Additive Decompositions, Random Allocations, and Threshold Phenomena.

Olivier Dubois[†], Guy Louchard[*], Jacques Mandler[†]

[†] LIP6, C.N.R.S.-Université Paris 6, 4 place Jussieu, 75252 Paris cedex 05, France,
email: Olivier.Dubois@lip6.fr, Jacques.Mandler@lip6.fr.
[*] Université Libre de Bruxelles, Département d'Informatique, CP 212, Boulevard du Triomphe,
B-1050 Bruxelles, Belgium, email: louchard@ulb.ac.be.

August 13, 2003

Abstract

An *additive decomposition* of a set I of nonnegative integers is an expression of I as the arithmetic sum of two other such sets. If the smaller of these has p elements, we have a *p-decomposition*. If I is obtained by randomly removing n^α integers from $\{0, ..., n-1\}$, decomposability translates into a balls-and-urns problem which we start to investigate (for large n) by first showing that the number of p-decompositions exhibits a threshold phenomenon as α crosses a p-dependent critical value. We then study in detail the distribution of the number of 2-decompositions. For this last case we show that the threshold is sharp and we establish the threshold function.

1 Definitions, Notations and basic Properties.

For integers i and sets of integers J and K, we use the fairly standard notations

$$i + J = J + i = \{i + j : j \in J\}, \qquad J + K = \{j + k : (j, k) \in J \times K\}.$$

In this paper, we restrict ourselves mostly to sets of *nonnegative* integers, though we do allow ourselves the first notation with negative i, so long as $-i$ is less than $\min J$.

Definition 1.1 *A set I of nonnegative integers is said to be decomposable if there exist sets J, K, each with at least two elements, such that $I = J + K$. We also say that I is J-decomposable, that (J, K) is a decomposition of I, and that J is a decomposition factor for I.*

Thus, a set of two integers is never decomposable; a set of three integers is decomposable iff they are in arithmetic progression; and a set $\{a, b, c, d\}$ of four integers $a < b < c < d$ is decomposable iff $a + d = b + c$. (In this paper, we consider the decomposability of finite sets only.)

Issues of decomposability of integer sets arise occasionally in mathematics and computer science, most immediately in the context of the factorization of polynomials, since plainly a necessary condition for a polynomial in one indeterminate with coefficients in some field to split into two factors is that its set of exponents be decomposable. The more restrictive notion of *irredundent* decomposition, where it is required that the map $(j, k) \mapsto j + k$ be one-to-one, gives a necessary and *sufficient* condition for a factorization to exist within the class of Minkowski polynomials, i.e., polynomials in $\mathbb{R}[X]$ with 0-1 coefficients.

Decompositions are, of course, generally not unique; further, if (J, K) is a decomposition of I, then $\min I = \min J + \min K$, and since any $0 \le i \le \min I$ can be written as $j + k$ with $0 \le j \le \min J$ and $0 \le k \le \min K$, resulting in $(I - i) = (J - j) + (K - k)$, decomposability is seen to be, in effect, a property of equivalence classes of subsets of $\mathbb{N}$ under translation.

Equally, if $I = J + K$ then also $I = (J - \min J) + (K + \min J)$. Therefore:

Lemma 1.2 *If I is decomposable, then it admits a decomposition factor containing* 0.

So in investigating whether a set J may be a factor of decomposition of I, we need, and shall, only consider sets containing 0. This remark lies at the root of the critical phenomena to be investigated. Also note that for one of the factors to be a subset of I, it is necessary and sufficient that the other factor contain 0 ($I = J+K$ with $K \subset I$ implies $\min I = \min J + \min K \geq \min J + \min I$, so $\min J \leq 0$); and for both factors to be subsets of I, it is necessary and sufficient that I contain 0.

Lemma 1.3 *A necessary and sufficient condition for a set J containing 0 to be a decomposition factor of I is: for any i in I, there is some (not necessarily unique) j in J such that $i - j + J \subset I$.*

Proof If $I = J + K$, then for any $i \in I$ there is a $j \in J$ such that $i - j \in K$, implying $i - j + J \subset K + J \subset I$. Conversely, if the condition is true, let $K = \{k \in \mathbb{N} : k + J \subset I\}$. By construction $J + K \subset I$, and by hypothesis $I \subset J + K$. ■

Note that for an arbitrary J, the condition in Lemma 1.3 is necessary and sufficient for $J - \min J$ to be a decomposition factor of I; J itself is one, iff the condition holds with the added requirement that $j \leq i$. Of course, for J containing 0, the definition of K might as well read $K = \{k \in I : k + J \subset I\}$.

For a given J, we now characterize the sets I which are *not* J-decomposable. This follows by negating the condition in Lemma 1.3:

Proposition 1.4 *Let Φ_J be the set of fixed-point free maps from J to itself, and for $\phi \in \Phi_J$ set:*

$$\mathcal{A}_\phi^J = \{I \subset \mathbb{N} : (\exists i \in I)\,(\forall j \in J)\,(i - j + \phi(j) \notin I)\}.$$

Then

$$\{I \subset \mathbb{N} : J \textit{ is not a decomposition factor for } I\} = \bigcup_{\phi \in \Phi_J} \mathcal{A}_\phi^J.$$

(We exclude maps ϕ with fixed points simply because they all give empty $\mathcal{A}_\phi$'s). To gain familiarity with Proposition 1.4, the reader may wish, before proceeding, to turn to the preliminary of Section 3, where the case $|J| = 2$ is worked out in explicit detail.

To simplify the study of the problem of decomposition as defined above, a main restriction is to consider the decomposition for a fixed size of the decomposition factor J.

Definition 1.5 *A set I of nonnegative integers is said to be p-decomposable if there exist sets J, K, one of which, say J, has exactly p elements, such that $I = J + K$.*

In this paper we will focus on the p-decomposability of a set I of integers. In the next section we define a natural probabilistic model and we show that there exists a threshold phenomenon for p-decomposability as a function of an appropriate *control parameter*. In the following sections we investigate 2-decomposability more specifically and we give a detailed study, providing asymptotic distributions dependent on the control parameter, and proving in particular the existence of a sharp threshold for which we establish the threshold function.

2 Random model and critical behavior for p-decomposability.

We show in this section that for a given integer $p \geq 2$, the p-decomposability of *random* sets of integers according to an appropriate probability distribution exhibits a critical behavior similar to that observed in random graphs [4], random Boolean formulae [5], and the partitioning of random sets of numbers [ibid.].

Consider n urns $U[0], U[1], ..., U[n-1]$, into which we throw $H = H(n)$ balls with equal probability to land in each urn; and associated binary r.v.'s, also denoted by $U[i]$. An empty urn is indicated by $U[i]=0$, an occupied one by $U[i]=1$ (irrespective of how many balls it has received). Depending on the outcome of this experiment, a set of integers $I \subset \{0,1,...,n-1\}$ is defined as follows:

$$I=\{i: 0 \leq i \leq n-1, U[i]=0\}$$

Otherwise expressed, we *remove* from $\{0,1,...,n-1\}=[n]$ the integers corresponding to *occupied* urns.

We are interested in the limiting behaviour of the p-decomposability of I for large n, with $H=n^{\alpha}(1+o(1))$ for some real (constant) $\alpha \in]0,1[$, the *control parameter;* though occasionally we assume a more general form for H, the number of integers removed from $[n]$, or of occupied urns. According to the value of α, two very different types of behavior arise as a function of a critical value of α.

To show this, we first establish two facts, one regarding almost sure decomposability, the other almost sure non-decomposability. Throughout the remainder of this section we set

$$\alpha_c=\alpha_c(p)=1-1/p. \tag{1}$$

Lemma 2.1 *There is a fixed subset with p elements, $J_0 \subset \mathbb{N}$, which, provided $\alpha<\alpha_c$, a.s. is a decomposition factor of I.*

Proof We may take J_0 to consist of any p numbers $0=j_0<j_1<j_2<...<j_{p-1}$, such that $j_{\nu+1}>2j_{\nu}$ for all ν; e.g., $j_{\nu}=3^{\nu}$ for $\nu \geq 1$ will do nicely. (The condition may be somewhat stronger than really required, though it would be unsuitable merely to specify sequences from which no arithmetic progression may be extracted.)
Observe that for any integer i, the intersection of any two of the p sets $i-j+J_0$, $j \in J_0$, reduces to $\{i\}$. Indeed, supposing $-j+j'=-\tilde{j}+\tilde{j}'$, $j \neq \tilde{j}$, $j \neq j'$, (hence $\tilde{j} \neq \tilde{j}'$), we may assume, without loss of generality, that the largest of the four j's is $\tilde{j}'$; since this cannot be j_0 (nor j_1 or j_2 actually), we have $\tilde{j}'=j_{\nu+1}$ for some ν, and all three other j's are $\leq j_{\nu}$. Therefore, $\tilde{j}'-\tilde{j} \geq j_{\nu+1}-j_{\nu}>j_{\nu} \geq j'-j$, a contradiction.

We may assume $n \geq 2j_{p-1}$. For an arbitrary $\phi \in \Phi_{J_0}$, it follows from the definition of $\mathcal{A}_{\phi}^{J_0}$ that the event $I \in \mathcal{A}_{\phi}^{J_0}$ is contained in the union of the events D_1^{ϕ}, D_2^{ϕ} and D_3^{ϕ} defined as follows:

$$\begin{array}{lll}
D_1^{\phi} & : & (\exists i \in [n])\ \ (j_{p-1} \leq i \leq n-1-j_{p-1}\ \wedge\ U[i]=0\ \wedge\ (\forall j \in J_0)\, U[i-j+\phi(j)]=1) \\
D_2^{\phi} & : & (\exists i \in [n])\ \ (i<j_{p-1}\ \wedge\ U[i]=0\ \wedge\ U[i-0+\phi(0)]=1) \\
D_3^{\phi} & : & (\exists i \in [n])\ \ (i \geq n-j_{p-1} \geq j_{p-1}\ \wedge\ U[i]=0\ \wedge\ U[i-j_{p-1}+\phi(j_{p-1})]=1)
\end{array}$$

Consequently $\Pr\left(I \in \mathcal{A}_{\phi}^{J_0}\right) \leq \Pr\left(D_1^{\phi}\right)+\Pr\left(D_2^{\phi}\right)+\Pr\left(D_3^{\phi}\right)$, and the three probabilities on the right are estimated 'by decomposing wrt (not necessarily disjoint) events of the form $i=\nu$'. More precisely, each of D_1^{ϕ}, D_2^{ϕ} and D_3^{ϕ} is an event $\mathcal{E}$ of the form $(\exists i \in [n])\, \Pi(i)$, where $\Pi(i)$ is as just indicated. We express $\mathcal{E}$ as the disjunction $\Pi(0) \vee \Pi(1) \vee \Pi(2) \vee ... \vee \Pi(n-1)$, where a great many terms are obviously (deterministically) false and are therefore dropped. Since the $\Pi(\nu)$'s are rather cumbersome to write, we will use the more concise notation $\mathcal{E} \wedge i:=\nu$ instead of $\Pi(\nu)$.
For D_1^{ϕ}, we draw on our preliminary observation to conclude that the p urns $U[i-j+\phi(j)]$ are distinct, hence

$$\begin{aligned}
\Pr\left(D_1^{\phi}\right) &\leq \Pr\left(D_1^{\phi} \wedge i:=j_{p-1}\right)+\Pr\left(D_1^{\phi} \wedge i:=j_{p-1}+1\right)+...+\Pr\left(D_1^{\phi} \wedge i:=n-1-j_{p-1}\right) \\
&\leq (n-2j_{p-1}) \Pr(p \text{ given urns are occupied and another one is empty}) \\
&\leq (n-2j_{p-1}) \frac{(n-1)^{H-p} H(H-1)...(H-p+1)}{n^H}=\mathcal{O}\left(\frac{H^p}{n^{p-1}}\right).
\end{aligned}$$

The other two are similar and somewhat simpler:

$$\begin{aligned}\Pr\left(D_2^{\phi}\right) &\leq \Pr\left(D_2^{\phi}\wedge i:=0\right)+\Pr\left(D_2^{\phi}\wedge i:=1\right)+...+\Pr\left(D_2^{\phi}\wedge i:=j_{p-1}-1\right),\\ \Pr\left(D_3^{\phi}\right) &\leq \Pr\left(D_3^{\phi}\wedge i:=n-j_{p-1}\right)+\Pr\left(D_3^{\phi}\wedge i:=n-j_{p-1}+1\right)+...+\Pr\left(D_3^{\phi}\wedge i:=n-1\right),\end{aligned}$$

and both are no larger than

$$j_{p-1}\times\Pr\left(\text{a given urn is occupied and another one is empty}\right),$$

so that

$$\Pr\left(D_2^{\phi}\right)+\Pr\left(D_3^{\phi}\right)\leq 2\times j_{p-1}\frac{(n-1)^{H-1}H}{n^H}=\mathcal{O}\left(\frac{H}{n}\right).$$

Finally, since Φ_{J_0} has $(p-1)^p=\mathcal{O}(1)$ elements,

$$\begin{aligned}\Pr\left(I\text{ is not }J_0\text{-decomposable}\right) &\leq \sum_{\phi\in\Phi_{J_0}}\Pr\left(I\in\mathcal{A}_{\phi}^{J_0}\right)\\ &= \mathcal{O}\left(\frac{H^p}{n^{p-1}}\right)+\mathcal{O}\left(\frac{H}{n}\right).\end{aligned}$$

It follows that if H^p/n^{p-1} tends to zero as n goes to infinity, the specific set J_0 is almost surely a p-decomposition factor of I, and *a fortiori* I is almost surely p-decomposable. In particular, p-decomposability succeeds a.s. if H (the number of balls, or the number of integers randomly withdrawn from $[n]$) is asymptotically n^{α} with $\alpha<\alpha_c$. ■

Lemma 2.2 *We have, unconditionally w.r.t. the behaviour of H:*

$$\mathbf{Pr}(I\text{ is }p\text{-decomposable})\leq n^p\exp\left[\frac{-1}{(p+1)^2}\left(\frac{H}{n^{\alpha_c}}\right)^p\right],$$

so that I is a.s. non-p-decomposable for $H\sim n^{\alpha}$ with $\alpha>\alpha_c$, or, more generally, whenever

$$\underline{\lim}\frac{H}{n^{\alpha_c}\left(p(p+1)^2\ln n\right)^{1/p}}>1. \tag{2}$$

Proof We take into account all potential p-decomposition factors of I, i.e. all sets $J\subset[n]$ with p elements (recall that we assume $\min J$ to be 0). However, in order to prove that a given J is *not* a p-decomposition factor of I, it suffices, according to Proposition 1.4, to exhibit a single $\phi_0\in\Phi_J$ such that $I\in\mathcal{A}_{\phi_0}^J$. We systematically select $\phi_0=\phi_0^J$ given by

$$\phi_0(0)=\min\left(J\backslash\{0\}\right),\quad \phi_0(j)=0\ \text{ for } j>0.$$

