Journal of Number Theory 129 (2009) 2267–2273

Contents lists available at ScienceDirect

Journal of Number Theory

www.elsevier.com/locate/jnt

A note on "Euclidean algorithms are Gaussian" by V. Baladi and B. Vallée

E. Cesaratto [a,b,*]

[a] *Instituto de Desarrollo Humano, Universidad Nac. de Gral. Sarmiento, J.M. Gutiérrez 1150, (1613) Los Polvorines, Pcia. de Buenos Aires, Argentina*
[b] *CONICET, Argentina*

ARTICLE INFO

Article history:
Received 6 July 2007
Revised 31 January 2009
Available online 4 June 2009
Communicated by Carl Pomerance

Keywords:
Distributional analysis
Euclidean algorithms

ABSTRACT

The paper "Euclidean algorithms are Gaussian" [V. Baladi, B. Vallée, Euclidean algorithm are Gaussian, J. Number Theory 110 (2005) 331–386], is devoted to the distributional analysis of three variants of Euclidean algorithms. The Central Limit Theorem and the Local Limit Theorem obtained there are the first ones in the context of the "dynamical analysis" method. The techniques developed have been applied in further various works (e.g. [V. Baladi, A. Hachemi, A local limit theorem with speed of convergence for Euclidean algorithms and Diophantine costs, Ann. Inst. H. Poincaré Probab. Statist. 44 (2008) 749–770; E. Cesaratto, J. Clément, B. Daireaux, L. Lhote, V. Maume, B. Vallée, Analysis of fast versions of the Euclid algorithm, in: Proceedings of Third Workshop on Analytic Algorithmics and Combinatorics, ANALCO'08, SIAM, 2008; E. Cesaratto, A. Plagne, B. Vallée, On the non-randomness of modular arithmetic progressions, in: Fourth Colloquium on Mathematics and Computer Science. Algorithms, Trees, Combinatorics and Probabilities, in: Discrete Math. Theor. Comput. Sci. Proc., vol. AG, 2006, pp. 271–288]). These theorems are proved first for an auxiliary probabilistic model, called "the smoothed model," and after, the estimates are transferred to the "true" probabilistic model. In this note, we remark that "the smoothed model" described in [V. Baladi, B. Vallée, Euclidean algorithm are Gaussian, J. Number Theory 110 (2005) 331–386] is not adapted to this transfer and replaces it by an adapted one. However, the results remain unchanged.

* Address for correspondence: Instituto de Desarrollo Humano, Universidad Nac. de Gral. Sarmiento, J.M. Gutiérrez 1150, (1613) Los Polvorines, Pcia. de Buenos Aires, Argentina.
E-mail address: ecesarat@ungs.edu.ar.

0022-314X/$ – see front matter
doi:10.1016/j.jnt.2009.02.018

1. Introduction

The paper "Euclidean algorithms are Gaussian" by V. Baladi and B. Vallée [2] appeared in the Journal of Number Theory in 2005. It performs a distributional analysis of three Euclidean algorithms. A Central Limit Theorem and a Local Limit Theorem are obtained for a wide variety of additive costs. These results are the first ones about distributional analysis in the context of the "dynamical analysis" method, which mixes tools from analysis of algorithms and dynamical systems theory. The Central and Local Limit Theorems are proved first for an auxiliary probabilistic model, called "the smoothed model," and after, the estimates are transferred to the "true" probabilistic model. This work is done in Section 4.2, more specifically in Lemmata 11, 12, 13 and 14. In this note, we remark that "the smoothed model" described in [2] is not adapted to this transfer and replaces it by an adapted one. However, the statements of Lemmata 11, 12, 13 and 14 remain unchanged when the new "smoothed probabilistic model" replaces the old one. Moreover, only the proof of Lemma 14 needs a modification. This note is motivated by the fact that the techniques introduced in [2] and, in particular the smoothed model, have been used in further various works. Such instances are the analysis of the Knuth–Schönage algorithm [3], the study of Arnold's constant for modular arithmetic progressions [5], and also various extensions of the results of [2] for more general costs [1]. An alternative method which also "repairs" Baladi–Vallée's paper is due to L. Lhote and is explained in [4]. It does not introduce an auxiliary probabilistic model and uses other computations, of different style.

