Journal of Mathematical Chemistry 24 (1998) 187–221

Statistics of aggregates

Thierry Huillet

LIMHP-CNRS, Université Paris XIII, Institut Galilée, 93430 Villetaneuse, France
E-mail: huillet@limhp.univ-paris13.fr

Received 28 May 1998

Aggregation phenomena of elementary particles into clusters has received considerable attention during the past few decades. We adopt here a stochastic approach for the modeling of these phenomena. More precisely, we formulate the problem in the following dynamical setup: given a population of n discernible atoms partitioned into p discernible (model 1) or indiscernible (model 2) groups, how does a new atom *eventually* connect to any of these p groups forming up a new partition of $n+1$ atoms into a certain amount of clusters? Nucleation is said to occur when the inserted atom does not connect (it nucleates), whereas aggregation takes place if it does (clusters coalesce). Depending on this local "logic" of pattern formation, the asymptotic structure of groups can be quite different, in the thermodynamic limit $n \to \infty$. These studies are the main purpose of this work. Understanding these aggregation phenomena requires first to derive the fragment size distributions (that is, the number P of fragments, or clusters, and the number N_m of size-m fragments with m constitutive atoms), as a function of the *control* parameter which is chosen here to be the average number of atoms $\langle N \rangle$. As $\langle N \rangle$ approaches infinity, we derive the study of these variables in the thermodynamic limit $n \to \infty$. It is shown, making extensive use of combinatorics, that two regimes are to be distinguished: the one of *weakly connected aggregates* where nucleation dominates aggregation and the one of *strongly connected aggregates* where aggregation dominates nucleation. In the first ("diluted") regime, the number of clusters $P(n)$ always diverges as $n \to \infty$, the asymptotic equivalent of which being under control in most cases. Large deviation results are shown to be available. Concerning $N_m(n)$, distinct behaviours are observed in models 1 and 2. In the second ("condensed") regime, the number of groups $P(n)$ and size-m groups $N_m(n)$ may converge in the thermodynamic limit, with a special role played by the geometric and Poisson distributions. The asymptotic variables become observable macroscopically. This work is therefore aimed toward a better understanding of the fundamentals involved in clusters' formation processes.

1. Introduction

Aggregation phenomena manifest themselves on a wide variety of physical scales, from the large structures of the Universe to the elementary particles and are still poorly understood. The purpose of the recent work [14] was to introduce a natural way of modeling nucleation–aggregation phenomena of elementary particles into clusters at the statistical physics' level, regardless of the nature of their constitutive elements or of the properties of their binding force. This field has recently received considerable interest,

in an attempt to understand processes such as coagulation of fine particles or coalescence of droplets: although there are many alternative approaches to this question, the most notable being mean-field population-balance Schmoluchowski's equations (see, e.g., [3,10,22]), we adopt here a microscopic approach for the representation of these phenomena, which avoids the need for certain unknown parameters such as coagulation and fragmentation rates ....

More precisely, we formulated the problem as follows: nucleation–aggregation phenomena consists of problems where "elementary" particles (atoms) are given the opportunity of forming "assemblies" (groups or clusters). Inspecting more closely this problem amounts to asking for the "connection politics" of a new atom when it "sees" a certain previously-formed group-pattern of n similar atoms. In this approach, nucleation of a cluster occurs when the inserted atom does not connect at all, whereas aggregation takes place when it joins any existing cluster, or more. Depending on this local "logic" of pattern formation, the asymptotic structure of groups (in the thermodynamic limit $n \to \infty$) can be quite different; also the group sizes' distributions may vary widely. These asymptotic studies are the main purpose of this work. We first (section 2) illustrate our ideas on two "monomer addition" models which in fact are two different and basic "connection logic". In these simplistic models, the additional atom, if it connects, connects to a single group: clusters grow by the addition of single particles. In section 3 we shall identify the model class to work with when clusters themselves are allowed to aggregate, which should match with more realistic situations. We shall make an extensive use of the notion of generating (partition) functions from combinatorics and more precisely of the one of "forests" (ordered or not) of increasing "trees" [9,12,16,19,21].

