([0, u]), u \geq 0)$ is a Poisson process with parameter 1 (see the Appendix for a quick presentation): for $u \geq 0$, the variable $N([0, x])$ has a Poisson distribution with parameter x . The function $x \rightarrow r(x)$ can therefore be expressed as

$$r(x) = \mathbb{E}(r_{N([0, x])}),$$

for $x \geq 0$. The law of large numbers for Poisson processes states that, almost surely, $N([0, x])/x \sim 1$ as x gets large, this suggests that $r_{N([0, x])} \sim r_{[x]}$ provided that the sequence

(r_n) does not vary too much: Recall that, due to the central limit theorem, the approximation $N([0, x]) \sim x$ is valid with an error of the order of \sqrt{x} . Analytically, some conditions on the sequence (r_n) can be formulated so that an equivalence between the asymptotic behaviors of the sequence and of the Poisson transform can be established. See Jacquet and Szpankowski [12] for example.

A direct method

The approach presented here relies on a convenient use of Fubini's Theorem combined, sometimes, with some elementary properties of Poisson processes. The purpose of the paper is to show that some asymptotic results for algorithms on trees can be obtained in an elementary way. To show the effectiveness of the method, expansions with oscillating behaviors, which are usually analyzed with quite technical results of complex analysis, are obtained with this approach.

With the analytical approach, oscillating expansions are described with a periodic function which shows up through its Fourier coefficients. It occurs when the Mellin transform of the sequence has poles on an imaginary axis at the points $(a + inb, n \in \mathbb{Z})$. The method presented here has the advantage of being more direct and to give an explicit expression of the mysterious periodic function which, in the end, is not mysterious at all.

Rather than setting up a framework with formal results, the presentation of some important and interesting algorithms already analyzed in the literature has been chosen. Section 2 studies one of the first algorithms for which a curious oscillating behavior has been proved (by Knuth). Section 3 considers the basic algorithm of the Ethernet protocol. It is not, strictly speaking, a law of large numbers setting but a similar oscillating behavior occurs also in this case. Moreover, it is not only true for the averages but also for the distribution of the variables. Section 4 gives other examples where similar methods can be used. In particular, a treatment of Equation (4) is proposed. The examples of Section 3 and Section 4 show also that the method does not give only a first order term of the asymptotic expansion (i.e. at the level of the law of large numbers) but can also give subsequent terms of the expansion. A more complicated splitting algorithm is analyzed in Mohamed and Robert [19].

Acknowledgements

The author is grateful to Philippe Flajolet for uncountable interesting conversations on tree algorithms and Mellin transforms. The paper also benefited from two anonymous referees' comments.

2 A Binary Splitting algorithm

This section analyzes an algorithm investigated by Knuth in 1973, see Knuth [15] page 131-132. It consists in splitting randomly and recursively a group of n initial items into subgroups until each of the subgroups has cardinality 0 or 1. Latter, it has been used by Capetanakis [3]

and Tsybakov and Mikhailov [22] in the design of algorithms for access protocols to communication channels. See also the surveys Flajolet and Jacquet [10] and Ephremides and Hajek [7].

BINARY SPLITTING ALGORITHM $\mathcal{K}(n)$

- $n = 0$ or $1 \longrightarrow$ STOP.
 - $n \geq 2$. The n elements are randomly, equally, divided into two subgroups. If n_1 and $n - n_1$ denote the cardinalities of these subgroups then
 \longrightarrow APPLY $\mathcal{K}(n_1)$ and $\mathcal{K}(n - n_1)$.
-

The cost R_n of the algorithm $\mathcal{K}(n)$ starting with n elements is defined as the number of steps it requires to stop, in particular, $R_0 = R_1 = 1$. The following proposition establishes a recurrence relation for the other values of n . It is simply a formal rephrasing of the description of the algorithm.

Proposition 1 (Stochastic Recurrence Relation). *For $n \geq 2$, the random variable R_n satisfies the relation*

$$R_n \stackrel{\text{dist.}}{=} 1 + R_{n-S_n}^1 + R_{S_n}^2, \quad n \geq 2,$$

with $S_n = B_1 + \dots + B_n$ and

- the i.i.d. variables B_i , $i \geq 1$ are Bernoulli with parameter $1/2$;
- for $0 \leq k \leq n$, the variables R_k^1 and R_{n-k}^2 are independent and R_k^1 and R_k^2 have the same distribution as R_k .

The Bernoulli variables have the following interpretation: For $i \geq 1$, if $B_i = 0$ [resp. $B_i = 1$] the i th element goes in the first [resp. second] subgroup.

The recurrence relation for (R_n) and the boundary conditions at $n = 0$ and 1 can be integrated in the following way,

$$R_n \stackrel{\text{dist.}}{=} 1 + R_{n-S_n}^1 + R_{S_n}^2 - 2_{\{n \leq 1\}}. \quad (5)$$

The rest of the section is devoted to the analysis of the asymptotic behavior of the average value $\mathbb{E}(R_n)$ of the random variable R_n when n is large.