1.1. Euclidean algorithms, costs and probabilistic setting

The paper [2] analyzes three different variants of the Euclid algorithm: the standard, the centered, and the odd Euclidean algorithms. Briefly, on the integer pair (u, v) so that $\frac{u}{v} \in \mathcal{I}$ (for some interval $\mathcal{I} \subset [0, 1]$ which depends on the algorithm), each algorithm performs a sequence of Euclidean divisions for computing the gcd of the pair. Each Euclidean division, of the form $v = mu + \varepsilon r$ with $r/u \in \mathcal{I}$, creates a digit (m, ε), with conditions on digits which also depend on the algorithm. Each algorithm applied to the rational u/v builds a specific continued fraction

$$\frac{u}{v} = \cfrac{1}{m_1 + \cfrac{\epsilon_1}{m_2 + \cfrac{\epsilon_2}{\ddots + \cfrac{\epsilon_{P-1}}{m_P}}}},$$

of depth $P = P(u, v)$.

The costs considered on the execution of any of the Euclidean algorithm over a pair (u, v) are defined as follows: Given a digit-cost function $c : \mathbb{N} \to \mathbb{R}^+$, the cumulative cost $C(u, v)$ is equal to

$$C(u, v) = \sum_{i=1}^{P(u,v)} c\big(m_i(u, v)\big).$$

This kind of costs includes a variety of costs of great interest for algorithmical studies; for instance, the number P of steps of the algorithm is obtained by setting $c = 1$.

1.2. The probabilistic model of interest

In probabilistic analysis of algorithms, the set of inputs with size bounded by a given natural N is endowed with a probability measure. The cost C is, then, considered as a random variable. Average-case analysis estimates the asymptotic expectation of C (when $N \to \infty$), whereas distributional analysis determines the asymptotic distribution of C.

In our framework, the inputs of the Euclidean algorithms are the integer pairs (u, v) with $u/v \in \mathcal{I}$ and the size of an input (u, v) is the positive integer v. The set of the inputs with size bounded by N is endowed with the uniform probability measure. Notice that the cumulative cost C only depends

on the continued fraction expansion of the rational u/v, and does not depend on the gcd of the pair (u, v). This entails that the study of the cost C can be restricted to the set

$$\Omega_N = \left\{ (u, v) \in \mathbb{N}_\star^2;\ \gcd(u, v) = 1,\ \frac{u}{v} \in \mathcal{I},\ v \leqslant N \right\} \tag{1}$$

endowed with the uniform probability $\mathbb{P}_N$.

Results about the average number of steps in the standard Euclidean algorithm were obtained independently by Heilbronn [7] and Dixon [6] and for the centered algorithm by Rieger [9]. Later, Hensley [8] proved that the number of steps of the standard Euclidean algorithm follows asymptotically a Gaussian distribution. The dynamical analysis method introduced by Vallée in the nineties provided a unified framework for the average-case analysis of Euclidean algorithms for a large class of costs. The paper [2] gave rise to the distributional analysis in the context of the Dynamical Analysis Method. The techniques developed in that paper allow to revisit the previous results about the average-case in order to obtain results about the distribution.

2. The rôle of the "smoothed probabilistic model"

The methods developed in [2] mainly rely on precise estimates for the moment generating function $\mathbb{E}_N[\exp(wC)]$, when w belongs to some complex neighborhood of 0.

The hard technical work of the paper is devoted to obtain estimates of the function

$$\Psi_w(N) := \sum_{Q \leqslant N} \sum_{n \leqslant Q} c_n(w) = \sum_{Q \leqslant N} \Phi_w(Q), \tag{2}$$

where

$$\Phi_w(N) := \sum_{n \leqslant N} c_n(w), \quad \text{and} \quad c_n(w) := \sum_{(u,n)\,\gcd(u,n)=1} \exp\bigl[wC(u, n)\bigr].$$

The double sum $\Psi_w(N)$ appears because the authors use as a main tool the Perron formula of order two. It does not seem possible to obtain direct estimates on the "simple" sum $\Phi_w(N)$. However, this is the final purpose of the authors, since the moment generating function $\mathbb{E}_n(\exp[wC])$ is expressed with $\Phi_w(N)$ under the form

$$\mathbb{E}_N\bigl[\exp(wC)\bigr] = \frac{\Phi_w(N)}{\Phi_0(N)}. \tag{3}$$

Then, they proceed in three steps.

(1) They first obtain estimates about $\Psi_w(N)$. At the end of Section 3, it is proved that

$$\Psi_w(N) := A(w) N^{2\sigma(w)+1} \bigl(1 + O\bigl(N^{-\alpha}\bigr)\bigr), \tag{4}$$

where A and σ are analytic functions of w, α is a positive number and the O term is uniform for w in a neighborhood of 0.