The set $W_J=\{\phi_0(j)-j:j\in J\}$ also has p elements, and the probability that J is a p-decomposition factor of I is bounded by the probability of the event $E_{J,I}$ that every translate $i+W_J, i\in I$, intersects I. Now, it follows from standard Poissonization, de-Poissonization arguments (e.g., [8]) that the urns are asymptotically independent, so we may assume n to be large enough that each urn is considered to have the probability H/n of being occupied, and $1-H/n$ of being empty; and κ urns, to have the probability $(H/n)^{\kappa}$ of being all occupied. Moreover, a set of q integers can intersect at most

q^2 of its translates, so for any $i_0 \in [n]$ there can be no more than $(p+1)^2$ values of $i \in [n]$ for which $i + (W_J \cup \{0\})$ intersects $i_0 + (W_J \cup \{0\})$. We can therefore find a set $H_J \subset [n]$ with at least $n/(p+1)^2$ elements, such that for $i \in H_J$ the sets $\{i\} \cup (i + W_J)$ are mutually disjoint. Because of this disjointness and of the fact that W_J does not contain 0, the $2|H_J|$ events $i \notin I, i \in H_J$ and $I \cap (i + W_J) \neq \emptyset, i \in H_J$, may be regarded as mutually independent. Let $\kappa_J(i)$ be the number of elements of $i + W_J$ that are in $[n]$, so $0 \leq \kappa_J(i) \leq p$. Then

$$\begin{aligned}
\Pr(E_{J,I}) &= \Pr\{\forall i \in [n] : i \notin I \vee I \cap (i + W_J) \neq \emptyset\} \\
&\leq \Pr\{\forall i \in H_J : i \notin I \vee I \cap (i + W_J) \neq \emptyset\} \\
&\leq \prod_{i \in H_J} \Pr\{i \notin I \vee I \cap ((i + W_J) \cap [n]) \neq \emptyset\} \\
&\leq \prod_{i \in H_J} \left[1 - \left(1 - \frac{H}{n}\right) \Pr(I \cap ((i + W_J) \cap [n]) = \emptyset)\right].
\end{aligned}$$

The probability appearing in the last product is the probability that $\kappa_J(i)$ urns are occupied, or $(H/n)^{\kappa_J(i)} \geq (H/n)^p$, so

$$\Pr(E_{J,I}) \leq \left[1 - \left(\frac{H}{n}\right)^p + \left(\frac{H}{n}\right)^{p+1}\right]^{\frac{n}{(p+1)^2}} = \exp\left(\frac{-1}{(p+1)^2} \frac{H^p}{n^{p-1}}\right)(1 + o(1)).$$

Hence, for large enough n and *any* $J \subset [n]$ with p elements,

$$\Pr(E_{J,I}) \leq 2 \exp\left(\frac{-1}{(p+1)^2}\left(\frac{H}{n^{\alpha_c}}\right)^p\right). \tag{3}$$

Since there are $\binom{n}{p} < n^p/2$ such sets, the bound on the probability that I is p-decomposable follows, as does the statement about the case $\alpha > \alpha_c$. As for (2), we need only rewrite this bound as

$$\mathbf{Pr}(I \text{ is } p\text{-decomposable}) \leq \exp\left\{-p \ln n \left[\frac{1}{p(p+1)^2 \ln n}\left(\frac{H}{n^{\alpha_c}}\right)^p - 1\right]\right\}.$$

■

Lemmas 2.1 and 2.2 imply

Theorem 2.3 *Let α_c be defined by (1). Then if $\alpha < \alpha_c$, I is a.s. p-decomposable, whereas if $\alpha > \alpha_c$, I is a.s. non-p-decomposable.*

We will see in Section 7 that, up to the constant, for $p = 2$ the upper bound deduced from (2) actually corresponds to a sharp threshold function for 2-decomposability. Indeed, by a detailed analysis we will establish :

Theorem 2.4 *Let the set I be obtained by randomly removing H integers from $[n]$. We have:*

$$\lim_{n \to \infty} \mathbf{Pr}(I \text{ is 2-decomposable}) = \begin{cases} 1 & if \quad \overline{\lim} \frac{H}{\sqrt{1/2\, n \ln n}} < 1 \\[2ex] 0 & if \quad \underline{\lim} \frac{H}{\sqrt{1/2\, n \ln n}} \geq 1 \end{cases}$$

This result for 2-decomposability and the above general upper bound lead us to suggest as a conjecture that $n^{\alpha_c}\left(p^{-1} \ln n\right)^{1/p}$ is a sharp threshold function for p-decomposability.

3 2-decomposability. Preliminary analysis

Since our main concern in this paper will be 2-decomposability, it is worth examining in more detail how to recognize whether some set $I \subset [n]$ admits $J = \{0, a\}$ as a decomposition factor. As it turns out, we need only ascertain, for all $i \in I$, whether the integers $i \pm a$ lie in I.

Indeed, of the 4 maps from J to itself, only one is fixed-point free, namely ϕ given by $\phi(0) = a$ and $\phi(a) = 0$. Hence, the general condition excluding J as a decomposition factor for I,

$$(\exists i \in I)\,(\exists \phi \in \Phi_J)\,(\forall j \in J) \qquad i - j + \phi(j) \notin I$$

just reads in this special case:

$$(\exists i \in I) \qquad i + a \notin I \ \wedge \ i - a \notin I. \tag{4}$$

Now when we look at the random model, integers outside $[n]$ are deterministically excluded from I, while for those in $[n]$, inclusion in I is random. Thus from a probabilistic viewpoint, it is appropriate to distinguish four mutually exclusive cases,

$$\begin{array}{ll} A: & i \pm a \text{ are both in } [n], \\ B: & \text{only } i - a \text{ is in } [n], \\ C: & \text{only } i + a \text{ is}, \\ D: & \text{neither is;} \end{array}$$

that is,

$$\begin{array}{lll} A: i & \geq & a \wedge i + a \leq n-1, \quad \text{i.e.:} \quad 2a \leq n-1, \quad a \leq i \leq n-a-1; \\ B: i & \geq & a \wedge i + a \geq n, \quad \text{i.e.:} \quad \max(a, n-a) \leq i \leq n-1; \\ C: i & < & a \wedge i + a \leq n-1, \quad \text{i.e.:} \quad 0 \leq i \leq \min(a-1, n-1-a); \\ D: i & < & a \wedge i \geq n-a, \quad \text{i.e.:} \quad a > n/2, \quad n-a \leq i \leq a-1. \end{array} \tag{5}$$

According to which of these obtains, we have distinct (and immediate) translations of (4) in terms of empty and occupied urns.

3.1 The singletons and the doublets

In this section, we analyze for large n, the number of singletons and doublets in our urn model. This will be useful in the next sections. We recall the model: we have n urns $U[0]..U[n-1]$ and we throw n^α balls, $0 < \alpha < 1$. For fixed a, let $H_{2,i}$ denote the event $\{U[i]$ and $U[i+a]$ are occupied $\}$. Let ν_i denote the event $U[i]$ is empty and similarly for ν_{ij} . Then, by Inclusion-Exclusion,

$$H_{2,0} = [\mathbf{1} - \nu_0][\mathbf{1} - \nu_a] = \mathbf{1} - \nu_0 - \nu_a + \nu_{0a}$$

Let us symbolically denote by V^j the event that j *different* given urns are empty. Then

$$PV[j] := \Pr[V^j] = (1 - j/n)^{n^\alpha} \sim 1 - jn^\alpha/n + \cdots$$

(Maple gives expressions to the desired precision, we will only give the first terms). Symbolically, $H_{2,.} = (\mathbf{1} - V)^2$, i.e.

$$p_2 := \Pr[H_{2,.}] = 1 - 2PV[1] + PV[2]$$

(we simply replace V^j by $PV[j]$). We obtain

$$p_2 = n^\alpha(-1 + n^\alpha)/n^2 + n^\alpha(-2 + 3n^\alpha - n^{2\alpha})/n^3 + \cdots.$$

Similarly, the event $H_{1,i} := \{U[i]$ is occupied $\}$ has the probability $1 - PV[1]$ and

$$p_1 := \Pr[H_{1,.}] = n^\alpha/n + n^\alpha(1 - n^\alpha)/(2n^2) + n^\alpha(1/3 - n^\alpha/2 + n^{2\alpha}/6)/n^3 + \cdots. \tag{6}$$

The event $H_{3,i} := \{U[i] \text{ and } U[i+a] \text{ and } U[i+2a] \text{ are occupied }\}$ is given by $(\mathbf{1}-V)^3$ and this leads to

$$p_3 := \Pr[H_{3,.}] = n^\alpha(2-3n^\alpha+n^{2\alpha})/n^3 + n^\alpha(9-33/2n^\alpha+9n^{2\alpha}-3/2n^{3\alpha})/n^4 + \cdots$$

and, finally, the event $H_{4,i,j} := \{U[i] \text{ and } U[i+a] \text{ and } U[j] \text{ and } U[j+a] \text{ are occupied, } j > i+a\}$, is given by $(\mathbf{1}-V)^4$ and this leads to

$$p_4 = n^\alpha(-6+11n^\alpha-6n^{2\alpha}+n^{3\alpha})/n^4 + \cdots$$

We recall that $|I| = n - S$, where the random variable

$$S := \text{number of singletons} = \sum_0^{n-1} H_{1,i}$$

has mean

$$M_S := \mathbb{E}(S) = np_1 = n^\alpha + n^\alpha(1-n^\alpha)/(2n) + n^\alpha(1/3 - n^\alpha/2 + n^{2\alpha}/6)/n^2 + \cdots,$$

and

$$\mu_{2,S} := \mathbb{E}(S^2) = np_1 + n(n-1)p_2$$

with

$$\sigma_S^2 := \mathrm{VAR}(S) = n^\alpha(-1+n^\alpha)/(2n) + n^\alpha(-5/6n^{2\alpha}+2n^\alpha-7/6)/n^2 + \cdots$$

By standard asymptotic analysis,

$$S \sim \mathcal{N}(M_S^*, \sigma_S^{*2}),$$

where $\mathcal{N}$ is the classical Gaussian RV and M_S^*, σ_S^{*2} are the dominant terms of M_S, σ_S^2, i.e.

$$M_S^* = n^\alpha, \sigma_S^{*2} = n^{2\alpha}/(2n).$$

Let us now analyze the doublets. We see that, asymptotically, the doublets are independent: $p_4 \sim p_2^2$. Also, $p_i \ll 1$. The random variable

$$\Delta := \text{number of doublets} = \sum_0^{n-a-1} H_{2,i}$$

has mean

$$M := \mathbb{E}(\Delta) = (n-a)p_2 = n^\alpha(-1+n^\alpha)/n + n^\alpha(-2+3n^\alpha-n^{2\alpha}-a(-1+n^\alpha))/n^2 + \cdots, \quad (7)$$

and

$$\mu_2 := \mathbb{E}(\Delta^2) = (n-a)p_2 + (2n+\beta)p_3 + (n^2+\gamma n+\delta)p_4$$

where β, γ, δ are unimportant constants for the precision we need here, and

$$\sigma^2 := \mathrm{VAR}(\Delta) = n^\alpha(-1+n^\alpha)/n + n^\alpha(-3n^{2\alpha}+(-a+7)n^\alpha+(-4+a))/n^2 + \cdots \quad (8)$$

It is now clear that we must consider 3 regions: $\alpha < 1/2, \alpha = 1/2, \alpha > 1/2$.

3.1.1 $\alpha < 1/2$

In this region, we obtain, by inclusion-exclusion,

$$\begin{aligned}
\Pr[\Delta = 0] &= \Pr\left[\prod_{i=0}^{n-1-a}[\mathbf{1} - H_{2,i}]\right] \\
&\sim 1 - (n-a)p_2 + (n+\gamma)p_3 + (n^2/2 + \delta n + \tau)p_4 + \cdots \\
&= 1 + n^\alpha(1-n^\alpha)/n + n^\alpha(1 - 1/2n^\alpha - n^{2\alpha} + a(-1+n^\alpha) + 1/2n^{3\alpha})/n^2 + \cdots \\
\Pr[\Delta = 1] &= \Pr\left[\sum_{i=0}^{n-1-a} H_{2,i}\prod_{j\neq i}[\mathbf{1} - H_{2,j}]\right].
\end{aligned}$$

However we note that, as far as $H_{2,j}, j \neq i$ is concerned, several balls can fall into $U[i], U[i+a]$ (actually with a multinomial law). So we must use $n^\alpha - c$ instead of n^α, with $c(\text{random}) = \mathcal{O}(1)$. To check this effect, we have expanded $PV[j]$ with $n^\alpha - c$, and p_i becomes

$$\begin{aligned}
p_{2,c} &= [n^\alpha(-1+n^\alpha) - 2n^\alpha c + c(1+c)]/n^2 + \cdots \\
p_{3,c} &= [n^\alpha(2 - 3n^\alpha + n^{2\alpha}) + n^\alpha c(-3n^\alpha + 3c + 6) + c(-2 - 3c - c^2)]/n^3 + \cdots \\
p_{4,c} &= [n^\alpha(-6 + 11n^\alpha - 6n^{2\alpha} + n^{3\alpha}) + n^\alpha c(-4n^{2\alpha} + n^\alpha(18+6c) - 4c^2 - 18c - 22) \\
&+ c(6 + 11c + 6c^2 + c^3)]/n^4 + \cdots
\end{aligned}$$

We obtain

$$\begin{aligned}
\Pr[\Delta = 1] &\sim (n-a)p_2[1 - (n-1-a)p_{2,c} + (n+\gamma)p_{3,c} + (n^2/2 + \delta n + \tau)p_{4,c} + \cdots] \\
&\sim n^\alpha(-1+n^\alpha)/n - (n^{4\alpha} + \mathcal{O}(n^{3\alpha}))/n^2 + \cdots \\
\Pr[\Delta = 2] &= \Pr\left[\sum_i\sum_{j>i} H_{2,i}H_{2,j}\prod_{k\neq(i,j)}[\mathbf{1} - H_{2,k}]\right] \\
&\sim [(n+\gamma)p_3 + (n^2/2 + \delta n + \tau)p_4][1 - (n-2-a)p_{2,c} + \cdots] \\
&\sim n^\alpha(-2 + 5n^\alpha - 4n^{2\alpha} + n^{3\alpha})/(2n^2) + \cdots
\end{aligned}$$

Note that we have neglected the fact that $H_{2,i}$ has some correlation with $H_{2,i+a}$ and $H_{2,i-a}$. This will be taken into account in Sec. 3.1.3.

3.1.2 $\alpha = 1/2$

As $p_2 \sim 1/n$, we see that we are in the realm of asymptotic Poisson distribution, with mean 1. This can be checked by the method of moments (see, for instance, Kolchin et al. [10], Barbour et al. [2]). First, by Poissonization-dePoissonization as exposed in the nice paper by Jacquet and Szpankowski, [8], we can again consider all urns as independent. We obtain

$$\begin{aligned}
M &\sim 1, \\
\mu_2 &\sim 2, \\
\mu_3 &:= \mathbb{E}(\Delta^3) \\
&= (n-a)p_2 + (3n^2 + \beta_1 n + \beta_2)p_4 + (n^3 + \beta_3 n^2 + \beta_4 n + \beta_5)p_6 + (6n + \beta_7)p_3 + (6n + \beta_8)p_4,
\end{aligned}$$

where $\beta_\bullet$ are unimportant constants. This leads to $\mu_3 \sim 5$, which is the Poisson(1) third moment. The generalization to μ_i is tedious but obvious.