Proposition 2 (Poisson Transform). *For $x > 0$, the Poisson transform $r(x)$ of the sequence $(\mathbb{E}(R_n))$ is given by the series*

$$r(x) = \sum_{n=0}^{+\infty} \mathbb{E}(R_n) \frac{x^n}{n!} e^{-x} = 1 + 2 \sum_{k \geq 0} 2^k \mathbb{P}(t_2 \leq x/2^k), \quad (6)$$

where t_2 is the sum of two independent exponentially distributed random variables with parameter 1 .

Proof. If $\mathcal{N} = (N([0, t]))$, (t_n) is a Poisson process with intensity 1. See the appendix. The Poisson transform $r(x)$ can be expressed as

$$r(x) = \mathbb{E} (R_{N([0, x])}) ,$$

For $n \geq 0$, S_n denotes the n th partial sum of Bernoulli random variables with parameter $1/2$. Relation (5) gives the identity

$$R_{N([0, x])} \stackrel{\text{dist.}}{=} 1 + R_{S_{N([0, x])}}^1 + R_{N([0, x]) - S_{N([0, x])}}^2 - 2_{\{N([0, x]) \leq 1\}}.$$

The Poisson variable random variable $N([0, x])$ is split into two random variables $S_{N([0, x])}$ and $N([0, x]) - S_{N([0, x])}$. According to Proposition 10 of the appendix, the distribution of these two (independent) random variables is Poisson with parameter $x/2$, i.e. the distribution of $N([0, x/2])$. Since, by definition, t_2 is the second point of the Poisson process then $\{N([0, x]) \leq 1\} = \{t_2 > x\}$ and

$$\mathbb{E} (R_{S_{N([0, x])}}) = 2\mathbb{E} (R_{S_{N([0, x/2])}}) + 1 - 2\mathbb{P}(t_2 > x) .$$

If $f(x) = r(x) - 1$, the last identity can be written as $f(x) = 2f(x/2) + 2\mathbb{P}(t_2 \leq x)$. By iterating, one gets the expansion

$$f(x) = 2^n f(x/2^n) + 2 \sum_{k=0}^{n-1} 2^k \mathbb{P}(t_2 \leq x/2^k).$$

The limit of the sequence $(2^n f(x/2^n)/x)$ is the limit of $(r(h) - 1)/h$ as h goes to 0:

$$\lim_{n \rightarrow +\infty} 2^n f(x/2^n)/x = r'(0) = \mathbb{E}(R_0) - \mathbb{E}(R_1) = 0.$$

Representation (6) of the Poisson transform is thus established. □

The following proposition establishes a useful integral representation of the quantity $\mathbb{E}(R_n)$ which gives the key of the asymptotic expansion. Its proof uses a probabilistic trick of independent interest to invert the Poisson transform.

Proposition 3 (Probabilistic de-Poissonization). *For $n \geq 2$,*

$$\mathbb{E}(R_n) = 4n \int_0^1 2^{-\{-\log_2(x)\}} (n-1) (1-x)^{n-2} dx - 1, \tag{7}$$

with $\{y\} = y - \lfloor y \rfloor$, the fractional part of $y \in \mathbb{R}$.

Proof. As in the proof of Proposition 2, $\mathcal{N} = (N([0, x])) = (t_n)$ denotes a Poisson point process with intensity 1. By using Equation (6) and by decomposing according to the

number of t_n 's in the interval $[0, x]$, one gets

$$\begin{aligned} r(x) &= 1 + 2 \sum_{k \geq 0} 2^k \mathbb{P}(t_2 \leq x/2^k) \\ &= 1 + 2 \sum_{k \geq 0} 2^k \sum_{n \geq 2} \mathbb{P}(t_2 \leq x/2^k, N([0, x]) = n) \\ &= 1 + 2 \sum_{k \geq 0} 2^k \sum_{n \geq 2} \mathbb{P}(t_2 \leq x/2^k \mid N([0, x]) = n) \frac{x^n}{n!} e^{-x}. \end{aligned}$$

According to Proposition 12, conditionally on the event $\{N([0, x]) = n\}$, the variable (t_1, t_2, \dots, t_n) has the same distribution as $(xU_{(1)}^n, xU_{(2)}^n, \dots, xU_{(n)}^n)$, where $(U_i, 1 \leq i \leq n)$ are independent random variables uniformly distributed on $[0, 1]$ and $(U_{(i)}^n, 1 \leq i \leq n)$ denotes their non-decreasing reordering. In particular, conditionally on the event $\{N([0, x]) = n\}$, the variable t_2 has the same distribution as $xU_{(2)}^n$, hence the quantity

$$\mathbb{P}(t_2 \leq x/2^k \mid N([0, x]) = n) = \mathbb{P}(xU_{2,n} \leq x/2^k) = \mathbb{P}(U_{2,n} \leq 1/2^k)$$

does not depend on x . The Poisson transform can thus be written as

$$r(x) = 1 + 2 \sum_{k \geq 0} 2^k \sum_{n \geq 2} \mathbb{P}(U_{2,n} \leq 1/2^k) \frac{x^n}{n!} e^{-x} \quad (8)$$