Understanding these aggregation phenomena requires first (in section 4) to derive the fragment size distributions (that is, the number P of fragments, or clusters, and the number N_m of size-m fragments with m constitutive atoms), as a function of the *control* parameter which is chosen here to be the average number of atoms $\langle N \rangle$.

As $\langle N \rangle$ approaches infinity, aggregates become unobservable and we derive the study of these variables in the thermodynamic limit $n \to \infty$, making extensive use of singularity analysis techniques [11].

This approach allows one to introduce (in section 5) the notions of *weakly connected aggregates* where nucleation dominates aggregation and the one of *strongly connected aggregates* where the contrary holds true. As conventional wisdom suggests, it happens that a strong statistical variability holds in the different models introduced; we shall indicate in which precise sense.

Roughly speaking, in the first ("diluted") regime, the number of clusters $P(n)$ always diverges as $n \to \infty$, the asymptotic equivalent of which being under control in most cases. The speed of divergence fixes the asymptotic behavior of the auxiliary number of size-m groups variable $N_m(n)$. Large deviation results are shown to be available. In the second ("condensed") regime, the number of groups $P(n)$ and size-m groups $N_m(n)$ may converge in the thermodynamic limit, with a special role played by

the geometric and Poisson distributions. The asymptotic structure becomes observable macroscopically.

2. Simple monomer-addition models

Let us start with two basic models: assume a population of n atoms is partitioned into p distinct non-empty groups. There can be at least one group of size n and at most n groups of size one. Let thus $n_j,\ j = 1, \ldots, n$, denote the number of clusters of size j in such a partition. Of course, if all this is to be consistent,

$$\sum_{j=1}^{n} j n_j = n \tag{1}$$

and

$$\sum_{j=1}^{n} n_j = p, \tag{2}$$

expressing (respectively) the number of atoms' (groups') conservation.

We shall next concentrate on the number of configurations with n (labeled) atoms and p (unlabeled) clusters.

Let $\sigma_n(p),\ p = 1, \ldots, n$, denote this quantity.

In order to clarify what these numbers really are, we now discuss the fate of an additional atom being added up to this structure (in the transition $n \to n + 1$):

- This new atom may not "connect" to any of the p existing groups, thereby forming itself a new group: it nucleates a new cluster.
- This new atom connects preferentially to (no more than) one of the p existing groups; the question is what group?

In order to answer the above question, let us make more precise the state-space of our models' class.

Given a population of n atoms, let P_n denote the number of groups in the partition, and let $(S_1(n), \ldots, S_{P_n}(n))$ be the group sizes' vector. Of course,

$$\sum_{p=1}^{P_n} S_p(n) = n. \tag{3}$$

Next, in the transition $n \to n + 1$, consider the (random) event $C_n = p$ that connection C_n of the additional atom is established with group $p \in \{1, \ldots, P_n\}$: the indicator function $\mathbf{1}_{C_n = p}$ will thus be one if its argument is true, zero otherwise. Lack of connection will be represented by the event $C_n = 0$, connection by the event $C_n > 0$.

We then have the obvious following state-space evolution equations:

$$\begin{bmatrix} S_1(n+1) \\ \vdots \\ S_p(n+1) \\ \vdots \\ \vdots \\ S_{P_{n+1}}(n+1) \\ 0 \\ \vdots \end{bmatrix} = \sum_{p=1}^{P_n} \begin{bmatrix} S_1(n) \\ \vdots \\ S_p(n)+1 \\ \vdots \\ S_{P_n}(n) \\ 0 \\ 0 \\ \vdots \end{bmatrix} \mathbf{1}_{C_n=p} + \begin{bmatrix} S_1(n) \\ \vdots \\ S_p(n) \\ \vdots \\ S_{P_n}(n) \\ 1 \\ 0 \\ \vdots \end{bmatrix} \mathbf{1}_{C_n=0}, \tag{4}$$

$$P_{n+1} = P_n \cdot \mathbf{1}_{C_n>0} + (P_n+1) \cdot \mathbf{1}_{C_n=0}. \tag{5}$$

Equation (4) indicates how to increment the group sizes' vector whenever a connection is established, whereas equation (5) is concerned with the (non-decreasing) number of such groups.