3.1.3 $\alpha > 1/2$

As $np_2 \gg 1$, this leads to a Gaussian distribution, with mean M and variance σ^2. This can be checked as follows. For simplicity, we assume that n is an integer multiple of a. Let us divide the urns into

a subsets $\Omega_i = [U(i), U(i+a)\cdots], i = 0\ldots a-1$. Again, asymptotically, we can consider the urns as independent, with occupation probability p_1. The a subsets are obviously independent. In each subset Ω_i, we obtain a Markov chain on doublets, with states $[0,0],[0,1],[1,0],[1,1]$, where the urn index represents the (discrete) time of the chain, with transition probability matrix ($q_1 := 1-p_1$):

$$\begin{pmatrix} q_1 & p_1 & 0 & 0 \\ 0 & 0 & q_1 & p_1 \\ q_1 & p_1 & 0 & 0 \\ 0 & 0 & q_1 & p_1 \end{pmatrix}$$

The stationary measure is given by $(q_1^2, q_1p_1, q_1p_1, p_1^2)$. Let Δ_i be the total occupation time of state $[1,1]$ in Ω_i. The stationay increase in Δ_i is given by $q_1p_1.p_1 + p_1^2.p_1 = p_1^2$ as expected. Δ is now given by a sum of i.i.d RV $\Delta_i, i = 0\ldots a-1$, Set $n^* := n/a$. Our Markov chain is finite, φ-mixing and we can apply the Functional Central Limit Theorem, Billingsley([3, p. 174, Thm. 20.1]). We obtain the following result, where $B(t)$ is the standard Brownian Motion (BM).

$$\frac{\Delta_i(\lfloor n^*t\rfloor) - M^*n^*t}{\sigma^*\sqrt{n^*}} \Rightarrow B(t), \quad n^* \to \infty,\ t\in[0,1],$$

where n^*M^* and $n^*\sigma^{*2}$ are the dominant terms of the mean $M(n^*)$ and variance $\sigma^2(n^*)$ of $\Delta_i(n^*)$. We have $M^* \sim p_1^2$ and $\sigma^{*2} \sim p_1^2$ by (7), (8). We conclude that Δ is indeed a Gaussian RV.

In the next Sections the 2-decomposability is analyzed in detail, when $n\to\infty$, with numerous asymptotic expressions. In particular, we introduce a random variable ND which is a measure of decomposability.

3.2 Preliminary analysis for asymptotic expressions.

We have here $J = \{0, a\}$. We first make an heuristic analysis, in order to extract the main characteristics of our constraints. For each a, let $Z_a :=$ [Some suitable combination exists], i.e:

$$\Pr[Z_a = 1] = \Pr[A^a \vee B^a \vee C^a \vee D^a],$$

and $Z_a = 1$ corresponds to *non-decomposability* according to $\{0,a\}$. The constraints are given by (value 0 corresponds to an empty urn and 1 to an occupied urn)

$$\begin{aligned}
A^a &:= \vee A_i^a, \quad \text{with} \quad A_i^a := [U[i-a]=1 \wedge U[i]=0 \wedge U[i+a]=1],\\
B^a &:= \vee B_i^a, \quad \text{with} \quad B_i^a := [U[i-a]=1 \wedge U[i]=0],\\
C^a &:= \vee C_i^a, \quad \text{with} \quad C_i^a := [U[i]=0 \wedge U[i+a]=1],\\
D^a &:= \vee D_i^a, \quad \text{with} \quad D_i^a := [U[i]=0].
\end{aligned}$$

Let us look for the possible ranges for a and i. From (5), and defining an exclusion range $\rho(a)$ where no urn can be occupied, we obtain

$$\begin{aligned}
A_i^a &: \text{ we must have } 2a \le n-1, \quad i\in[a, n-1-a],\\
B_i^a &: \text{ if } \ 2a \le n, \quad i\in[n-a, n-1], \text{ with an exclusion range size } \rho(a)\sim a,\\
&: \text{ if } \ 2a > n, \quad i\in[a, n-1], \text{ with an exclusion range size } \rho(a)\sim n-a,\\
C_i^a &: \text{ if } \ 2a \le n, \quad i\in[0, a-1], \text{ with an exclusion range size } \rho(a)\sim a,\\
&: \text{ if } \ 2a > n, \quad i\in[0, n-1-a], \text{ with an exclusion range size } \rho(a)\sim n-a,\\
D_i^a &: \text{ if } \ 2a > n, \quad i\in[n-a, a-1].
\end{aligned}$$

Set

$$Y_a := [Z_a = 0] = 1 - Z_a,$$

and

$$ND := \sum_{a=1}^{n-1} Y_a,$$

i.e., Non-decomposability $\equiv [ND = 0]$. Actually, ND is a measure of decomposability.

Now we first make an heuristic analysis. If we neglect the events $[U[i] = 0]$ (which each have Probability ~ 1), and if we neglect also the possible A, B and C overlappings, we see that

$$\Pr[B^a \vee C^a] \sim 1 - (1 - 2\rho(a)/n)^{n^\alpha} \sim 1 - e^{-2\rho(a)n^\alpha/n},$$

which means that

$$\begin{aligned}\Pr(B^a \vee C^a) \sim 0, \text{ if } \quad 2\rho(a) < n^{1-\alpha},\\ \Pr(B^a \vee C^a) \sim 1, \text{ if } \quad 2\rho(a) > n^{1-\alpha}.\end{aligned}$$

Our 3 usual α ranges are roughly analyzed as follows: (and we assume, in first approximation, independence between all Z_a),

- $\alpha < 1/2, a < n/2 : \rho(a) = a$. So we set $a_1^* := n^{1-\alpha}/2$.
  By Sec 3.1.1, $\Pr(A^a) \sim 0$, and

  $$\begin{aligned}\Pr[B^a \vee C^a] &\sim 0, \quad \text{if} \quad a < a_1^*,\\ \Pr[B^a \vee C^a] &\sim 1, \quad \text{if} \quad a_1^* < a < n/2,\end{aligned}$$

  $\alpha < 1/2, a > n/2 : \rho(a) = (n - a)$. So we set $a_2^* := n - n^{1-\alpha}/2$.

  $$\begin{aligned}\Pr[B^a \vee C^a] &\sim 1 \quad \text{if} \quad n/2 < a < a_2^*,\\ \Pr[B^a \vee C^a] &\sim 0, \quad \Pr[D^a] \sim 1, \quad \text{if} \quad a > a_2^*.\end{aligned}$$

  Hence $ND \sim n^{1-\alpha}/2$. More precisely, with

  $$\begin{aligned} m &:= \mathbb{E}(ND) = \sum_{a=1}^{n/2} e^{-2\rho(a)n^\alpha/n} \sim \int_0^\infty e^{-2\rho(a)n^\alpha/n} da \sim n/(2n^\alpha),\\ \sigma^2 &:= \mathrm{VAR}(ND) = \sum_{a=1}^{n/2} e^{-2\rho(a)n^\alpha/n}(1 - e^{-2\rho(a)n^\alpha/n})\\ &\sim \int_0^\infty e^{-2\rho(a)n^\alpha/n}(1 - e^{-2\rho(a)n^\alpha/n}) da \sim n/(4n^\alpha),\end{aligned}$$

  we have $ND \sim \mathcal{N}(m, \sigma^2)$.

- $\alpha = 1/2, a < n/2 : \rho(a) = a$. So we set $a_1^* := \sqrt{n}/2$.
  By Sec. 3.1.2, for fixed a, $\Pr(\neg A^a) \sim e^{-1}$.

  $$\begin{aligned}\Pr[B^a \vee C^a] &\sim 0, \quad \text{if} \quad a < \sqrt{n}/2,\\ \sum_{a=1}^{\sqrt{n}/2} [Z_a = 0] &\sim Bin(\sqrt{n}/2, e^{-1}), \quad \text{if} \quad a < \sqrt{n}/2, \quad \text{taking only } A \text{ into account},\\ \Pr[B^a \vee C^a] &\sim 1 \quad \text{if} \quad \sqrt{n}/2 < a < n/2.\end{aligned}$$

  $\alpha = 1/2, a > n/2 : \rho(a) = (n - a)$. So we set $a_2^* := n - \sqrt{n}/2$.

  $$\begin{aligned}\Pr[B^a \vee C^a] &\sim 1, \quad \text{if} \quad n/2 < a < n - \sqrt{n}/2,\\ \Pr[B^a \vee C^a] &\sim 0, \quad \Pr[D^a] \sim 1, \quad \text{if} \quad a > n - \sqrt{n}/2.\end{aligned}$$

  Hence $ND \sim Bin(\sqrt{n}/2, e^{-1})$, or $ND \sim \mathcal{N}(\sqrt{n}/2e^{-1}, \sqrt{n}/2e^{-1}(1 - e^{-1}))$.

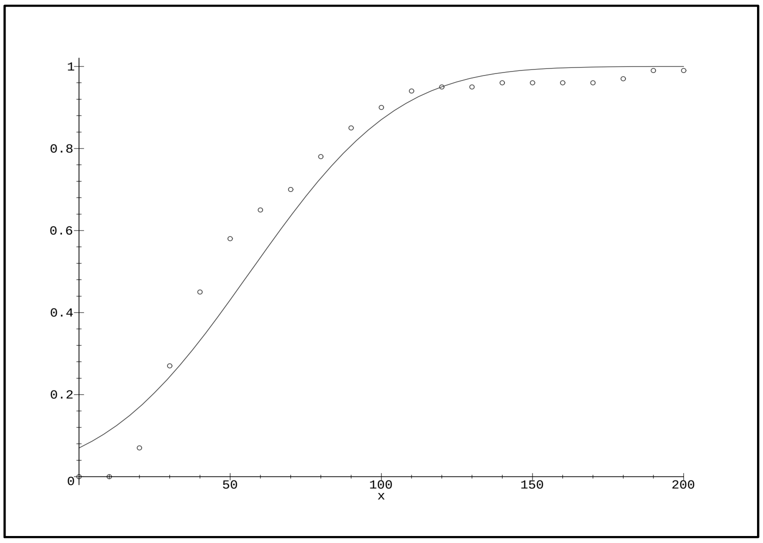

Figure 1: Observed and Gaussian ND distribution function (with observed parameters), $\alpha = .45$

- $\alpha > 1/2, a < n/2 : \Pr(A^a) \sim 1$ for fixed a, by Sec 3.1.3, and, with $a_1^* := n^{1-\alpha}/2$,

$$\begin{aligned}\Pr[B^a \vee C^a] &\sim 0, \quad \Pr(A^a) \sim 1, \quad \text{if} \quad a < a_1^*,\\ \Pr[B^a \vee C^a] &\sim 1, \quad \text{if} \quad a_1^* < a < n/2.\end{aligned}$$

  $\alpha > 1/2, a > n/2 : \rho(a) = (n-a)$. So we set $a_2^* := n - n^{1-\alpha}/2$.

$$\begin{aligned}\Pr[B^a \vee C^a] &\sim 1, \quad \text{if} \quad n/2 < a < a_2^*,\\ \Pr[B^a \vee C^a] &\sim 0, \quad \Pr[D^a] \sim 1, \quad \text{if} \quad a > a_2^*.\end{aligned}$$

  Hence $ND \sim 0$ if $\alpha > 1/2$.

$\alpha = 1/2$ corresponds to a critical value, where we observe a phase transition.

In the first 2 cases, Pr[Non-decomposability] ~ 0, in the third case, Pr[Non-decomposability] ~ 1 . The simulations lead to the following observations. For $n = 10000, \alpha = .45$ of $T = 100$ trials, we obtain Figure 1, with observed =circle, asymptotic Gaussian=line. The observed mean is 56.69, the asymptotic mean is 79, the observed variance is 1474. The fit is unsatisfactory. Obviously, we cannot ignore dependencies.

4 Main results on the ND distribution

In this section, we present the main results of our paper on the ND distribution, the proofs will be given in the next sections. Set

$$\begin{aligned}Q(j) &:= \prod_{k=1}^{j}(1-2^{-k}), \quad Q(0) = 1,\\ Q(\infty) &= 0.288788095087\ldots, \quad Q^-(j) := 1/Q(j),\\ C(j) &:= \frac{(-1)^j}{2^{j(j+1)/2}} Q^-(j),\\ K(n,\alpha) &:= \frac{e^{-np_1^2}}{p_1}, \text{ with } p_1 \text{ given in (6), } p_1 \sim n^{\alpha-1},\end{aligned}$$

$$\begin{aligned}
\beta &:= K(n,\alpha)/4, \\
Z_a &:= \text{[Some suitable combination exists], i.e:} \\
\Pr[Z_a = 1] &= \Pr[A^a \vee B^a \vee C^a \vee D^a], \\
Y_a &:= [Z_a = 0] = 1 - Z_a, \\
ND &:= \sum_{a=1}^{n-1} Y_a.
\end{aligned}$$

Non-decomposability $\equiv [ND = 0]$. Actually, ND is a measure of decomposability.

If $\beta < 1$, we have the following result

Proposition 4.1

$$\Pr[ND = k] \sim (-1)^k \sum_{i=k}^{\infty} (-1)^i \binom{i}{k} \beta^i Q^-(i). \tag{9}$$

For all β, we have

Proposition 4.2

$$\Pr[ND = k] \sim Q^-(\infty) \sum_{j=0}^{\infty} C(j) \frac{(\beta/2^j)^k}{(1+\beta/2^j)^{k+1}}. \tag{10}$$

Set

$$s_i = \sum_{j=0}^{\infty} \frac{1}{(\beta + 2^j)^i},$$

we obtain

$$\Pr[ND = 0] \sim \prod_{k=0}^{\infty} \frac{1}{1+\beta/2^k}, \tag{11}$$

and, for all β,

Proposition 4.3

$$\begin{aligned}
\Pr[ND = k] &\sim \beta^k \Pr[ND = 0] \sum_{\alpha_1 + 2\alpha_2 + \ldots + k\alpha_k = k} \frac{1}{\alpha_1!\alpha_2!\ldots\alpha_k! 2^{\alpha_2}\ldots k^{\alpha_k}} s_1^{\alpha_1} s_2^{\alpha_2} \ldots s_k^{\alpha_k} \\
&\sim \beta^k \Pr[ND = 0] \sum_{\alpha_1 + 2\alpha_2 + \ldots + k\alpha_k = k} \frac{1}{\alpha_1!\alpha_2!\ldots\alpha_k!} \left(\frac{s_1}{1}\right)^{\alpha_1} \left(\frac{s_2}{2}\right)^{\alpha_2} \cdots \left(\frac{s_k}{k}\right)^{\alpha_k}.
\end{aligned} \tag{12}$$

The factorial moments of (10) are given by

$$\tilde{m}_l = l!\beta^l Q^-(l), \tag{13}$$

5 A more precise approach. The correlation problem

In this section, we analyze the range overlappings, the inter-level correlations and the effect of empty urns. The exclusion ranges will be denoted by $\rho(a)$. As, from Sec.3.2, we only need to consider the range $[0, a_1^*]$, we will limit ourselves to the case $a < n/6$. We will use the following notations:

$$\begin{aligned}
P^A(a) &:= \Pr^A[Z_a = 0] = \Pr[\neg A^a], \\
P^{BC}(a) &:= \Pr^{B\vee C}[Z_a = 0] = \Pr[\neg(B^a \vee C^a)]
\end{aligned}$$

5.1 Ranges overlapping for $Z_a = 0$

Let us first look (for fixed a), for the possible overlappings of $B \vee C$ ranges. We neglect, in first approximation, the events $U[.] = 0$, and we concentrate on the *exclusion* of the events $U[k] = 1$. A more precise analysis is provided later on. After a detailed evaluation, we see that 4 different cases must be considered for k.