Now, by switching the two sums, one gets

$$\begin{aligned} r(x) &= \sum_{n \geq 0} \frac{x^n}{n!} e^{-x} + 2 \sum_{n \geq 2} \left(\sum_{k \geq 0} 2^k \mathbb{P}(U_{2,n} \leq 1/2^k) \right) \frac{x^n}{n!} e^{-x} \\ &= e^{-x} + xe^{-x} + \sum_{n \geq 2} \left(1 + 2 \sum_{k \geq 0} 2^k \mathbb{P}(U_{2,n} \leq 1/2^k) \right) \frac{x^n}{n!} e^{-x}. \end{aligned}$$

The last expression is clearly a Poisson transform, by identifying the coefficients, this yields the following formula, for $n \geq 2$,

$$r_n = 1 + 2 \sum_{k \geq 0} 2^k \mathbb{P}\left(U_{2,n} \leq \frac{1}{2^k}\right) = 1 + 2 \sum_{k \geq 0} 2^k \mathbb{E}\left(\mathbb{1}_{\{U_{2,n} \leq 1/2^k\}}\right).$$

By Fubini's Theorem, the sum and the expectation can also be switched (the summands are non-negative), therefore,

$$r_n = 1 + 2 \mathbb{E}\left(\sum_{k \geq 0} 2^k \mathbb{1}_{\{k \leq -\log_2(U_{2,n})\}}\right) = 1 + 2 \left(\mathbb{E}\left(2^{\lceil -\log_2(U_{2,n}) \rceil}\right) - 1\right).$$

Since, for $x > 0$, $\mathbb{P}(U_{2,n} \leq x) = 1 - (1-x)^n - nx(1-x)^{n-1}$, the density function of the variable $U_{2,n}$ is the function $x \rightarrow n(n-1)x(1-x)^{n-2}$ on $[0, 1]$. This gives the relation

$$\begin{aligned} r_n &= 4n \int_0^1 2^{\lfloor -\log_2(x) \rfloor} (n-1)x(1-x)^{n-2} dx - 1 \\ &= 4n \int_0^1 2^{-\{-\log_2(x)\}} (n-1)(1-x)^{n-2} dx - 1. \end{aligned}$$

The proposition is proved. \square

With a similar use of Fubini's Theorem as in the above proof, one gets the following representation of the Poisson transform of (R_n) .

Proposition 4. *The Poisson transform of the sequence (R_n) can be represented as*

$$r(x) = 4x \int_0^x 2^{-\{\log_2(x) - \log_2(y)\}} e^{-y} dy + 2(1+x)e^{-x} - 1, \quad x \geq 0. \quad (9)$$

If Equation (9) is more explicit than Equation (6), it has nevertheless no use in the asymptotic analysis, Equation (6) is the key identity.

Proof. By Equation (6) and Fubini's Theorem, one gets

$$\begin{aligned} \frac{1}{2}(r(x) - 1) &= \sum_{k \geq 0} 2^k \mathbb{E}(\mathbb{1}_{\{t_2 \leq x/2^k\}}) = \mathbb{E}\left(\sum_{k \geq 0} 2^k \mathbb{1}_{\{t_2 \leq x/2^k\}}\right) \\ &= \mathbb{E}\left[\left(2^{\lfloor \log_2(x/t_2) \rfloor + 1} - 1\right) \mathbb{1}_{\{t_2 \leq x\}}\right], \end{aligned}$$

the proof is concluded by trite calculations and by using the fact that the random variable t_2 has the density $(x \exp(-x))$ on \mathbb{R}_+ . \square

Knuth's result on the asymptotic behavior of this algorithm is given by the following theorem.

Theorem 5 (Asymptotic oscillations). *The average of R_n satisfies the expansion*

$$\mathbb{E}(R_n) = nF(\log_2(n)) - 1 + O(ne^{-n}), \quad n \geq 1, \quad (10)$$

with

$$F(y) = 4 \int_0^{+\infty} 2^{-\{y - \log_2(x)\}} e^{-x} dx. \quad (11)$$

The function F is periodic with period 1, in particular the sequence $(\mathbb{E}(R_n)/n)$ is not converging.

In communication networks, the quantity $\mathbb{E}(R_n)/n$ is related to an average transmission time when n messages are initially in the network. The practical consequence of the oscillations of the sequence $(\mathbb{E}(R_n)/n)$ is nevertheless quite limited since the amplitude of the oscillations of the function F is of the order of 10^{-5} . See Figure 1.