(2) Then, the authors introduced an auxiliary model, the so-called smoothed model, in order to exploit the estimates of the Cesàro sums $\Psi_w(N)$. This model depends on some function $T \mapsto \epsilon(T)$ and is denoted by $(\overline{\Omega}_N(\epsilon), \overline{\mathbb{P}}_N(\epsilon))$. Lemmata 11, 12, and 13 of the paper [2] prove that, with the smoothed model, the cost C follows an asymptotic Gaussian law as $N \to \infty$.

(3) Under some hypothesis on $\epsilon(T)$, Lemma 14 states that the smoothed model is close enough to the "true" probabilistic model $(\Omega_N, \mathbb{P}_N)$ to ensure the transfer of estimates from $(\overline{\Omega}_N(\epsilon), \overline{\mathbb{P}}_N(\epsilon))$ to $(\Omega_N, \mathbb{P}_N)$.

Baladi and Vallée describe their intermediary model as follows (see [2]): "*Associate to some non-negative function $T \mapsto \epsilon(T)$, with $\epsilon(T) \leqslant 1$, the probabilistic models $(\overline{\Omega}_N(\epsilon), \overline{\mathbb{P}}_N(\epsilon))$ as follows: For any integer N, set $\overline{\Omega}_N(\epsilon) = \Omega_N$; next, choose uniformly an integer Q between $N - \lfloor N\epsilon(N)\rfloor$ and N, and draw uniformly an element (u, v) of Ω_Q. Slightly abusing language, we refer to the function C in the model $(\overline{\Omega}_N(\epsilon), \overline{\mathbb{P}}_N(\epsilon))$ as the smoothed cost. The cumulative value of* $\exp[wC]$ *for $\overline{\mathbb{P}}_N(\epsilon)$ is*

$$\overline{\Phi}_w(N) := \frac{1}{\lfloor N\epsilon(N)\rfloor} \sum_{Q=N-\lfloor N\epsilon(N)\rfloor}^{N} \sum_{n\leqslant Q} c_n(w), \tag{5}$$

so that the moment generating function of the smoothed cost is just

$$\overline{\mathbb{E}}_N\big[\exp(wC)\big] = \frac{\overline{\Phi}_w(N)}{\overline{\Phi}_0(N)}."\tag{6}$$

Eqs. (5), (6) are exactly like Eqs. (4.4), (4.5) of [2]. With the relation

$$\overline{\Phi}_w(N) = \frac{1}{\lfloor N\mathcal{E}(N)\rfloor}\big[\Psi_w(N) - \Psi_w\big(N - \lfloor N\epsilon(N)\rfloor\big)\big],$$

together with equality (6), Baladi and Vallée exploit the estimates $\Psi_w(N)$, and transfer them into estimates on $\overline{\mathbb{E}}_N[\exp(wC)]$.

However, Eqs. (5), (6) are false. The actual expected value of the random variable $\exp[wC]$ computed with the smoothed probabilistic model is

$$\overline{\mathbb{E}}_N\big[\exp(wC)\big] = \frac{1}{\lfloor N\epsilon(N)\rfloor} \sum_{Q=N-\lfloor N\epsilon(N)\rfloor}^{N} \frac{1}{|\Omega_Q|} \sum_{n\leqslant Q} c_n(w), \tag{7}$$

and it does not coincide with the value computed in (6).

3. An adapted smoothed probabilistic model

We now introduce another smoothed probabilistic model for which the key equation (6) is true. The new smoothed probabilistic model is denoted with an underline. Consider, for the same function ϵ as in [2], the (disjoint) union

$$\underline{\Omega}_N(\epsilon) := \bigcup_{N-\lfloor N\epsilon(N)\rfloor\leqslant Q\leqslant N} \Omega_Q \times \{Q\}$$

endowed with the uniform probability $\underline{\mathbb{P}}_N$. The cost function C (defined on Ω_N) is extended to $\underline{\Omega}_N(\epsilon)$ by the relation $C(u, v, Q) := C(u, v)$ for any Q which satisfies $N - \lfloor N\epsilon(N)\rfloor \leqslant Q \leqslant N$. In this probabilistic model, the cumulative value of $\exp(wC)$ is

$$\underline{\Phi}_w(N) := \sum_{Q=N-\lfloor N\epsilon(N)\rfloor}^{N} \sum_{n\leqslant Q} c_n(w), \quad \text{with } \underline{\Phi}_0(N) = \big|\underline{\Omega}_N(\epsilon)\big|,$$

so that the expectation of $\exp(wC)$ in our smoothed model is:

$$\underline{\mathbb{E}}_N\big[\exp(wC)\big] = \frac{\underline{\Phi}_w(N)}{\underline{\Phi}_0(N)}. \tag{8}$$

There is a close relation between $\underline{\Phi}_w$ and $\overline{\Phi}_w$:

$$\underline{\Phi}_w(N) = \lfloor N\epsilon(N) \rfloor \cdot \overline{\Phi}_w(N),$$

and finally, $\underline{\Phi}_w$ can be expressed as a function of Ψ_w, via

$$\underline{\Phi}_w(N) = \Psi_w(N) - \Psi_w\big(N - \lfloor N\epsilon(N) \rfloor\big). \tag{9}$$

Now we can exploit the estimates (4), and then, in the new model, Lemma 10 implies Lemma 11 in the same lines as [2].