- $a < n/4$: $U[k] = 1$ *must be excluded* for $k \in [a, 2a] \cup [n-2a, n-a]$, and the joint exclusion range is given by $\rho^{BC}(a) = 2a$,
- $n/4 < a < n/3$: $U[k] = 1$ *must be excluded* for $k \in [a, n-a]$, and the joint exclusion range is given by $\rho^{BC}(a) = n - 2a$,
- $n/3 < a < n/2$: $U[k] = 1$ *must be excluded* for $k \in [n-2a, 2a]$, and the joint exclusion range is given by $\rho^{BC}(a) = 4a - n$,
- $n/2 < a$: $U[k] = 1$ *must be excluded* for $k \in [0, n-a] \cup [a, n]$, and the joint exclusion range is given by $\rho^{BC}(a) = 2(n-a)$.

Now we look for the overlapping between $B \vee C$ and A. Again we first assume that the events $[U[i] = 0]$ can be neglected. Given the constaints due to $B \vee C$, the different ranges for the *exclusion* of $[U[i+a] = 1 \wedge U[i-a] = 1]$ are given by (we give only the first two ranges, the other ones are just more complicated and will not be used in the sequel).

- $a < n/6$: $\rho^A(a) = n - 4a$,
- $n/6 < a < n/5$: $\rho^A(a) = 2a$.

5.2 The events $U[.] = 0$ in A

Let $p_{j,k}$ correspond to the event: [j occupied urns and k empty urns, all distinct]. We have

$$\begin{aligned}
\Pr[A_i^a] &= p_{2,1} = \Pr[V(\mathbf{1}-V)^2] = PV[1] - 2PV[2] + PV[3] \\
&= n^\alpha(-1+n^\alpha)/n^2 + n^\alpha(-4+6n^\alpha-2n^{2\alpha})/n^3 + \cdots \sim p_2 \sim p_1^2.
\end{aligned}$$

Similarly, $p_{3,2} = \Pr[V^2(\mathbf{1}-V)^3]$, $\quad p_{4,2} = \Pr[V^2(\mathbf{1}-V)^4]$.
As clear from Sec.3.2, we can limit our analysis to the range $[0, a_1^*]$, and the exclusion range $\rho^A(a) = n - 4a = \mathcal{O}(n)$.

- If $\alpha < 1/2$, we obtain by inclusion-exclusion

$$\begin{aligned}
P^A(a) &= 1 - \rho^A(a)p_{2,1} + (\rho^A(a) - 2a)^+ p_{3,2} \\
&+ \left(\binom{\rho^A(a)}{2} - (\rho^A(a)-2a)^+ - (\rho^A(a)-a)^+\right) p_{4,2} + \cdots \\
&\sim 1 - \rho^A(a)p_1^2 \sim e^{-\rho^A(a)p_1^2}.
\end{aligned} \tag{14}$$

- If $\alpha = 1/2$, the Poisson approximation leads to

$$P^A(a) \sim e^{-\rho^A(a)p_{2,1}} \sim e^{-\rho^A(a)p_1^2} \sim e^{-\rho^A(a)/n} \sim e^{-1}. \tag{15}$$

- If $\alpha > 1/2$, we have

$$\begin{aligned}
P^A(a) &= \Pr\left[\prod_{j \in \rho^A(a)} [\mathbf{1} - \nu_j[\mathbf{1}-\nu_{j+a}][\mathbf{1}-\nu_{j-a}]]\right] \\
&= \Pr\left[\prod_{j \in \rho^A(a)} [\mathbf{1} - [\mathbf{1}-\nu_{j+a}][\mathbf{1}-\nu_{j-a}] + [\mathbf{1}-\nu_j][\mathbf{1}-\nu_{j+a}][\mathbf{1}-\nu_{j-a}]]\right].
\end{aligned}$$

Set

$$\begin{aligned} F_j &:= \mathbf{1} - [\mathbf{1} - \nu_{j+a}][\mathbf{1} - \nu_{j-a}], \\ G_j : &= [\mathbf{1} - \nu_j][\mathbf{1} - \nu_{j+a}][\mathbf{1} - \nu_{j-a}], \end{aligned}$$

which are *disjoints* events. We obtain

$$P^A(a) = \Pr\left[\prod_j F_j + \sum_i G_i \prod_{j \neq i} F_j + \cdots\right].$$

Let us first deal with $\Psi_1 := \Pr[\prod_j F_j]$. We divide the range $[0..\rho^A(a)]$ into $2a$ subranges $\{i, i+2a, i+4a, \cdots\}, i = 0..2a-1$. In each range, of size $\lfloor \rho^A(a)/(2a) \rfloor = \mathcal{O}(n^\alpha)$, (in the neighbourhood of a_1^*), we must exclude the events $[U[k] = 1 \wedge U[k+1] = 1]$, but, as $p_1 = \Pr[U[k] = 1] \ll 1$, we can safely consider all urns as asymptotically independent.

Let us denote by $P(m)$ the probability that there does not appear 2 successive occupied urns among m urns. This is a well known problem. We have (with $q_1 := 1 - p_1$),

$$P(m) = p_1 q_1 P(m-2) + q_1 P(m-1), \text{with} \quad P(0) = P(1) = 1.$$

The generating function(GF) of $P(m)$ is given by $F(z) := \sum_0^\infty P(m) z^m$. We obtain

$$F - 1 - z = p_1 q_1 z^2 F + q_1 z (F - 1),$$

or

$$F := \frac{1 + z p_1}{1 - p_1 z^2 + p_1^2 z^2 - z + p_1 z}. \tag{16}$$

We denote by z_1^* the root of the denominator of (16), with smallest module, (the other one is denoted by z_2^*). This leads to

$$F = \frac{1 + z p_1}{p_1 q_1 z_1^* (z - z_2^*)(1 - z/z_1^*)},$$

and by standard singularity analysis, we derive $P(m) \sim C/z_1^{*m}$, with $C \sim 1 + p_1^2, \quad 1/z_1^* \sim 1 - p_1^2 + p_1^3 + \cdots$. This leads to $P(m) \sim e^{-mp_1^2}$, which is probabilistically obvious. So finally,

$$\Psi_1 \sim e^{-\rho^A(a) p_1^2} \ll 1. \tag{17}$$

Now we consider $\Psi_2 := \sum_i G_i \prod_{j \neq i} F_j$. Asymptotically,

$$\Psi_2 \sim \rho^A(a) p_3 e^{-(\rho^A(a)-1)p_1^2} = e^{-(\rho^A(a)-1)p_1^2} \mathcal{O}(n^{3\alpha}/n^2).$$

If $\alpha < 2/3$ then $\Psi_2 = o(\Psi_1)$, and $P^A(a) \sim \Psi_1 + \Psi_2 \sim \Psi_1 \sim e^{-\rho^A(a) p_1^2}$. If $\alpha \geq 2/3$, by a similar analysis we obtain $P^A(a) \sim e^{-\rho^A(a) p_{2,1}}$.

5.3 The events $U[.] = 0$ in $B \vee C$

- $\alpha < 1/2$. Let us first look for the simplest case: $a < n/4$. We have

$$\begin{aligned} P^{BC}(a) &= \Pr\left[\prod_{i=0}^{a} [\mathbf{1} - \nu_i[\mathbf{1} - \nu_{i+a}]] \prod_{i=n-a}^{n} [\mathbf{1} - \nu_i[\mathbf{1} - \nu_{i-a}]]\right] \\ &= \Pr\left[\prod_{i=0}^{a} [\nu_{i+a} + [\mathbf{1} - \nu_i][\mathbf{1} - \nu_{i+a}]] \prod_{i=n-a}^{n} [\nu_{i-a} + [\mathbf{1} - \nu_i][\mathbf{1} - \nu_{i-a}]]\right]. \end{aligned}$$

Set

$$\begin{aligned} F_j &:= \nu_j \\ G_j^{(1)} &:= [\mathbf{1}-\nu_{j-a}][\mathbf{1}-\nu_j], \quad G_j^{(2)} := [\mathbf{1}-\nu_j][\mathbf{1}-\nu_{j+a}], \end{aligned}$$

which are again *disjoint* events. Proceeding as before, we see that

$$\begin{aligned} \Psi_1 &:= \Pr\left[\prod_{j=a}^{2a}\prod_{j=n-2a}^{n-a} F_j\right] \sim e^{-2ap_1}, \\ \Psi_2 &:= \Pr\left[\sum_{j=a}^{2a} G_j^{(1)}\prod_{j\neq i} F_j + \sum_{j=n-2a}^{n-a} G_j^{(2)}\prod_{j\neq i} F_j\right] \sim 2ap_2 e^{-2ap_1} = \mathcal{O}(n^{2\alpha}/n)e^{-2ap_1} = o(\Psi_1). \end{aligned}$$

So $P^{BC}(a) \sim \Psi_1 \sim e^{-2ap_1}$.

To show the increasing complexity of the asymptotics, let us consider the case $n/4 < a < n/3$. (But we will not need these results in the sequel). After a detailed analysis, this leads to

$$P^{BC}(a) = \prod_{j=a}^{n-2a}\left[F_j + G_j^{(1)}\right] \cdot \prod_{j=2a}^{n-a}\left[F_j + G_j^{(2)}\right] \cdot \prod_{j=n-2a}^{2a}\left[F_j + G_j^{(3)}\right],$$

where

$$G_j^{(3)} := [\mathbf{1}-\nu_j][\mathbf{1}-\nu_{j+a}][\mathbf{1}-\nu_{j-a}].$$

This leads to

$$\begin{aligned} \Psi_1 &\sim e^{-(n-2a)p_1} \\ \Psi_2 &\sim e^{-(n-2a-1)p_1}[2(n-3a)p_2 + (4a-n)p_3] = o(\Psi_1). \end{aligned}$$

Therefore, $P^{BC}(a) \sim e^{-(n-2a)p_1}$.

- $\alpha = 1/2$. We limit ourselves to the range $[0, a_1^* = \sqrt{n}/2]$. By a similar analysis, we obtain $P^{BC}(a) \sim e^{-2a/\sqrt{n}}$.
- $\alpha > 1/2$. In the range $[0, a_1^*]$, we derive $P^{BC}(a) \sim e^{-2ap_1}$.

5.4 Dependencies between the different a levels

Again, as we use only the range $[0, a_1^*]$, we will consider the case $a < n/6$. From Sec. 5.3, we must have $Z_a^C = 0$, i.e., asymptotically, no occupied urn in $[a, 2a]$, (and similarly for B). If we consider some level $b > a$, this implies that, if $b < 2a$, only the urns between $2a$ and $2b$ must be empty, which leads to a conditional probability

$$P^{BC}(b|a) := Pr^{B\vee C}[Z_b = 0|Z_a = 0] \sim e^{-2(2b-2a)p_1}.$$

If $b \geq 2a$, no correlation exists for BC

Let us now turn to the dependency of A on BC.

$$\begin{aligned} &\text{If} \quad b \leq 2a, \text{then} \quad \rho^A(b|a) \sim n - 6b + 2a, \\ &\text{If} \quad b \geq 2a, \text{then} \quad \rho^A(b|a) \sim n - 4b - 2a, \end{aligned} \tag{18}$$

and $P^A(b|a) := \Pr^A[Z_b = 0|Z_a = 0]$ is given by (14) or (15) or (17), (depending on α), with now $\rho^A(b|a)$ instead of $\rho^A(a)$.

So our different cases are analyzed as follows (we can of course neglect the correlation between A at a and A at b).

- $\alpha < 1/2$. The mean of ND is asymptotically given by

$$m \sim \sum_{a=1}^{n/6} e^{-2ap_1} P^A(a) \sim \frac{e^{-np_1^2}}{2p_1(1-2p_1)} \sim \frac{1}{2p_1}, \tag{19}$$

where $P^A(a)$ is given by (14), with $\rho^A(a) = n - 4a$.

$$m_2 := \mathbb{E}(ND^2) = m + 2S_2,$$

where $S_2 := \mathbb{E}(\sum_a \sum_{b>a} Y_a Y_b)$, and

$$\begin{aligned} S_2 \quad &\sim \quad \sum_{a=1}^{n/12} e^{-2ap_1} P^A(a) \left[\sum_{b=a+1}^{2a} e^{-2(2b-2a)p_1} P^A(b|a) + \sum_{b=2a+1}^{n/6} e^{-2bp_1} P^A(b|a) \right] \\ &+ \quad \sum_{a=n/12+1}^{n/6} e^{-2ap_1} P^A(a) \sum_{b=a+1}^{n/6} e^{-2(2b-2a)p_1} P^A(b|a) \end{aligned} \tag{20}$$

where $P^A(b|a)$ is given by (14), (with $\rho^A(b|a)$), $\rho^A(b|a)$ is given by (18). This leads to

$$S_2 \sim \frac{e^{-2np_1^2}}{6p_1^2} \sim \frac{1}{6p_1^2},$$

and

$$\sigma^2 := \mathrm{VAR}(ND) = m_2 - m^2 \sim \frac{1}{12p_1^2}.$$

- $\alpha = 1/2$. $p_1 \sim 1/\sqrt{n}, p_2 \sim 1/n$ and , by (15), $P^A(a) \sim e^{-(n-4a)/n}$. m and S_2 are given by (19) and (20), with

$$P^A(b|a) \sim e^{-\rho^A(b|a)/n},$$

and $\rho^A(b|a)$ is given by (18). This leads to $m \sim \frac{e^{-1}}{2p_1}$ and $S_2 \sim \frac{e^{-2}}{6p_1^2}$.

- $\alpha > 1/2$. This is more delicate and we observe a phase transition: $\alpha = 1/2$ is the critical value.

We now have

$$m \sim \sum_{a=1}^{n/6} e^{-2ap_1} e^{-(n-4a)p_1^2} \sim \int_{a=0}^{\infty} e^{-2ap_1} e^{-(n-4a)p_1^2} da = \frac{e^{-np_1^2}}{2(p_1 - 2p_1^2)} \sim \frac{e^{-np_1^2}}{2p_1}.$$

We also have $ND \sim \sum_{a=1}^{n/6} Y_a$, and we must compute the joint distribution of $Y_i, Y_j, \ldots$ So we obtain, with

$$S_1 := m, \quad S_i := \mathbb{E}(\sum_{a_1} \sum_{a_2 > a_1} \cdots \sum_{a_i > a_{i-1}} Y_{a_1} Y_{a_2} \cdots Y_{a_i}),$$

$$\Pr[ND = 0] = \Pr[(\mathbf{1} - Y_1)(\mathbf{1} - Y_2) \cdots (\mathbf{1} - Y_{n-a})] \sim 1 - m + S_2 - S_3 + \cdots = \sum_0^{\infty} (-1)^i S_i \tag{21}$$

Note that if all Y_i were independent, we should derive

$$S_2 \sim m^2/2, S_3 \sim m^3/3!, \cdots$$

and $\Pr[ND = 0]$ would be given by e^{-m} which is exactly the Poisson approximation.