Proof. By Equation (7), it is enough to look at the asymptotic behavior of

$$\int_0^1 2^{-\{-\log_2(x)\}} n (1-x)^{n-1} dx = \int_0^n 2^{-\{\log_2(n)-\log_2(x)\}} \left(1 - \frac{x}{n}\right)^{n-1} dx.$$

The elementary inequality

$$\begin{aligned} \left| \int_0^1 2^{-\{-\log_2(x)\}} n (1-x)^{n-1} dx - \int_0^{+\infty} 2^{-\{\log_2(n)-\log_2(x)\}} e^{-x} dx \right| \\ \leq \int_n^{+\infty} e^{-x} dx + \int_0^n \left| \left(1 - \frac{x}{n}\right)^{n-1} - e^{-x} \right| dx = 2e^{-n} \end{aligned}$$

concludes the proof. \square

Remark

By evaluating the integral of Equation (7), one gets that the quantity $\mathbb{E}(R_n)$ can be expressed as

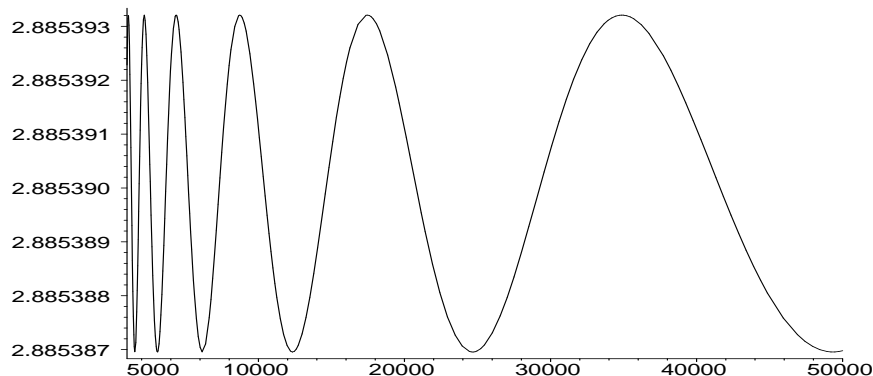
$$\mathbb{E}(R_n) = 1 + 2 \sum_{k \geq 0} 2^k \left(1 - \left(1 - \frac{1}{2^k}\right)^n - \frac{n}{2^k} \left(1 - \frac{1}{2^k}\right)^{n-1} \right).$$

This is the starting point of most of analyses of this algorithm. It is followed by some exponential approximations, a Mellin transform of the residual series and finally some complex analysis arguments to derive the asymptotic behavior of the original sequence. The periodic function F appears then through its Fourier transform. By inversion, it is expressed in term of the values of the Gamma function Γ on a vertical axis of the complex plane,

$$F(x) = -\frac{2}{\log 2} \sum_{k \in \mathbb{Z} - \{0\}} \xi_k \Gamma(\xi_k - 1) e^{2ik\pi x}$$

with $\xi_k = 2ik\pi/\log 2$ for $k \in \mathbb{Z}$.

Clearly, it is much easier to evaluate the asymptotic behavior of the integral (7). Moreover, as a benefit, the periodic function F shows up quite naturally and with a direct explicit expression which is apparently new.

Figure 1: The interpolated sequence $n \rightarrow \mathbb{E}(R_n)/n$

3 Ethernet

The context of the algorithm analyzed in this section is the following: In a communication network with only one channel, at most one message can be transmitted each unit of time. If two transmitters (or stations) try at the same time, this is a failure (a collision) and both of them have to retransmit later. In an Ethernet network, to each station waiting for transmission is associated an integer L which is the number of collisions it has already experienced. At time $t + 1$, a station with a counter equal to k tries to transmit with probability $1/2^k$. See Metcalf and Boggs [18] and Aldous [1]. For $t \geq 0$, $L(t)$ denotes the value of the counter of a given station which is waiting for transmission for t units of time. In a completely congested network, the sequence $(L(t))$ evolves as follows: $L(0) = 1$ and

$$L(t+1) = \begin{cases} L(t) & \text{with probability } 1 - a^{L(t)} \\ L(t) + 1 & a^{L(t)}, \end{cases}$$

where $a \in (0, 1)$. For Ethernet the value of a is $1/2$. An approximated counting algorithm proposed by Flajolet and Martin [11] uses also such a sequence $(L(t))$, see Flajolet [8].

For $k \geq 1$, G_k denotes the sojourn time of $(L(t))$ in state k , G_k is a geometrically distributed with parameter a^k :

$$\mathbb{P}(G_k \geq n) = (1 - a^k)^n.$$

Consequently, $G_1 + \dots + G_{k-1}$ is the hitting time of k for the process $(L(t))$. Note that the random variables (G_k) are, of course, independent.