3.1. A new proof for Lemma 14

We now provide a new proof for our version of Lemma 14 of [2], where we replace their smoothed model (overlined) by our model (underlined). Our version of Lemma 14 is the following:

Lemma 14. *Suppose that* $\lim_{N\to\infty} \epsilon(N) = 0$ *with* $\epsilon(N)^{-1} = O(N/\log N)$. *Then the distance between the distributions* $\mathbb{P}_N$ *and* $\underline{\mathbb{P}}_N(\epsilon)$ *on* Ω_N *is* $O(\epsilon(N))$.

Proof. In the following, we denote by $N' := N - \lfloor N\epsilon(N) \rfloor$. First recall (as in [2]) that there is $K > 0$ so that

$$|\Omega_N| = KN^2\left(1 + O\left(\frac{\log N}{N}\right)\right), \qquad \mathbb{P}_N(u, v) = \frac{1}{KN^2}\left(1 + O\left(\frac{\log N}{N}\right)\right), \tag{10}$$

for all $(u, v) \in \Omega_N$. There exist precise estimates for K (see [2]). Now, thanks to the hypothesis on ϵ, we know that

$$\frac{\log Q}{Q} \leqslant \frac{\log N}{N'} = \frac{\log N}{N}\big(1 + O\big(\epsilon(N)\big)\big) = O\big(\epsilon(N)\big), \quad \text{for any } Q \in]N', N]. \tag{11}$$

The definition of $\underline{\Omega}_N(\epsilon)$ entails that

$$\big|\underline{\Omega}_N(\epsilon)\big| = \sum_{Q=N'}^{N} |\Omega_Q| = K\big(1 + O\big(\epsilon(N)\big)\big) \sum_{Q=N'}^{N} Q^2 \sim KN^3\epsilon(N). \tag{12}$$

Let us observe that now $\mathbb{P}_N$ and $\underline{\mathbb{P}}_N$ are not defined on the same probability space. However, we are only interested in dealing with sets $\underline{A} \subset \underline{\Omega}_N(\epsilon)$ which come from subsets $A \subset \Omega_N$: we deal with sets $\underline{A}$ of the form

$$\underline{A} = \bigcup_{N' \leqslant Q \leqslant N} (A \cap \Omega_Q) \times \{Q\}$$

where A is a subset of Ω_N. By abuse of language, for such a subset, $\underline{\mathbb{P}}_N(A)$ will mean $\underline{\mathbb{P}}_N(\underline{A})$. There are two cases of interest for sets $A \subset \Omega_N$: The ordinary subsets included in $\Omega_{N'}$, for which, with (12),

$$\underline{\mathbb{P}}_N(A) := \underline{\mathbb{P}}_N(\underline{A}) = \frac{\lfloor N\epsilon(N) \rfloor \cdot |A|}{|\underline{\Omega}_N(\epsilon)|} = \frac{|A|}{KN^2} \cdot \big[1 + O\big(\epsilon(N)\big)\big] \tag{13}$$

and, the exceptional subsets A not included in $\Omega_{N'}$.

Now, Ω_N decomposes into the ordinary subset $\mathcal{O}_N := \Omega_{N'}$ and the exceptional subset $\mathcal{E}_N := \Omega_N \setminus \mathcal{O}_N$. Both probabilities of the exceptional subset are linked by the following inequality

$$\underline{\mathbb{P}}(\mathcal{E}_N) < \mathbb{P}(\mathcal{E}_N), \tag{14}$$

which is proved at the end.