Actually, we have

$$S_2 \sim \sum_{a=1}^{n/6} \varphi(a) \sum_{b=a+1}^{n/6} \psi(b|a),$$

with

$$\begin{aligned}
\varphi(a) &= e^{-2ap_1}e^{-(n-4a)p_1^2}, \\
\psi(b|a) &= e^{-2bp_1}P^A(b|a), \text{if} \quad b \geq 2a, \\
\psi(b|a) &= e^{-2(2b-2a)p_1}P^A(b|a), \text{if} \quad b \leq 2a.
\end{aligned} \tag{22}$$

$P^A(b|a)$ (with $\rho^A(b|a)$) is given by (17) and $\rho^A(b|a)$ is given by (18). Asymptotically, we obtain (as $p_1^2 \ll p_1$)

$$S_2 \sim \int_{a=1}^{n/6-1} \varphi(a)da \int_{b=a+1}^{n/6} \psi(b|a)db \sim \int_{b=0}^{\infty} db \int_{a=0}^{b} \varphi(a)\psi(b|a)da \sim \frac{e^{-2np_1^2}}{6p_1^2}.$$

More generally, set

$$\begin{aligned}
\rho^A(x_i|x_1\cdots x_{i-1}) &:= \text{ exclusion range for } A \text{ at level } i, \text{ given all values at previous levels,} \\
\psi(x_i|x_1\cdots x_{i-1}) &:= e^{-2x_ip_1}e^{-\rho^A(x_i|x_1\cdots x_{i-1})p_1^2}, \text{if} \quad x_i \geq 2x_{i-1}, \\
\psi(x_i|x_1\cdots x_{i-1}) &:= e^{-2(2x_i-2x_{i-1})p_1}e^{-\rho^A(x_i|x_1\cdots x_{i-1})p_1^2}, \text{if} \quad x_i \leq 2x_{i-1}, \\
\Theta_i(x_1\cdots x_i) &:= 1, \\
\Theta_{l-1}(x_1\cdots x_{l-1}) &:= \int_{x_l=x_{l-1}}^{\infty} \psi(x_l|x_1\cdots x_{l-1})\Theta_l(x_1\cdots x_l)dx_l, l>1.
\end{aligned} \tag{23}$$

Then

$$S_i \sim \int_0^{\infty} \varphi(x_1)\Theta_1(x_1)dx_1.$$

The numerical values grow too rapidly with n. Moreover, Expressions for $\rho^A(.)$ seem difficult to compute.
We also have

$$\Pr[ND=1] = m - 2S_2 + 3S_3 + \cdots = \sum_{i=1}^{\infty}(-1)^{i+1}iS_i \tag{24}$$

and, again, the Poisson approximation (if all Y_i were independent) would lead to $e^{-m}m$. More generally, we have the *inclusion-exclusion principle*, also called Jordan's formula

Proposition 5.1

$$\Pr[ND=k] = (-1)^k\sum_{i=k}^{\infty}(-1)^i\binom{i}{k}S_i. \tag{25}$$

6 Asymptotic results

In this section, we present and prove several asymptotic expressions for $\Pr(ND=k)$.

6.1 First asymptotic expressions

In this section, we give a first asymptotic expression for $\Pr[ND=k]$. Let us first remark that in the term $e^{-(n-4a)p_1^2}$ of $\varphi(a)$ (as given by (22) and in the term $e^{-\rho^A p_1^2}$ of ψ (as given by (23)), we have a first contribution $e^{-np_1^2}$ and a second contribution of the form $exp[(\sum_i c_ix_i)p_1^2]$. Actually, as $p_1^2 \ll p_1$, this second part is negligible wrt the term e^{-2ap_1} of $\varphi(a)$ and wrt the term $e^{-2x_ip_1}$ or $e^{-2(2x_i-2x_{i-1})p_1}$ of ψ. Hence we obtain the equivalent

$$S_2 \sim \int_{b=0}^{\infty} db \int_{a=0}^{b} \varphi^*(a)\psi^*(b|a)da.\frac{e^{-2np_1^2}}{p_1^2},$$

where

$$\begin{aligned}
\varphi^*(a) &= e^{-2a}, \\
\psi^*(b|a) &= e^{-2b}, \text{if} \quad b \geq 2a, \\
\psi^*(b|a) &= e^{-2(2b-2a)}, \text{if} \quad b \leq 2a.
\end{aligned}$$

This leads to

$$S_2 \sim \frac{e^{-2np_1^2}}{6p_1^2}.$$

More generally, set

$$\begin{aligned}
\Theta_2^*(x_2) &:= \int_{x_1=0}^{x_2} \varphi^*(x_1)\psi^*(x_2|x_1)dx_1, \\
\Theta_i^*(x_i) &:= \int_{x_{i-1}=0}^{x_i} \psi^*(x_i|x_{i-1})\Theta_{i-1}^*(x_{i-1})dx_{i-1}, i > 2.
\end{aligned}$$

Then

$$\begin{aligned}
S_2 &\sim W_2 \frac{e^{-2np_1^2}}{p_1^2}, \text{with} \quad W_2 = \int_0^\infty \Theta_2^*(x_2)dx_2, \\
S_i &\sim W_i \frac{e^{-inp_1^2}}{p_1^i}, \text{with} \quad W_i = \int_0^\infty \Theta_i^*(x_i)dx_i. \qquad (26)
\end{aligned}$$

We obtain

$$W[0..\infty] = [1, 1/2, 1/6, 1/21, 4/315, 32/9765, 512/615195, 16384/78129765, \ldots].$$

An explicit expression for W_i is given by the following proposition:

Proposition 6.1 *For all $i \geq 0$, we have:*

$$W_i = \prod_{k=0}^{i-1} \frac{1}{2\left(2 - 2^{-k}\right)} = \frac{1}{2^{2i}} \prod_{k=1}^{i} \frac{2^k}{2^k - 1} = \frac{2^{i(i-3)/2}}{\prod_{k=1}^{i}(2^k - 1)}, \qquad (27)$$

where by convention an empty product equals 1 (*i.e.* $W_0 = 1$).

The proof is given in the appendix. Set

$$\begin{aligned}
Q(j) &:= \prod_{k=1}^{j}(1 - 2^{-k}), \quad Q(0) = 1, \\
Q(\infty) &= 0.288788095087\ldots, \quad Q^-(j) := 1/Q(j).
\end{aligned}$$

These functions appear in many analyses of algorithms (see, for instance, Louchard [11]). We obtain $W_i = Q^-(i)/4^i$. This leads to $W_i \sim Q^-(\infty)/4^i$. Hence (26) shows that we must have

$$K(n, \alpha) := \frac{e^{-np_1^2}}{p_1} < 4$$

in order to use (25) . We now have

$$S_i \sim W_i K^i; \qquad (28)$$

Equ.(25) becomes

$$\Pr[ND = k] \sim (-1)^k \sum_{i=k}^{\infty} (-1)^i \binom{i}{k} (K/4)^i Q^-(i). \qquad (29)$$

Which is Proposition 4.1. For $n = 10000, \alpha = .58$, we obtain $K(n, \alpha) = 0.608$. A simulation of $T = 100$ trials leads to Figure 2(with observed mean) . The asymptotic mean is given by 0.317, the observed mean is given by 0.44.

The Poisson approximation(with the observed mean) is given by Figure 3. Curiously enough, this is the best fit! (This is explained in the next section).

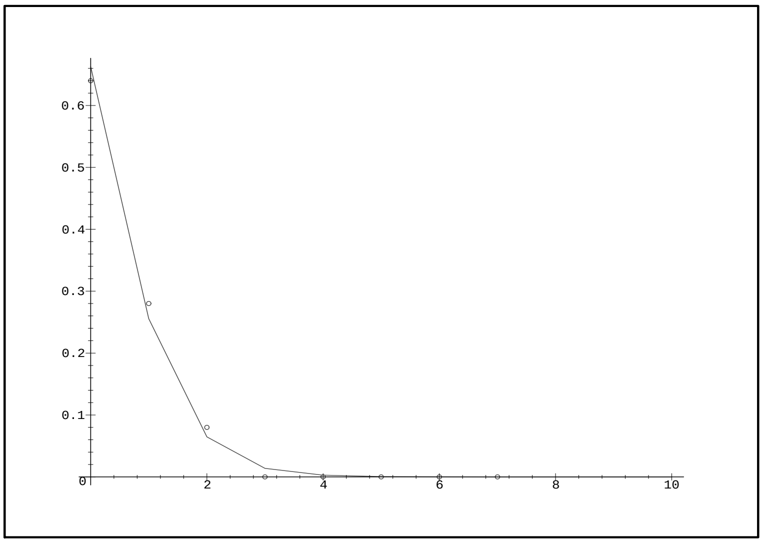

Figure 2: Observed and limiting (with (9) and observed mean) ND distribution , $\alpha = .58$

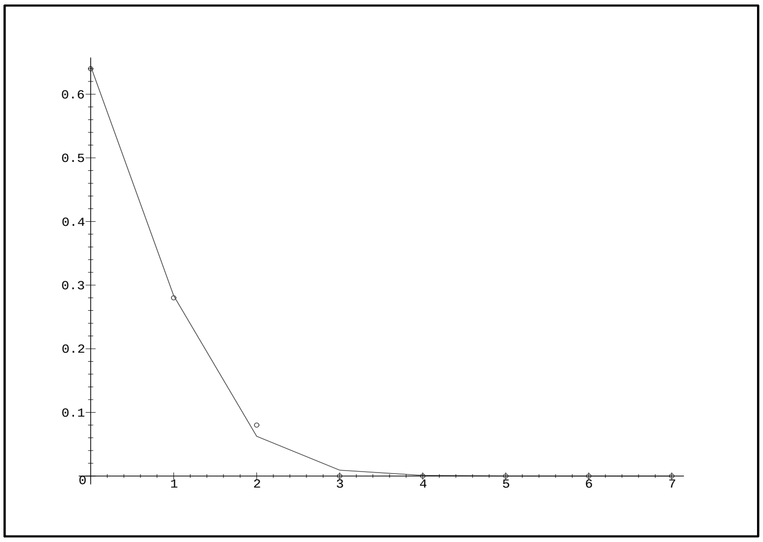

Figure 3: Observed and Poisson(with observed mean) ND distribution , $\alpha = .58$

6.2 More efficient asymptotic expressions

In this section, we present several more asymptotic expressions for $Pr[ND = k]$. The factorial moments are also analyzed.

First of all, we obtain

$$\Pr[ND = k] \sim Q^{-}(\infty) \sum_{j=0}^{\infty} C(j) \frac{(\beta/2^j)^k}{(1+\beta/2^j)^{k+1}}$$

which is Proposition 4.2. The proof is given in the appendix.

This series is absolutely convergent. Of course, (10) and (9) do have the same moments $m_k := \mathbb{E}(ND^k)$. But it is easy to check that, when $\beta > 1$, $m_k \sim k! S_k \sim k! \beta^k Q^{-}(k)$. Even when $\beta \gg 1$, these moments satisfy the relation

$$\limsup_{k\to\infty} \frac{1}{k} m_k^{1/k} < \infty.$$

This can be seen by using Stirling asymptotic formula. Hence, by Feller [6], p.228, the distribution is uniquely determined by its moments and (10) is valid for all β. Actually, we can derive a precise expression for the factorial moments of (10):

$$\begin{aligned}
\tilde{m}_l &:= \mathbb{E}(ND^{\underline{l}}) = Q^{-}(\infty) \sum_{j=0}^{\infty} C(j) \sum_{k=0}^{\infty} k^{\underline{l}} \frac{(\beta/2^j)^k}{(1+\beta/2^j)^{k+1}} \\
&= Q^{-}(\infty) \sum_{j=0}^{\infty} C(j) l! (\beta/2^j)^l \\
&= Q^{-}(\infty) l! \beta^l \prod_{k=l+1}^{\infty} (1-2^{-k}) = l! \beta^l Q^{-}(l),
\end{aligned}$$

by (42). This proves (13). In particular,

$$\begin{aligned}
\tilde{m}_0 &= 1, \\
\tilde{m}_1 &= m, \\
\tilde{m}_2 &= 4m^2/3, \\
\tilde{m}_3 &= 16m^3/7.
\end{aligned}$$

Remember that, by (28), $S_l \sim \beta^l Q^{-}(l)$.

Still another expression for $\Pr[ND = k]$ can be derived. First we obtain

$$\Pr[ND = 0] \sim \prod_{k=0}^{\infty} \frac{1}{1+\beta/2^k}$$

and, setting

$$s_i = \sum_{j=0}^{\infty} \frac{1}{(\beta + 2^j)^i},$$

we obtain

$$\begin{aligned}
\Pr[ND = k] &\sim \beta^k \Pr[ND = 0] \sum_{\alpha_1 + 2\alpha_2 + \ldots + k\alpha_k = k} \frac{1}{\alpha_1! \alpha_2! \ldots \alpha_k! 2^{\alpha_2} \ldots k^{\alpha_k}} s_1^{\alpha_1} s_2^{\alpha_2} \ldots s_k^{\alpha_k} \\
&\sim \beta^k \Pr[ND = 0] \sum_{\alpha_1 + 2\alpha_2 + \ldots + k\alpha_k = k} \frac{1}{\alpha_1! \alpha_2! \ldots \alpha_k!} \left(\frac{s_1}{1}\right)^{\alpha_1} \left(\frac{s_2}{2}\right)^{\alpha_2} \cdots \left(\frac{s_k}{k}\right)^{\alpha_k}
\end{aligned}$$

which is Proposition 4.3. The proofs are given in the appendix.

The number of terms in (12) is equal to the number of *partitions* of the integer k, which is known [1] to have the asymptotic value $\exp\left(\pi\sqrt{2k/3}\right) / \left(4k\sqrt{3}\right)$. For example, the first four terms are

$$
\begin{aligned}
\Pr[ND=1] &\sim \beta \Pr[ND=0]\, s_1, \\
\Pr[ND=2] &\sim \frac{\beta^2}{2}\Pr[ND=0]\left(s_2+s_1^2\right), \\
\Pr[ND=3] &\sim \frac{\beta^3}{6}\Pr[ND=0]\left(2s_3+3s_1s_2+s_1^3\right), \\
\Pr[ND=4] &\sim \frac{\beta^4}{24}\Pr[ND=0]\left(6s_4+8s_1s_3+3s_2^2+6s_1^2s_2+s_1^4\right).
\end{aligned}
\tag{30}
$$

7 A reanalysis of the different cases

We can now reanalyze our different cases, with appropriate asymptotics.