The average hitting time of k is therefore a^{-k} , this suggests that the value of $L(t)$ is of the order of $\log_{1/a}(t)$ when t is large. This approximation is indeed true : it is not difficult to show that the quantity $|\mathbb{E}(L(t)) - \log_{1/a}(t)|$ is bounded with respect to t . This estimation

suggests that there should be a convergence in distribution of the random variable $L(t) - \log_{1/a}(t)$ as t goes to infinity. Surprisingly, this convergence does not hold. With calculus on some alternating series, Mellin transforms and complex analysis methods, Flajolet [8] has shown that the variable $L(t) - \log_{1/a}(t)$ exhibits an asymptotic oscillating behavior with respect to t . For an extension of these results, see also Kirschenhofer *et al.* [14].

In this section, a simple proof of this result is presented, together with an explicit description of the asymptotic behavior of the distribution of the random variable $L(t) - \log_{1/a}(t)$. The following elementary proposition simplifies a lot the analysis of the algorithm.

Proposition 6. *The convergence in distribution*

$$a^n \sum_{k=0}^{n-1} G_{n-k} \longrightarrow H \stackrel{\text{dist.}}{=} \sum_{k=0}^{+\infty} a^k E_k \quad (12)$$

holds as n goes to infinity, where (E_n) is a sequence of independent random variables exponentially distributed with parameter 1. The random variable H has a density h on \mathbb{R}_+ given by

$$h(x) = \frac{1}{\prod_{k=1}^{+\infty} (1 - a^k)} \sum_{n=0}^{+\infty} \frac{1}{\prod_{k=1}^n (1 - a^{-k})} a^{-n} \exp(-a^{-n}x), \quad x \geq 0. \quad (13)$$

Moreover, Convergence (12) is true for the total variation norm, i.e.

$$\lim_{n \rightarrow +\infty} \sup_{x \geq 0} \left| \mathbb{P} \left(\sum_{k=0}^{n-1} a^n G_{n-k} \geq x \right) - \mathbb{P}(H \geq x) \right| = 0. \quad (14)$$

The random variable H is already known in the domain of communication networks, but in a very different framework: The congestion avoidance phase of the Transmission Control Protocol (TCP) of the Internet. In this case, H is related to the stationary distribution of the throughput of a long TCP connection. See Dumas *et al.* [6]. The variable H also appeared in mathematical finance models to describe Asian Options, see Carmona *et al.* [4].

Proof. For $k \geq 0$ and $x > 0$, then

$$\begin{aligned} \mathbb{P}(a^n G_{n-k} \geq x) &= (1 - a^{n-k})^{\lceil x/a^n \rceil} \\ &= \exp\left(\lceil x/a^n \rceil \log(1 - a^{n-k})\right) \sim \exp\left(-x/a^k\right), \end{aligned}$$

hence the random variable $a^n G_{n-k}$ converges in distribution to $a^k E_k$ where E_k is exponentially distributed with parameter 1. Therefore, by independence of the variables (G_k) , the convergence (12) holds. For the explicit expression of the density of H , see the proof of Proposition 13 of Dumas *et al.* [6].

Chebychev's Inequality gives, for $x > 0$,

$$\mathbb{P} \left(a^n \sum_{k=0}^{n-1} G_{n-k} \geq x \right) \leq \frac{1}{x} \mathbb{E} \left(a^n \sum_{k=0}^{n-1} G_{n-k} \right) \leq \frac{1}{x} \frac{a^2}{1-a},$$

hence, the uniform convergence (14) has only to be proved in a compact interval. Since the sequence $n \rightarrow \mathbb{P}(a^n G_{n-k} \geq x)$ is non-decreasing, the same property holds for the function f_n defined as follows

$$n \rightarrow f_n(x) \stackrel{\text{def.}}{=} \mathbb{P}\left(a^n \sum_{k=0}^{n-1} G_{n-k} \geq x\right).$$

By Dini's Theorem, the non-decreasing sequence of functions (f_n) converges *uniformly* on compact sets to $x \rightarrow \mathbb{P}(H \geq x)$. The uniform convergence (14) is established . \square

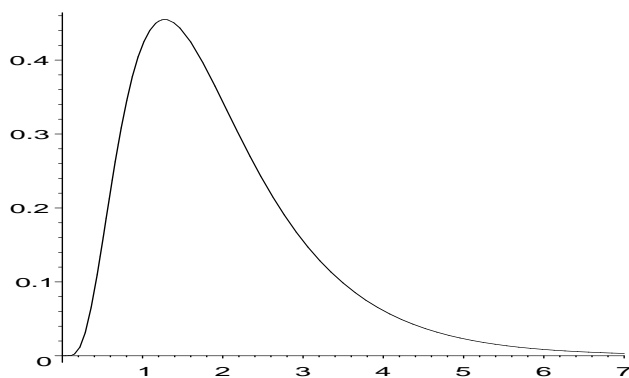


Figure 2: The density function of H when $a = 1/2$

For $t \geq 1$ and $u \in [0, 1]$, it is easy to check that

$$\mathbb{E}\left(u^{L(t)}\right) = \mathbb{E}\left(\sum_{n \geq 0} u^n (1-u) 1_{\{L(t) \leq n\}}\right) = \sum_{n \geq 0} u^n (1-u) \mathbb{P}(L(t) \leq n)$$

holds and, since $\mathbb{P}(L(t) \leq n) = \mathbb{P}(G_1 + \dots + G_n \geq t)$, for $n \geq 1$. One thus gets the relation

$$\mathbb{E}[\exp(-\lambda L(t))] = \sum_{n \geq 0} e^{-\lambda n} (1 - e^{-\lambda}) \mathbb{P}\left(a^n \sum_{i=0}^{n-1} G_{n-i} \geq a^n t\right), \quad \lambda > 0.$$