The cardinality of $|\mathcal{E}_N|$ is $O(N^2\epsilon(N))$ (due to (11)). Then we have

$$\underline{\mathbb{P}}(\mathcal{E}_N) < \mathbb{P}(\mathcal{E}_N) := \frac{|\mathcal{E}_N|}{|\Omega_N|} = O\left(\frac{|\mathcal{E}_N|}{N^2}\right) = O\big(\epsilon(N)\big).$$

For any pair (u, v) of the ordinary subset $\mathcal{O}_N$, relations (13) and (10) entail the estimate

$$\frac{\underline{\mathbb{P}}_N(u, v)}{\mathbb{P}_N(u, v)} = 1 + O\big(\epsilon(N)\big).$$

Finally, for any $A \subset \Omega_N$, the difference $|\,\mathbb{P}_N(A) - \underline{\mathbb{P}}_N(A)|$ is less than

$$\big|\mathbb{P}_N(A \cap \mathcal{O}_N) - \underline{\mathbb{P}}_N(A \cap \mathcal{O}_N)\big| + \big|\mathbb{P}_N(A \cap \mathcal{E}_N) - \underline{\mathbb{P}}_N(A \cap \mathcal{E}_N)\big| = O\big(\epsilon(N)\big).$$

To conclude, we prove (14).

Proof of (14). First, set $a_j := |\Omega_j - \Omega_{j-1}|$ with $\Omega_0 = \emptyset$ and $1 \leqslant j \leqslant N$. Let $A_k := \sum_{j=1}^{k} a_j$, and $A'_k := \sum_{j=k+1}^{N} a_j$. With this notation we have that

$$\mathbb{P}(\mathcal{E}_N) := \frac{|\Omega_N - \Omega_{N'}|}{|\Omega_N|} = \frac{A'_{N-N'}}{A_N}$$

and, in the same manner,

$$\underline{\mathbb{P}}(\mathcal{E}_N) := \frac{\sum_{Q=N'}^{N} |\Omega_Q - \Omega_{N'}|}{\sum_{Q=N'}^{N} |\Omega_Q|} = \frac{\sum_{j=N-N'+1}^{N} (N-j+1)a_j}{(N-N'+1)A_{N-N'} + \sum_{j=N-N'+1}^{N} (N-j+1)a_j}.$$

We obtain the inequality

$$\frac{A'_{N-N'}}{A_N} > \frac{\sum_{j=N-N'+1}^{N} (N-j+1)a_j}{(N-N'+1)A_{N-N'} + \sum_{j=N-N'+1}^{N} (N-j+1)a_j}$$

by cross multiplying, canceling and finally observing that $N - N' + 1 > N - j + 1$ in all cases. □

4. Conclusion

We provide here an intermediary model which is adapted to the method of [2], whereas there was not the case for the intermediary model proposed in [2]. Replacing in the paper [2] their auxiliary model by the present one corrects the paper [2]. The results and methods of the rest of the paper [2] remain exact and unchanged.

Acknowledgments

This work was done when the author spent one year in the GREYC laboratory with a post-doctoral position funded by CNRS. The author is very grateful to Viviane Baladi and Brigitte Vallée for many and valuable discussions. I also thank the referee for useful comments.

References

[1] V. Baladi, A. Hachemi, A local limit theorem with speed of convergence for Euclidean algorithms and Diophantine costs, Ann. Inst. H. Poincaré Probab. Statist. 44 (2008) 749–770.
[2] V. Baladi, B. Vallée, Euclidean algorithm are Gaussian, J. Number Theory 110 (2005) 331–386.
[3] E. Cesaratto, J. Clément, B. Daireaux, L. Lhote, V. Maume, B. Vallée, Analysis of fast versions of the Euclid algorithm, in: Proceedings of Third Workshop on Analytic Algorithmics and Combinatorics, ANALCO'08, SIAM, 2008.
[4] E. Cesaratto, J. Clément, B. Daireaux, L. Lhote, V. Maume, B. Vallée, Regularity of the Euclid algorithm. Application to the analysis of fast gcd algorithms, J. Symbolic Comput. 44 (2009) 726–767.
[5] E. Cesaratto, A. Plagne, B. Vallée, On the non-randomness of modular arithmetic progressions, in: Fourth Colloquium on Mathematics and Computer Science. Algorithms, Trees, Combinatorics and Probabilities, in: Discrete Math. Theor. Comput. Sci. Proc., vol. AG, 2006, pp. 271–288.
[6] J.D. Dixon, The number of steps in the Euclidean algorithm, J. Number Theory 2 (1970) 414–422.
[7] H. Heilbronn, On the average length of a class of continued fractions, in: P. Turan (Ed.), Number Theory and Analysis, Plenum, New York, 1969, pp. 87–96.
[8] D. Hensley, The number of steps in the Euclidean algorithm, J. Number Theory 49 (2) (1994) 142–182.
[9] G.J. Rieger, Über die mittlere Schrittanzahl bei Divisionalgorithmen, Math. Nachr. (1978) 157–180.