- $\alpha < 1/2$. $K \sim 1/p_1 \sim 2m, \beta = K/4$. The distribution function of ND is given, from (10) by

$$\Pr[ND \le k] \sim 1 - Q^{-}(\infty)\sum_{j=0}^{\infty} C(j)\left(\frac{\beta/2^j}{1+\beta/2^j}\right)^k. \tag{31}$$

We know that $\beta \gg 1$. Hence

$$\left(\frac{\beta/2^j}{1+\beta/2^j}\right)^k = \left(\frac{1}{1+2^j/\beta}\right)^k \sim (1-2^j/\beta)^k \sim e^{-2^j k/\beta}.$$

So we set $k = x\beta$ and we obtain

$$\Pr[ND \le k] \sim 1 - Q^{-}(\infty)\sum_{j=0}^{\infty} C(j)e^{-x2^j}. \tag{32}$$

The fit of (32) with the observations ($\alpha = 0.45$), with the observed mean in β is given in Figure 4. The fit is not completely ideal, but much better than the Gaussian approximation.

Still two other asymptotic results can be obtained.
i) From (11), we derive

$$L(\beta) := \ln(\Pr[ND=0]) \sim -\sum_{k=0}^{\infty}\ln(1+\beta/2^k),$$

and, by Euler-Mc Laurin, for large β,

$$\ln(\Pr[ND=0]) \sim -\int_0^{\beta}\ln(1+y)dy/(y\ln 2) - \ln(1+\beta)/2.$$

The first term leads to

$$\sim -\int_1^{\beta}\ln(z)dz/((z-1)\ln 2) \sim -\int_1^{\beta}\ln(z)dz/(z\ln 2) + C_1,$$

for some cst C_1. We finally obtain

$$\frac{-\ln^2(\beta)}{2\ln 2} + C_1 - \ln(1+\beta)/2$$

and

$$\Pr[ND=0] \sim C_2\beta^{-(\log_2(\beta)+1)/2}.$$

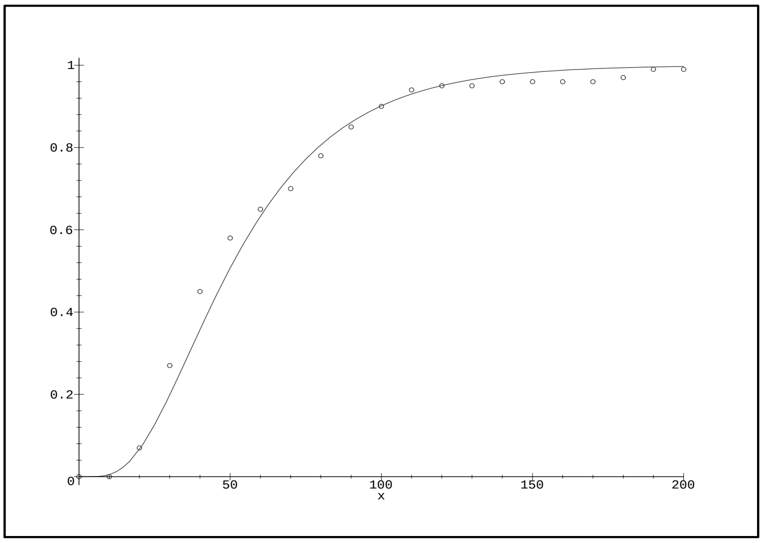

Figure 4: Observed and limiting (32) (with observed mean) ND distribution , $\alpha = .45$

Actually, we can get an equivalent of C_2 as follows: assume $\beta = 2^j$. We derive

$$\Pr[ND = 0] \sim \prod_{k=0}^{\infty} \frac{1}{1+\beta/2^k} = \frac{1}{2^{j(j+1)}} \frac{2^{j(j+1)/2}}{\prod_{k=0}^{j}(1+1/2^k)} \prod_{k=1}^{\infty} \frac{1}{1+1/2^k} \sim C_3^2/2.\beta^{-(\log_2(\beta)+1)/2}, \tag{33}$$

with

$$C_3 := \prod_{k=1}^{\infty} \frac{1}{1+1/2^k} = 0.419422441795\ldots$$

A still more precise analysis can be done. Indeed, $L(\beta)$ is a dyadic sum, which, for large β, can be analyzed using Mellin transforms: see Flajolet et al. [7]. It is well known that the dominant value is given by some function of $\log(\beta)$. The oscillatory part has a very small amplitude, usually of order 10^{-5}. Indeed, set $f(y) := \ln(1+y)$. We obtain

$$L(\beta) = -\sum_{k=0}^{\infty} f(1+\beta/2^k),$$

the Mellin transform of which is

$$L^*(s) = -\frac{\pi}{s\sin(\pi s)} \frac{1}{1-2^s}.$$

defined in the fundamental strip $< -1, 0 >$. To the right of this strip, the poles of $L^*(s)$ are a triple pole at $s = 0$, simple poles at $s = \chi_k := 2k\pi i/\ln 2 (k \neq 0)$ and simple poles at the positive integers. The singular expansion of $L^*(s)$ is given by

$$\begin{aligned} L^*(s) &\asymp \left[\frac{1}{\ln 2 s^3} - \frac{1}{2s^2} + \frac{\ln 2/12 + \pi^2/(6\ln 2)}{s}\right] \\ &+ \sum_{k\neq 0} \frac{-\pi}{\chi_k \sin(\pi\chi_k)} \frac{1}{\ln 2(s-\chi_k)} + \sum_{k=1}^{\infty} (-1)^k \frac{-1}{k(s-k)(1-2^k)}. \end{aligned}$$

This leads, by converse mapping, to

$$L(\beta) \sim \frac{-\ln^2(\beta)}{2\ln 2} - \ln(\beta)/2 + [-\ln 2/12 - \pi^2/(6\ln 2)] + \gamma_0(\beta) + \sum_{k=1}^{\infty} (-1)^k \frac{-1}{k\beta^k(1-2^k)}, \tag{34}$$

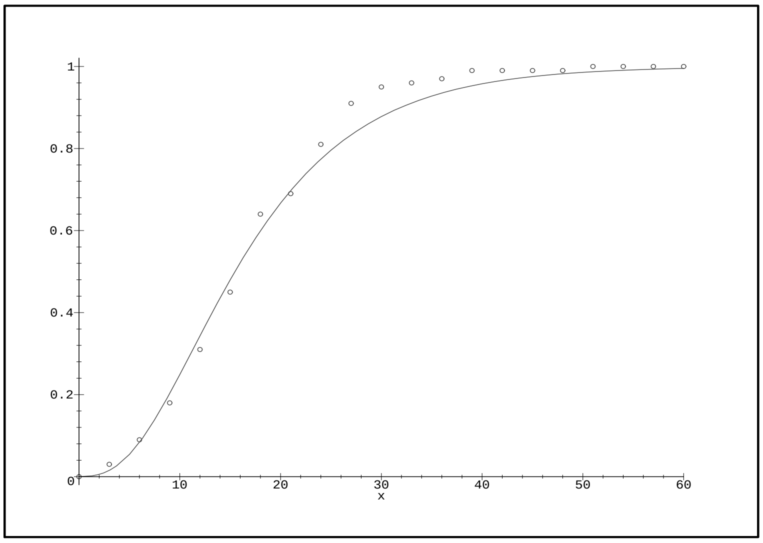

Figure 5: Observed and limiting (31) (with observed mean) ND distribution , $\alpha = .5$

where γ_0 is a small periodic function of $\log_2(\beta)$:

$$\gamma_0(\beta) := \sum_{k\neq 0} \frac{-\pi}{\chi_k \sin(\pi\chi_k)} \frac{\beta^{-\chi_k}}{\ln 2} = \sum_{k\neq 0} \frac{-\pi}{\chi_k \sin(\pi\chi_k)} \frac{e^{-2\pi i k \log_2(\beta)}}{\ln 2}$$

Here, the amplitude of γ_0 is of order 10^{-12}. Substituting $\beta = 2^j$ in (34), and comparing with (33), we arrive at the curious identity

$$-2\sum_{k=1}^{\infty} \ln(1+1/2^k) - \ln(2) = -\ln 2/12 - \pi^2/(6\ln 2) + \sum_{k=1}^{\infty} \frac{1}{k\sinh(2\pi^2 k/\ln(2))}.$$

ii) The moments of (32) are easily computed. By absolute convergence,

$$m \sim \beta Q^-(\infty) \sum_{j=0}^{\infty} C(j) \int_0^{\infty} e^{-x2^j}\, dx = \beta Q^-(\infty) \sum_{j=0}^{\infty} C(j)/2^j = 2\beta$$

by (43) . More generally,

$$\begin{aligned} m_k &:= \mathbb{E}(ND^k) \sim \beta^k Q^-(\infty) \sum_{j=0}^{\infty} C(j) \int_0^{\infty} e^{-x2^j} k x^{k-1} dx \\ &= \beta^k Q^-(\infty) k! \sum_{j=0}^{\infty} C(j)/2^{jk} = \beta^k k! Q^-(k) \end{aligned}$$

by (44) . However, this is exactly what we expected. Indeed, as $\beta \gg 1, m_k \sim k! S_k \sim k! \beta^k Q^-(k)$. Moreover this is of course compatible with (13).

- $\alpha = 1/2$. $K \sim e^{-1}/p_1 \sim 2m, \beta = K/4 \gg 1$. The fit of (31) with the observations($\alpha = 0.5$), with the observed mean in β is given in Figure 5. The fit of (32), with the observed mean in β is given in Figure 6. Our analysis given in the case $\alpha < 1/2$ is still valid.

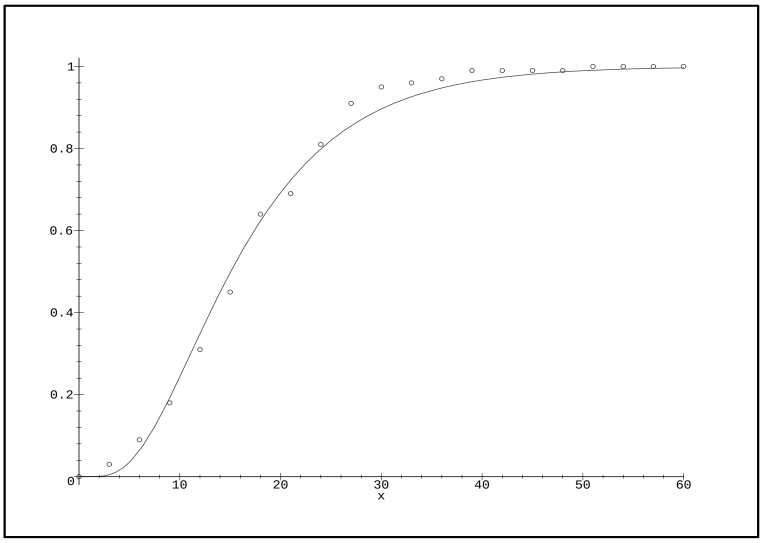

Figure 6: Observed and limiting (32) (with observed mean) ND distribution , $\alpha = .5$

- An interesting extension is to consider the case where we replace $H = n^{1/2}$ by $H = \sqrt{\delta n \log n}$. All our previous asymptotic analysis is still valid, with now

$$p_1 = \sqrt{\delta \log n/n}.$$

For instance, K becomes

$$K = \frac{e^{-\delta \log n}\sqrt{n}}{\sqrt{\delta \log n}} = \frac{n^{1/2-\delta}}{\sqrt{\delta \log n}}.$$

Hence,
if $\delta \geq 1/2$, $K \downarrow 0, n \to \infty$ and the almost sure non-decomposability analysis is available,
if $\delta < 1/2$, $K \uparrow \infty, n \to \infty$ and the almost sure decomposability analysis is valid.
This proves our Theorem 2.4.

- $\alpha > 1/2$. $K = e^{-np_1^2}/p_1 \sim 2m, \beta = K/4 > 1$. (A precise analysis of K is given in the appendix). The fit of (10) with the observations($\alpha = 0.55$), with the observed mean in β is given in Figure 7 . Apart from $k = 0, 2$, the fit is rather satisfactory. $\Pr[ND = 0]$ is asymptotically given by (10) or (11): this leads (with the observed mean in β) to 0.1344, the observed value of $\Pr[ND = 0] = 0.1180$. A more extended simulation is given in Figure 8, with $n = 10^6$ and $T = 12000$ trials. The fit near the origin is better.

- $\alpha > 1/2$. $K = e^{-np_1^2}/p_1 \sim 2m, \beta = K/4 < 1$. Now, we can use either (9) or (10). We can also explain the rather good fit of Figure 3 ($\alpha = 0.58$) with a Poisson distribution. Indeed, if $\beta \ll 1$, (11) leads to

$$\ln(\Pr[ND = 0]) \sim -\sum_{k=0}^{\infty} \ln(1 + \beta/2^k) \sim -\sum_{k=0}^{\infty} \beta/2^k \sim -2\beta \sim -m,$$

or

$$\Pr[ND = 0] \sim e^{-m}.$$

Similarly, in (46) , we use

$$s_1 = \sum_{k=0}^{\infty} \frac{1}{\beta + 2^k} \sim 2,$$

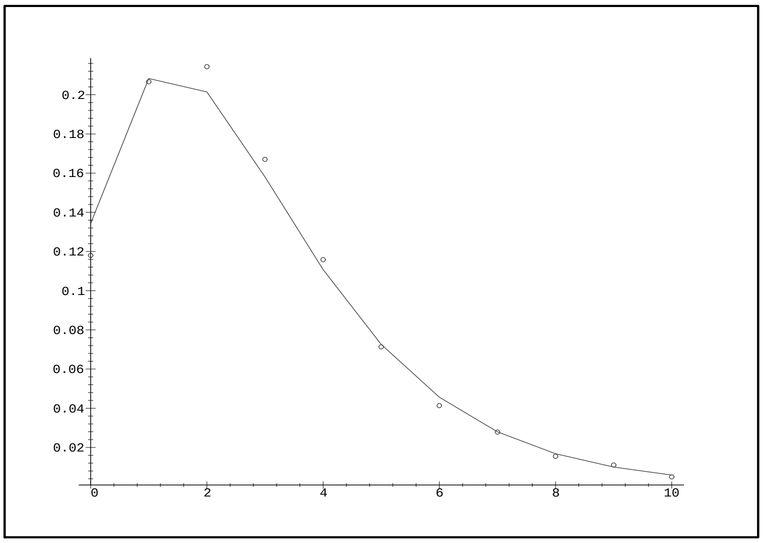

Figure 7: Observed and limiting (10) (with observed mean)ND distribution , $\alpha = .55$