The uniform convergence (14) gives the expansion, as t goes to infinity,

$$\mathbb{E}[\exp(-\lambda L(t))] = \sum_{n \geq 0} e^{-\lambda n} (1 - e^{-\lambda}) \mathbb{P}(H \geq a^n t) + o(1),$$

by switching the sum and the expectation, one gets

$$\begin{aligned}
\mathbb{E}[\exp(-\lambda L(t))] &\sim \mathbb{E} \left(\sum_{n \geq 0} e^{-\lambda n} (1 - e^{-\lambda}) \mathbf{1}_{\{H \geq a^n t\}} \right) \\
&= \mathbb{E} \left(\sum_{n \geq \lceil \log_{1/a}(t/H) \rceil} e^{-\lambda n} (1 - e^{-\lambda}) \mathbf{1}_{\{t \geq H\}} \right) \\
&= \mathbb{E} \left[\exp \left(-\lambda \lceil \log_{1/a}(t/H) \rceil \right) \mathbf{1}_{\{t \geq H\}} \right] \sim \mathbb{E} \left[\exp \left(-\lambda \lceil \log_{1/a}(t/H) \rceil \right) \right].
\end{aligned} \tag{15}$$

The following equivalence of Laplace transforms has therefore been obtained:

$$\begin{aligned}
\mathbb{E} \left(\exp \left[-\lambda \left(L(t) - \log_{1/a}(t) \right) \right] \right) &\sim \mathbb{E} \left(\exp \left[-\lambda \left(\lceil \log_{1/a}(t/H) \rceil - \log_{1/a}(t) \right) \right] \right) \\
&= \mathbb{E} \left(\exp \left[-\lambda \left(1 - \log_{1/a}(H) - \{ \log_{1/a}(t/H) \} \right) \right] \right).
\end{aligned}$$

The above relation implies that if $x \geq 0$ and $Z(t) = L(t) - \log_{1/a}(t)$ then the sequence $(Z(a^{-n-x}))$ converges in distribution. The main result of Flajolet [8] can now be stated.

Theorem 7 (Asymptotic Oscillating Distribution). *Asymptotically, the distribution of $L(t) - \log_{1/a}(t)$ is equivalent to the distribution of $F(\log_{1/a}(t))$ where, for $x \geq 0$, $F(x)$ is defined as*

$$F(x) = \log_{1/a} \left(\frac{1}{aH} \right) - \left\{ x - \log_{1/a}(H) \right\}, \quad x \in \mathbb{R},$$

where H is the random variable introduced in Proposition 6.

Note that the random function $x \rightarrow F(x)$ is periodic with period 1. The variable H is in fact concentrated around its average away from 0 (see the detailed study by Litvak and den Zwet [16]) and from $+\infty$ (exponential decay). The phenomenon of moderate oscillations (but not as small) seen for Knuth's Algorithm is thus also true in this case.

4 Extensions

In Section 2, the asymptotic behavior of the sequence $(\mathbb{E}(R_n))$ is directly obtained from Relation (7). To obtain this identity, the key steps are 1) the de-Poissonization and 2) the use of Fubini's Theorem in Equation (8) to remove the series. In Section 3, the key step is also the use of Fubini's Theorem in Equation (15) to get rid of the series.

It may be thought that these derivations are nevertheless possible only because of the particular expressions involved: For example, in Section 2, the special properties of Poisson processes are critical in the solution of the problem. The purpose of this section is to propose a simple method along the same lines to study the asymptotic behavior of some series and functions. As it will be seen, it applies in various situation, even when there is no probabilistic interpretation of the sequence/function under study.

Dyadic Sums

This example is taken from Flajolet *et al.* [9] page 35. The behavior of the function

$$G(x) = \sum_{k \geq 1} g(x/2^k) \quad (16)$$

is investigated when x goes to infinity.

It is assumed that $x \rightarrow g(x)$ is differentiable and that $g(0) = 0$ so that the sum is well defined. These conditions are not the weakest possible. The goal here is to keep the presentation as simple as possible, not to get the most accurate result for this special case.

By using the elementary relation, for $x > 0$,

$$g(x) = \int_0^x g'(u) du,$$

the series (16) can be written as

$$G(x) = \sum_{k \geq 1} \int_0^x \mathbb{1}_{\{u < x/2^k\}} g'(u) du.$$

When the sum and the integral are permuted in the last expression, it yields

$$\int_0^x \sum_{k \geq 1} \mathbb{1}_{\{u < x/2^k\}} g'(u) du = \int_0^x [\log_2(x/u)] g'(u) du,$$

Fubini's Theorem states that this last term is indeed $G(x)$ when the condition

$$\int_0^x |\log_2(u)| |g'(u)| du < +\infty$$

holds. In this case, the function G can thus be represented as

$$\begin{aligned} G(x) &= \int_0^x [\log_2(x/u)] g'(u) du \\ &= - \int_0^x \{\log_2(x/u)\} g'(u) du + \int_0^x \log_2(x/u) g'(u) du. \end{aligned}$$

The following proposition has been proved.

Proposition 8. *If the function g is differentiable and such that $g(0) = 0$, under the condition*

$$\int_0^{+\infty} |\log(u)| |g'(u)| du < +\infty, \quad (17)$$

the dyadic sum $G(x)$ can be expressed as

$$G(x) = F(\log_2(x)) + \int_x^{+\infty} \{\log_2(x/u)\} g'(u) du + \frac{1}{\log 2} \int_x^{+\infty} \frac{g(u)}{u} du. \quad (18)$$

where F is the periodic function, with period 1, defined by

$$F(y) = -\frac{1}{\log 2} \int_0^{+\infty} \frac{g(u)}{u} du - \int_0^{+\infty} \{y - \log_2(u)\} g'(u) du, \quad y \geq 0.$$

Provided that g has a monotone behavior in the neighborhood of 0 and $+\infty$, Condition (17) is equivalent to the fact that the integral

$$\int_0^{+\infty} \frac{|g(u)|}{u} du$$

exists, i.e. that the Mellin transform of g is defined at 0.

From Equation (18), it is then not difficult to derive an asymptotic expansion of $G(x)$ as x goes to infinity.

Harmonic Sums

More generally, the simple method developed above can be used for a more general class of functions:

$$G(x) = \sum_{k \geq 1} \lambda_k g(\mu_k x).$$

This is the main application of Flajolet *et al.* [9]. For simplicity, it is assumed that the sequence (μ_k) is non-increasing and converging to 0 for example. *Provided that Fubini's Theorem can be applied*, one gets the following representation:

$$G(x) = \int_0^{+\infty} \sum_{k \geq 1} \lambda_k \mathbb{1}_{\{\mu_k \geq u/x\}} g'(u) du.$$

If, for $y > 0$ and $n \geq 1$,

$$\tau(y) = \sup\{k : \mu_k > y\} \text{ and } \Lambda(n) = \sum_{k=1}^n \lambda_k,$$

with the convention that $\sup\{\emptyset\} = 0$, then the harmonic sum can be written as

$$G(x) = \int_0^{+\infty} \Lambda(\tau(x/u)) g'(u) du.$$

Hence, the rates of growth of Λ and τ give the key of the asymptotic behavior of the function G .

Final Remark

It must be noted that, when it can be applied, the method proposed in this paper will require weaker assumptions than the analytic approach. Indeed, the conditions to apply Fubini's Theorem are quite minimal (which does not mean that there is no condition at all). On the contrary, the use of Mellin transform implies the existence of the transform itself on some specified strip of the complex plane together with some growth conditions at infinity.

5 Appendix on Poisson Processes

To keep the paper self-contained, this section recalls the basic definitions and results concerning Poisson processes used in this paper. See also Kingman [13] and Chapter 1 of Robert [21] for a more detailed presentation of these important stochastic processes.

Definition 9. *A random variable X has the Poisson distribution with parameter λ whenever*

$$\mathbb{P}(X = n) = \frac{\lambda^n}{n!} e^{-\lambda}, \quad n \in \mathbb{N}.$$

The following proposition is the basic property that motivates the use of Poisson transform when dealing with splitting algorithms. It is a striking elementary property of the Poisson distribution.

Proposition 10 (Splitting Property). *If X is a Poisson random variable with parameter λ , X balls are thrown randomly among n urns and for $1 \leq i \leq n$, X_i denotes the number of balls in the i th urn. The variables (X_i) are independent with a common Poisson distribution with parameter λ/n .*

Definition 11. *A Poisson process with intensity λ is an increasing sequence (t_n) of positive random variables such that*

- *the increments $t_{n+1} - t_n$, $n \geq 1$, are independent;*
- *For $n \geq 1$ and $x \geq 0$, $\mathbb{P}(t_{n+1} - t_n \geq x) = \exp(-\lambda x)$.*

A Poisson process (t_n) is also represented as a non-decreasing integer valued function $(N([0, t]), t \geq 0)$ on \mathbb{R}_+ where, for $t \geq 0$, $N([0, t])$ is the number of t_n 's in the interval $[0, t]$: $N([0, t]) = n$ on the event $\{t_n \leq t < t_{n+1}\}$. For a Poisson process, the representations as a sequence (t_n) or a non-decreasing function $(N([0, t]), t \geq 0)$ are of course equivalent.