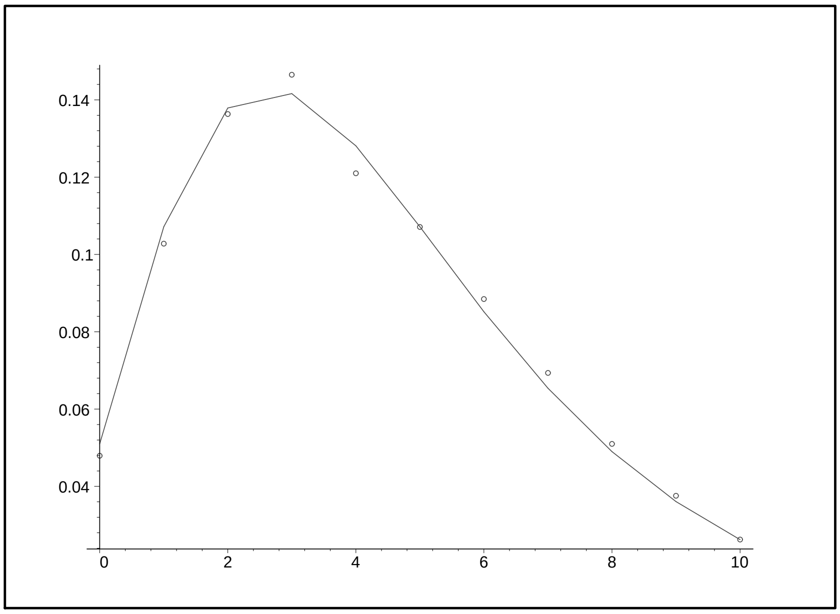

Figure 8: Observed and limiting (10) (with observed mean)ND distribution , $\alpha = .55, n = 10^6, T = 12000$

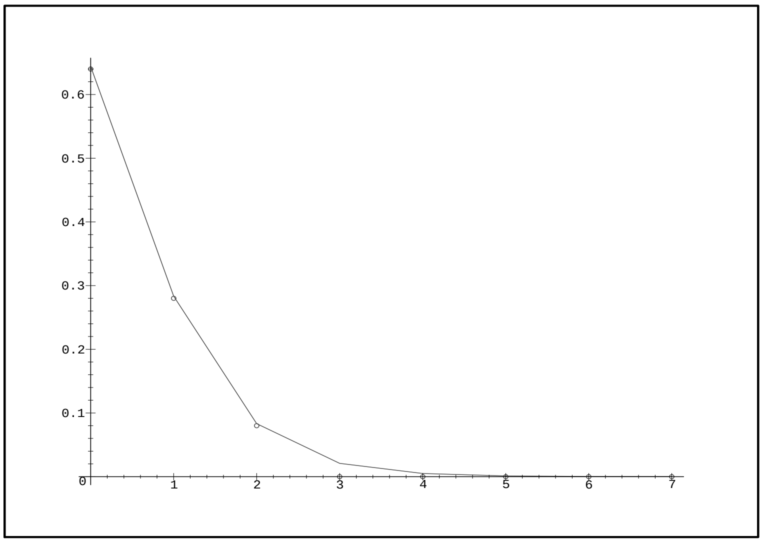

Figure 9: Observed and limiting (with (35) and observed mean) ND distribution , $\alpha = .58$

and

$$\Pr[ND = 1] \sim e^{-m} m.$$

More generally, we see that, by successively differentiating (40) wrt z, and setting $z = -\beta, q = 1/2$, (9) leads to

$$i! \Pr[ND = i]/\beta^i \sim e^{-m} i! [z^i] \prod_{k=0}^{\infty} \frac{1}{1-(1/2)^k z} = e^{-m} i! Q^-(i),$$

again by (40) , or

$$\Pr[ND = i] \sim e^{-m} (m/2)^i Q^-(i), \tag{35}$$

which corresponds to a form of **distorted** Poisson distribution. For instance, for $n = 10000, \alpha = .58$, the fit with (35) (with the observed mean) is given in Figure 9. The term $Q^-(i)$ in (35) can also be derived from (30).

Let us now compute the factorial moments of (35). We obtain

$$\tilde{m}_l := \mathbb{E}(ND^{\underline{l}}) = \sum_{i=0}^{\infty} e^{-m} i^{\underline{l}} \beta^i Q^-(i).$$

Again, successively differentiating (40) wrt z, setting now $z = \beta, q = 1/2$, we obtain, for small β,

$$\tilde{m}_l \sim e^{-m} \beta^l l! Q^-(l) e^m \sim l! \beta^l Q^-(l)$$

which is of course equivalent to (13).

8 Conclusion

The problem of the decomposability of a set of integers as the sum of two sets is a basic combinatorial problem related to the factorization of some classes of polynomials. In this paper we have addressed the problem of p-decomposability which arises when one of the two sets of the decomposition has a fixed size p. We have studied the limiting large-n probability of p-decomposability of a set drawn uniformly at random among the subsets of $[n]$ of size $n - n^\alpha$, showing that it changes abruptly for large n, decreasing from 1 to 0 as α increases through a critical value, $\alpha_c = (p-1)/p$. Then, focussing on 2-decomposability, we have provided a very detailed analysis of the asymptotic distribution of the number of possible decompositions as α varies. Particularly we have shown that there exists a sharp threshold function for 2-decomposability, namely $\sqrt{1/2\, n \ln n}$. A next step could be to adapt the study developed in this paper to 3-, 4-, ... decomposability, with the hope of eventually generalizing it to p-decomposability. However, the complexity of the analysis carried out for 2-decomposability casts doubt on the possibility of doing so. Probably, then, another approach is to be found, possibly inspired by this one.

Acknowledgements The first author wishes to thank Maurice Nivat for introducing him to the problem of decomposability. The pertinent comments of the referee led to improvements in the presentation.

References

[1] G.E. Andrews. *The Theory of Partitions*, volume 2 of *Encyclopedia of Mathematics and its Applications.* Addison–Wesley, 1976.

[2] A.D. Barbour, L. Holst, and S. Janson. *Poisson Approximation.* Oxford University Press, 1992.

[3] P. Billingsley. *Convergence of Probability Measures.* Wiley, 1968.

[4] B. Bollobás. *Random Graphs.* Academic Press, 1985.

[5] O. Dubois, R. Monasson, B. Selman, and R. Zecchina. Phase transition in combinatorial poblems. *Theoretical Computer Science*, 265(1-2):1–398, 2001.

[6] W. Feller. *Introduction to Probability Theory and its Applications. Vol II.* Wiley, 1971.

[7] P. Flajolet, X. Gourdon, and P. Dumas. Mellin transforms and asymptotics: Harmonic sums. *Theoretical Computer Science*, 144:3–58, 1995.

[8] P. Jacquet and W. Szpankowski. Analytic depoissonization and its applications. *Theoretical Computer Science*, 1-2:1–62, 1998.

[9] D.E. Knuth. *The Art of Computer Programming.V.3.* Addison-Wesley, 1998.

[10] V.K. Kolchin, A.S. Sevastyanov, and P.C. Chistiakov. *Random Allocations.* Wiley, 1978.

[11] G. Louchard. Exact and asymptotic distributions in digital and binary search trees. *RAIRO Theoretical Informatics and Applications*, 21(4):479–496, 1987.

[12] P.A. MacMahon. *Combinatorial analysis.* Cambridge University press (reprinted by Chelsea, New York, 1960), 1915.

[13] J. Riordan. *An Introduction to Combinatorial Analysis.* Wiley (Reprint: Princeton University Press, 1980), 1958.

[14] R. P. Stanley. *Enumerative Combinatorics Vol.II.* Cambridge University Press, 1999.

A Appendix: Proofs

A.1 Proof of Proposition 6.1

Proof We do have $W_0 = 1$ and $W_1 = 1/2$, so let i be ≥ 2. Our first (and main) step will be to prove by induction on i that for all such i and $x \geq 0$,

$$\Theta_i^*(x) = \sum_{k=-1}^{i-2} (-1)^{k+1} a_{i,k} e^{-\frac{2^{k+2}-1}{2^k}x} \tag{36}$$

where for $i \geq 2$ and $-1 \leq k \leq i-2$,

$$a_{i,k} = \frac{2^{(i-1)(i-2)/2+k(k+1)/2}}{\prod_{s=1}^{k+1}(2^s-1)\prod_{s=1}^{i-k-2}(2^s-1)}. \tag{37}$$

For $i = 2$,

$$\begin{aligned}
\Theta_2^*(x) &= \int_0^x \varphi^*(u)\,\psi^*(x|u)\,du \\
&= e^{-2x}\int_0^{x/2} e^{-2u}du + e^{-4x}\int_{x/2}^x e^{2u}du \\
&= e^{-2x} - e^{-3x}
\end{aligned}$$

does satisfy (36) with (37). Now suppose the same holds for some $i \geq 2$. We calculate Θ_{i+1}^* :

$$\begin{aligned}
\Theta_{i+1}^*(x) &= \int_0^x \Theta_i^*(u)\,\psi^*(x|u)\,du \\
&= \sum_{k=-1}^{i-2}(-1)^{k+1} a_{i,k}\left[e^{-2x}\int_0^{x/2} e^{-\frac{2^{k+2}-1}{2^k}u}du + e^{-4x}\int_{x/2}^x e^{\frac{u}{2^k}}du\right] \\
&= \sum_{k=-1}^{i-2}(-1)^{k+1} a_{i,k}\left\{\frac{2^k}{2^{k+2}-1}\left[e^{-2x} - e^{-\frac{2^{k+3}-1}{2^{k+1}}x}\right] + 2^k\left[e^{-\frac{2^{k+2}-1}{2^k}x} - e^{-\frac{2^{k+3}-1}{2^{k+1}}x}\right]\right\}.
\end{aligned}$$

Plainly then (agreeing, as we clearly may, that $a_{i,i-1} = 0$ for all i),

$$a_{i+1,k} = \begin{cases} 2^k a_{i,k} + \frac{2^{2k}}{2^{k+1}-1}a_{i,k-1}, & 0 \leq k \leq i-1; \\ a_{i,-1} + \sum_{l=0}^{i-2}(-1)^{l+1}\frac{2^l}{2^{l+2}-1}a_{i,l}, & k = -1. \end{cases} \tag{38}$$

(Note that the value for $k = -1$ *cannot* be written $\sum_{l=-1}^{i-2}(-1)^{l+1}2^l/\left(2^{l+2}-1\right)\,a_{i,l}$, since it actually exceeds this by $a_{i,-1}/2$.)
We deal with $k \geq 0$ first. For $k = i-1$, the quotient $2^{2k}/\left(2^{k+1}-1\right)$ equals $2^{(i-1)+k}/\left(2^{k+1}-1\right)$, which adds just the right factors to $a_{i,i-2}$ so that (38) extends (37). Suppose then $0 \leq k \leq i-2$, the induction hypothesis (37) implies

$$a_{i+1,k} = \frac{2^{(i-1)(i-2)/2}}{\prod_{s=1}^{k+1}(2^s-1)\prod_{s=1}^{(i+1)-k-2}(2^s-1)}\left[2^{k(k+1)/2+k}\left(2^{i-k-1}-1\right) + 2^{k(k-1)/2+2k}\right].$$

The quantity between brackets on the right equals $2^{k(k+1)/2+k}\left[\left(2^{i-k-1}-1\right)+1\right] = 2^{i-1}2^{k(k+1)/2}$, so that (38) again extends (37).
Let us turn now to $k = -1$. Using the q-factorial notation (see, e.g. [1]) :

$$(q)_s = (1-q)\left(1-q^2\right)\ldots(1-q^s),$$

the corresponding case of (38) reads

$$a_{i+1,-1} = (-1)^{i+1}\, 2^{(i-1)(i-2)/2} \left[\frac{1}{(2)_{i-1}} + \frac{1}{2} \sum_{l=0}^{i-2} (-1)^{l+1} \frac{2^{(l+1)(l+2)/2}}{2^{l+2}-1} \frac{1}{(2)_{l+1}\,(2)_{i-l-2}} \right].$$

Taking account of the fact that $\left(2^{l+2}-1\right)(2)_{l+1} = -(2)_{l+2}$ and applying the change of summation index $h = l+2$, we get

$$a_{i+1,-1} = (-1)^{i+1}\, 2^{(i-1)(i-2)/2} \left[\frac{1}{(2)_{i-1}} + \frac{1}{2} \sum_{h=2}^{i} (-1)^{h} \frac{2^{h(h-1)/2}}{(2)_h\,(2)_{i-h}} \right]. \tag{39}$$

Now from a special case of an identity due to Cauchy (Formula (3.3.6) of [1] with $z = 1$ and $q = 2$; see also [12], vol.II, p. 59), we know that

$$\sum_{h=0}^{i} (-1)^{h} \frac{2^{h(h-1)/2}}{(2)_h\,(2)_{i-h}} = 0.$$

Therefore, the quantity between brackets in (39) is

$$\frac{1}{(2)_{i-1}} - \frac{1}{2}\left(\frac{1}{(2)_i} - \frac{1}{(2)_1\,(2)_{i-1}}\right) = \frac{1}{2}\left(\frac{1}{(2)_{i-1}} - \frac{1}{(2)_i}\right) = \frac{1}{2\,(2)_i}\left[(1-2^i)-1\right] = -\frac{2^{i-1}}{(2)_i},$$

so that finally

$$a_{i+1,-1} = (-1)^i \frac{2^{i(i-1)/2}}{(2)_i},$$

which is indeed the same as (37) with i replaced by $i+1$. Thus (36) with (37) is established for all $i \geq 2$.