Proposition 12. *It (t_n) is a Poisson process with intensity λ , for $t > 0$,*

- *the variable $N([0, t])$ has a Poisson distribution with parameter λ ;*
- *conditionally on the event $\{N([0, t]) = n\}$, the variables*

$$0 \leq t_1 \leq t_2 \leq \dots \leq t_n \leq t$$

have the same distribution as the reordering of n independent, uniformly distributed random variables on $[0, t]$.

References

- [1] David Aldous, *Ultimate instability of exponential back-off protocol for acknowledgement-based transmission of control of random access communication channels*, IEEE Transactions on Information Theory **33** (1987), 219–223.
- [2] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, *Algorithms*, MIT Press, Cambridge, Massachusetts, 1990.
- [3] John I. Capetanakis, *Tree algorithms for packet broadcast channels*, IEEE Transactions on Information Theory **25** (1979), no. 5, 505–515.
- [4] Philippe Carmona, Frédérique Petit, and Marc Yor, *On the distribution and asymptotic results for exponential functionals of Lévy processes*, Exponential functionals and principal values related to Brownian motion, Rev. Mat. Iberoam., Madrid, 1997, pp. 73–130.
- [5] Hubert Delange, *Sur la formule sommatoire de la fonction somme des chiffres*, L'enseignement Mathématique **XXI** (1975), no. 1, 31–47.
- [6] Vincent Dumas, Fabrice Guillemin, and Philippe Robert, *A Markovian analysis of Additive-Increase Multiplicative-Decrease (AIMD) algorithms*, Advances in Applied Probability **34** (2002), no. 1, 85–111.
- [7] Anthony Ephremides and Bruce Hajek, *Information theory and communication networks: an unconsummated union*, IEEE Transactions on Information Theory **44** (1998), no. 6, 1–20.
- [8] Philippe Flajolet, *Approximate counting: a detailed analysis*, BIT. Nordisk Tidskrift for Informationsbehandling (BIT) **25** (1985), no. 1, 113–134.
- [9] Philippe Flajolet, Xavier Gourdon, and Philippe Dumas, *Mellin transforms and asymptotics: harmonic sums*, Theoretical Computer Science **144** (1995), no. 1-2, 3–58, Special volume on mathematical analysis of algorithms.
- [10] Philippe Flajolet and Philippe Jacquet, *Analytic models for tree communication protocols*, NATO Advance Science Institute Series (A. R. Odoni L. Bianco and G. Szegő, eds.), Series F: Computer and Systems Sciences, vol. 38, Springer Verlag, 1987, pp. 223–234.
- [11] Philippe Flajolet and Nigel Martin, *Probabilistic counting for data base applications*, Journal of Computer Systems **31** (1985), 182–209.
- [12] Philippe Jacquet and Wojciech Szpankowski, *Analytical depoissonization and its applications*, Theoretical Computer Science **201** (1998), no. 1–2, 1–62.
- [13] J. F. C. Kingman, *Poisson processes*, Oxford Studies in Probability, 1993.

-
- [14] Peter Kirschenhofer, Helmut Prodinger, and Wojciech Szpankowski, *Analysis of a splitting process arising in probabilistic counting and other related algorithms*, Random Structures and Algorithms **9** (1996), no. 4, 379–401.
 - [15] Donald E. Knuth, *The art of computer programming. Volume 3*, Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1973, Sorting and searching, Addison-Wesley Series in Computer Science and Information Processing.
 - [16] Nelly Litvak and Willem R. van Zwet, *On the minimal travel time needed to collect n items on a circle*, Annals of Applied Probability **14** (2004), no. 2, 881–902.
 - [17] Hosam M. Mahmoud, *Evolution of random search trees*, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons Inc., New York, 1992, A Wiley-Interscience Publication.
 - [18] R. Metcalf and D. Boggs, *Ethernet: Distributed packet switching for local computer networks*, Communications of the ACM **19** (1976), no. 7, 395–403.
 - [19] Hanène Mohamed and Philippe Robert, *A probabilistic analysis of some tree algorithms*, August 2004, <http://www-rocq.inria.fr/~robert/src/papers/2004-8.pdf>.
 - [20] Nicholas Pippenger, *An elementary approach to some analytic asymptotics*, SIAM Journal on Mathematical Analysis **24** (1993), no. 5, 1361–1377.
 - [21] Philippe Robert, *Stochastic networks and queues*, Stochastic Modelling and Applied Probability Series, vol. 52, Springer, New-York, June 2003.
 - [22] B. S. Tsybakov and V. A. Mikhaïlov, *Free synchronous packet access in a broadcast channel with feedback*, Problems Inform. Transmission **14** (1978), no. 4, 32–59.



Unité de recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)

Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)

Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
<http://www.inria.fr>
ISSN 0249-6399