Now for the short second step. Integrating (36) yields, in view of (38) and (37) :

$$\begin{aligned} W_i &= \sum_{k=-1}^{i-2} (-1)^{k+1} \frac{2^k}{2^{k+2}-1} a_{i,k} = a_{i+1,-1} - \frac{a_{i,-1}}{2} = (-1)^i \left[\frac{2^{i(i-1)/2}}{(2)_i} + \frac{2^{(i-1)(i-2)/2-1}\left(1-2^i\right)}{(2)_i} \right] \\ &= (-1)^i \frac{2^{i(i-3)/2}}{(2)_i}, \end{aligned}$$

which is just the last expression in (27), and trivially equal to the other two. ■

A.2 Proof of Proposition 4.2

Proof Equ.(9) can be written, with $\beta := K(n,\alpha)/4$,

$$\Pr[ND = k] \sim Q^-(\infty)(-1)^k \sum_{i=k}^{\infty} (-1)^i \binom{i}{k} \beta^i + (-1)^k \sum_{i=k}^{\infty} (-1)^i \binom{i}{k} \beta^i [Q^-(i) - Q^-(\infty)].$$

However,

$$\begin{aligned} Q^-(i) - Q^-(\infty) &= \prod_{k=1}^{i} \frac{1}{1-2^{-k}} - \prod_{k=1}^{\infty} \frac{1}{1-2^{-k}} \\ &= Q^-(\infty) \left[\prod_{k=i+1}^{\infty} (1-2^{-k}) - 1 \right]. \end{aligned}$$

Using two classical Euler identities, we will derive several summation formulae we need in the sequel. These identities are

$$\prod_{k=0}^{\infty} \frac{1}{1-q^k z} = 1 + \sum_{i=1}^{\infty} z^i / \prod_{k=1}^{i} (1-q^k), \tag{40}$$

$$\prod_{k=0}^{\infty} (1+q^k z) = 1 + \sum_{i=1}^{\infty} z^i q^{i(i-1)/2} / \prod_{k=1}^{i} (1-q^k). \tag{41}$$

(for a simple proof, see Knuth [9],Ex 5.1.1/16). First, set $z = -1/2^{i+1}, q = 1/2$ in (41). This leads to

$$\begin{aligned} \prod_{k=i+1}^{\infty} (1-2^{-k}) &= 1 + \sum_{j=1}^{\infty} (-1)^j \frac{1}{2^{(i+1)j}} \frac{1}{2^{j(j-1)/2}} Q^-(j) \\ &= \sum_{j=0}^{\infty} \frac{(-1)^j}{2^{ij} 2^{j(j+1)/2}} Q^-(j) = \sum_{j=0}^{\infty} \frac{C(j)}{2^{ij}}, \end{aligned} \tag{42}$$

where

$$C(j) := \frac{(-1)^j}{2^{j(j+1)/2}} Q^-(j).$$

For further use, we note that, setting $i = 0$ in (42), we obtain $Q(\infty) = \sum_{j=0}^{\infty} C(j)$. Setting $i = 1$, we derive

$$\frac{Q(\infty)}{1-1/2} = \sum_{j=0}^{\infty} \frac{C(j)}{2^j}. \tag{43}$$

(This is relation (A.4) in [11]). More generally,

$$Q(\infty) Q^-(i) = \sum_{j=0}^{\infty} \frac{C(j)}{2^{ij}}. \tag{44}$$

Putting everything together we obtain

$$\Pr[ND = k] \sim Q^-(\infty)(-1)^k \sum_{j=0}^{\infty} C(j) \sum_{i=k}^{\infty} (-1)^i \binom{i}{k} \frac{\beta^i}{2^{ij}}. \tag{45}$$

However, differentiating the GF $1/(1+\gamma z) = \sum (-1)^i \gamma^i z^i$, we obtain

$$\frac{\gamma^k}{(1+\gamma)^{k+1}} = (-1)^k \sum_{i=k}^{\infty} (-1)^i \binom{i}{k} \gamma^i,$$

hence (45) becomes

$$\Pr[ND = k] \sim Q^-(\infty) \sum_{j=0}^{\infty} C(j) \frac{(\beta/2^j)^k}{(1+\beta/2^j)^{k+1}}$$

■

A.3 Proof of Proposition 4.3

First, we set $z = -\beta, q = 1/2$ in (40): Equ.(9) leads to

$$\Pr[ND = 0] \sim \prod_{k=0}^{\infty} \frac{1}{1+\beta/2^k}$$

Now, differentiating (40) wrt z we obtain

$$\sum_{k=0}^{\infty}\frac{q^k}{1-q^kz}\prod_{u=0}^{\infty}\frac{1}{1-q^uz}=\sum_{i=1}^{\infty}iz^{i-1}/\prod_{k=1}^{i}(1-q^k).$$

Setting $z=-\beta, q=1/2$, we obtain

$$\sum_{k=0}^{\infty}\frac{(1/2)^k}{1+\beta/2^k}\prod_{u=0}^{\infty}\frac{1}{1+\beta/2^u}=\sum_{i=1}^{\infty}i(-1)^{i-1}\beta^{i-1}Q^-(i)),$$

i.e., by (9)

$$\Pr[ND=1]/\beta\sim\sum_{k=0}^{\infty}\frac{1}{\beta+2^k}\Pr[ND=0],$$

or

$$\Pr[ND=1]\sim\beta\sum_{k=0}^{\infty}\frac{1}{\beta+2^k}\Pr[ND=0]. \tag{46}$$

A direct check of (11) and (46) is given as follows: (10) leads to

$$\Pr[ND=0]\sim Q^-(\infty)\sum_{j=0}^{\infty}\frac{C(j)}{1+\beta/2^j}=Q^-(\infty)\sum_{j=0}^{\infty}C(j)\sum_{i=0}^{\infty}\frac{(-1)^i\beta^i}{2^{ij}}. \tag{47}$$

However, (40) gives

$$\prod_{k=0}^{\infty}\frac{1}{1+\beta/2^k}=\sum_{i=1}^{\infty}(-1)^i\beta^iQ^-(i),$$

and by (44), we recover immediately (11). Similarly, differentiating (11) and (47) wrt β, we derive

$$Q^-(\infty)\sum_{j=0}^{\infty}C(j)\frac{1/2^j}{(1+\beta/2^j)^2}=\prod_{k=0}^{\infty}\frac{1}{1+\beta/2^k}\sum_{k=0}^{\infty}\frac{1}{\beta+2^k}.$$

which immediately leads to (46).

Let us now proceed for all k. We first prove the following lemma:

Lemma A.1 *Setting*

$$s_i=\sum_{j=0}^{\infty}\frac{1}{(\beta+2^j)^i},$$

the following obtains, for any $k\geq 1$ (and $\beta<1$ at least):

$$s_k=k\sum_{\alpha_1+2\alpha_2+...+k\alpha_k=k}\frac{(-1)^{\alpha_1+...+\alpha_k+1}}{\alpha_1+...+\alpha_k}\binom{\alpha_1+...+\alpha_k}{\alpha_1\ \ ...\ \ \alpha_k}\prod_{i=1}^{k}\left(\frac{P_i}{\beta^iP_0}\right)^{\alpha_i}$$

with $P_i\sim\Pr[ND=i]$.

Proof Let A be the right-hand side of (40). By $k-$fold differentiation of the logs of both sides, using the Faà di Bruno formula for the kth derivative of a compound function (e.g., [13], p. 35):

$$\begin{aligned}
&(k-1)!\sum_{j=0}^{\infty}\frac{q^{jk}}{(1-q^jz)^k}\\
&=\sum_{\alpha_1+2\alpha_2+...+k\alpha_k=k}\left(D^{\alpha_1+...+\alpha_k}\log\right)(A)\frac{k!}{\alpha_1!...\alpha_k!1!^{\alpha_1}...k!^{\alpha_k}}\times(DA)^{\alpha_1}\left(D^2A\right)^{\alpha_2}...\left(D^kA\right)^{\alpha_k}\\
&=k!\sum_{\alpha_1+2\alpha_2+...+k\alpha_k=k}(-1)^{\alpha_1+...+\alpha_k+1}\frac{(\alpha_1+...+\alpha_k-1)!}{\alpha_1!...\alpha_k!A^{\alpha_1+...+\alpha_k}}\times\left(\frac{DA}{1!}\right)^{\alpha_1}\left(\frac{D^2A}{2!}\right)^{\alpha_2}...\left(\frac{D^kA}{k!}\right)^{\alpha_k}.
\end{aligned}$$

Now from (40) for $z=-\beta$ and $q=1/2$, and (9), we see that $D^iA=i!P_i/\beta^i$, where $P_i\sim\Pr[ND=i]$. The result follows. ∎

Although we are only interested here in the field of reals, the base field in the following two lemmas could of course be $\mathbb{C}$.

Lemma A.2 *For $1\le i\le k$, let the polynomial P_i in the indeterminates $X_1,...,X_i$ be of the form $c_iX_i+S_i(X_1,...,X_{i-1})$, with $c_i\neq 0$. [S_1 is understood to be a constant.]*
(i) There is (at least one) polynomial $Q(X_1,...,X_k)$, depending only on the P_i's, such that for any polynomials $A_i(X_1,...,X_k), B_i(X_1,...,X_k)$ verifying $A_i=P_i(B_1,...,B_i)$ $(1\le i\le k)$, we have $B_k=Q(A_1,...,A_k)$. [In particular, we have of course $X_k=Q(P_1,...,P_k)$.]
(ii) If there are particular *polynomials $U_i(X_1,...,X_k), V_i(X_1,...,X_k)$ $(1\le i\le k)$, with the U_i's* algebraically independent, *and $R(X_1,...,X_k)$, such that $U_i=P_i(V_1,...,V_i)$ and $V_k=R(U_1,...,U_k)$, then the polynomial Q in (i) is unique and equal to R.*
(iii) If (ii) holds, then for any real numbers $a_1,...,a_k$ and $b_1,...,b_k$, the equalities $a_i=P_i(b_1,...,b_i)$ $(1\le i\le k)$ imply $b_k=R(a_1,...,a_k)$.

Proof *(i)* We have $B_1=c_1^{-1}(A_1-S_1)$. Call this $Q_1(A_1)$. For any $r<k$, from $B_i=Q_i(A_1,...,A_i)$ $(1\le i\le r)$ there follows $B_{r+1}=c_{r+1}^{-1}(A_{r+1}-S_{r+1}(Q_1,...,Q_r))$ which we call $Q_{r+1}(A_1,...,A_{r+1})$. We then may take $Q=Q_k$.
(ii) For any Q as in *(i)* we have $V_k=Q(U_1,...,U_k)=R(U_1,...,U_k)$ Since the U_i's are algebraically independent, $Q-R$ is the null polynomial.
(iii) The a_i's and b_i's are just particular polynomials. [Note that we do *not* assert that R is necessarily the *only* polynomial such that for *arbitrary numbers* a_i,b_i the equalities $a_i=P_i(b_1,...,b_i)$ imply $b_k=R(a_1,...,a_k)$. This is certainly true in the event that the P_i's are algebraically independent, but in any case we do not need it.] ∎

Lemma A.3 *If real numbers a_i and b_i $(1\le i\le k)$ verify*

$$a_r=(-1)^{r+1}\sum_{\alpha_1+2\alpha_2+...+r\alpha_r=r}\frac{1}{\alpha_1+...+\alpha_r}\binom{\alpha_1+...+\alpha_r}{\alpha_1\ ...\ \alpha_r}b_1^{\alpha_1}...b_r^{\alpha_r}\ \textit{for}\ 1\le r\le k,$$

then we have

$$b_r=(-1)^{r+1}\sum_{\alpha_1+2\alpha_2+...+r\alpha_r=r}\frac{a_1^{\alpha_1}...a_r^{\alpha_r}}{\alpha_1!...\alpha_r!}\ \textit{for}\ 1\le r\le k. \tag{48}$$

Proof We apply Lemma A.2 with

$$P_r(X_1,...,X_r)=(-1)^{r+1}\sum_{\alpha_1+2\alpha_2+...+r\alpha_r=r}\frac{1}{\alpha_1+...+\alpha_r}\binom{\alpha_1+...+\alpha_r}{\alpha_1\ ...\ \alpha_r}X_1^{\alpha_1}...X_r^{\alpha_r}\ \text{for}\ 1\le r\le k$$

and

$$R(X_1,...,X_k)=(-1)^{k+1}\sum_{\alpha_1+2\alpha_2+...+k\alpha_k=k}\frac{X_1^{\alpha_1}...X_k^{\alpha_k}}{\alpha_1!...\alpha_k!}.$$

For $1\le r\le k$, let $\sigma_r=\sum_{1\le i_1<i_2<...<i_r\le k}X_{i_1}...X_{i_r}$ be the elementary symmetric polynomials in the indeterminates $X_1,...,X_k$, and $S_r=\sum_{i=1}^k X_i^r$ the Newton sums. From the classical Waring identities (e.g., [12],Vol.1,p.5,[14],esp.(7.23)), the polynomials P_i, R, $U_i=-S_i/i$, and $V_i=-\sigma_i$ are immediately seen to verify *(ii)*. The S_i's, thus also the U_i's, are known to be algebraically independent (loc.cit.). Hence (48) for $r=k$ is obtained from *(iii)*. Since k was arbitrary, we are done. ∎

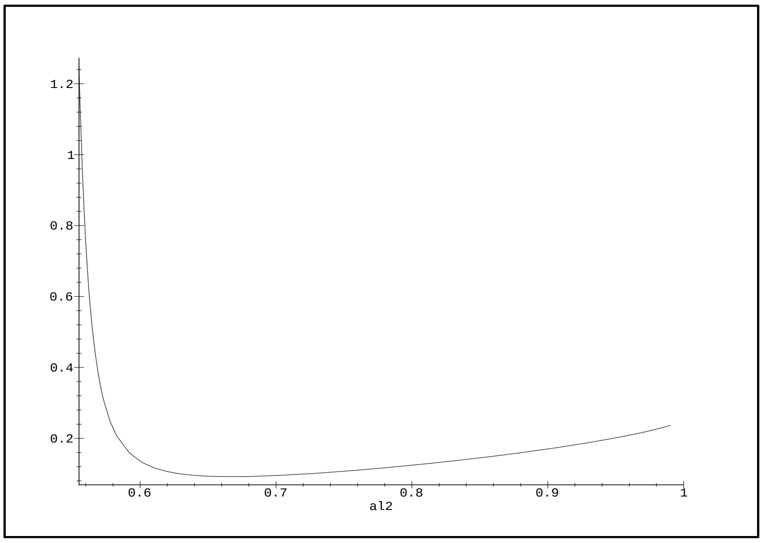

Figure 10: $S(\alpha)$

The proof of Proposition 4.3 goes as follows.
Proof Using Lemma A.1, we see that the a_i's and b_i's given, for $1 \le i \le k$, by

$$a_i = (-1)^i \, s_i/i \quad \text{and} \quad b_i = -\frac{P_i}{\beta^i P_0}$$

comply with the hypotheses of Lemma A.3. Writing down (48) we get (12). ■

A.4 A precise analysis of $\beta = K(n, \alpha)/4, \alpha > 1/2$

In this section, we show that there exists still another threshold value α^* delimiting the choice of asymptotic expressions. We consider $\beta = K(n, \alpha)/4 = e^{-n^{2\alpha-1}}/(4n^{\alpha-1})$. Differentiating wrt n, (considered as a real), we see that β possesses a unique maximum for

$$n = \left(\frac{1-\alpha}{2\alpha-1}\right)^{1/(2\alpha-1)}.$$

This maximum, $S(\alpha)$, say, is given in Figure 10 . $S(\alpha) \uparrow \infty$ as $\alpha \downarrow 1/2$. The unique threshold value giving $S(\alpha^*) = 1$ is $\alpha^* = 0.557023356433\ldots$. For $\alpha > \alpha^*$, we can use either (25) or (10). For $1/2 < \alpha < \alpha^*$, we observe two values $n_1(\alpha), n_2(\alpha)$ such that, for $n \in [1/2, n_1]$ or $n \in [n_2, \infty]$, we have $\beta < 1$ and we can use either (25) or (10). For $n \in [n_1, n_2]$, $\beta > 1$, we must use (10). The two functions n_1, n_2 can be computed as follows. Set

$$W := -(2\alpha - 1)[\ln n - 2\ln 2/(1-\alpha)].$$

This leads to

$$We^W = x, \quad \text{with} \quad x = \frac{2\alpha-1}{\alpha-1} e^{2\ln 2(2\alpha-1)/(1-\alpha)}.$$

This is the well-known equation of the Lambert function $W(x)$. W is related to the tree GF $T(x)$ by the relation $T(x) = -W(-x)$. It appears that two branches of W lead to the solutions we need here: the classical one $W(x)$, analytic at 0, and the other one, $W(-1, x)$. We obtain

$$\begin{aligned} n_1(\alpha) &= \exp\{-[W(x) - 2\ln 2(2\alpha-1)/(1-\alpha)]/(2\alpha-1)\}, \\ n_2(\alpha) &= \exp\{-[W(-1,x) - 2\ln 2(2\alpha-1)/(1-\alpha)]/(2\alpha-1)\}. \end{aligned}$$

For instance, with $\alpha = 0.55$, we obtain $n_1 = 3140, n_2 = 988126791$.