Adopting thus for a while a probabilistic language, we consider the three following random self-consistent connection rules of the additional atom giving the (random) probability, say $\mathbf{Q}$, that connection is established with group $p \in [1, \ldots, P_n]$:

(a) $\mathbf{Q}(C_n = 0) = 1/(P_n+1),\ \mathbf{Q}(C_n = p) = 1/(P_n+1),\ p = 1, \ldots, P_n,$

(b) $\mathbf{Q}(C_n = 0) = 1/(n+1),\ \mathbf{Q}(C_n = p) = S_p(n)/(n+1),\ p = 1, \ldots, P_n,$

for which, respectively,

(a) $\mathbf{Q}(C_n = 0) = 1/(P_n+1),\ \mathbf{Q}(C_n > 0) = P_n/(P_n+1),$

(b) $\mathbf{Q}(C_n = 0) = 1/(n+1),\ \mathbf{Q}(C_n > 0) = n/(n+1).$

Letting thus $\mathbf{Q}(P_n = p) \stackrel{\text{def}}{=} \mathbf{Q}_n(p)$ denote the probability that the (random) number of groups is p given a population of n atoms, it follows from the state-space equation (5) that

(a) $\mathbf{Q}_{n+1}(p) = 1/p \cdot \mathbf{Q}_n(p-1) + p/(p+1) \cdot \mathbf{Q}_n(p),$

(b) $\mathbf{Q}_{n+1}(p) = 1/(n+1) \cdot \mathbf{Q}_n(p-1) + n/(n+1) \cdot \mathbf{Q}_n(p).$

The answer to our question, therefore, lies in the three recurrences on $\sigma_n(p)$ we shall consider here (omitting normalization constants in the *transition probabilities*), skipping from the probabilistic language to the one of enumeration

$$\begin{aligned} &\text{(a)} \quad \sigma_{n+1}(p) = \sigma_n(p-1) + p\sigma_n(p), \\ &\text{(b)} \quad \sigma_{n+1}(p) = \sigma_n(p-1) + n\sigma_n(p), \end{aligned} \tag{6}$$

with common boundary conditions

$$\begin{aligned} &\sigma_1(1)=1,\\ &\sigma_n(0)=0, \quad \forall n \geqslant 1,\\ &\sigma_n(p)=0, \quad \forall p \geqslant n+1,\ \forall n \geqslant 1. \end{aligned} \tag{7}$$

Thus:

In model (a), all existing p groups are equally likely to form a new group (or not) with the additional atom, independently of the sizes of these p groups.

In model (b), the connection is more likely to occur with a group of large size: the additional atom behaves gregariously and preferentially moves towards larger groups.

Recurrences (7) identify the numbers' sequences under concern, namely,

(a) $\sigma_n(p) = S_n(p)$, the second kind Stirling's numbers,

(b) $\sigma_n(p) = |s_n(p)|$, the absolute values of the first kind Stirling's numbers.

Let us also introduce the numbers $\sigma_{_n} = \sum_{p=1}^{n} \sigma_n(p)$, giving the total number of partitions of n atoms, which are, respectively,

(a) $\sigma_{_n} = B_n$, the Bell numbers,

(b) $\sigma_{_n} = n!$.

Observe also that

$$\sigma_{_n}(p) \stackrel{\text{def}}{=} \#\big\{q \in \{1,\ldots,\sigma_n\}\colon\ P_{_n}(q) = p\big\}, \quad p = 1,\ldots,n, \tag{8}$$

where $\#\{\ldots\}$ is to be read as "the cardinal of the set ...".

3. Cluster–cluster aggregation: the generic model

We now come to the question of including clusters' aggregation into our models: first note that the hypothesis of section 2 that connection, if established, concerns a single group is very restrictive, although, as one can guess, far from elementary. In this sense, these aggregation models are what one may call monomer-addition since clusters grow by the addition of single particles only – there is no aggregation of two (or more) larger clusters together. We shall now indicate how to include aggregation of clusters which appears more realistic in practise.

We first introduce a combinatorial tree-structure which shall prove useful for our purpose. Suppose j atoms (or nodes) have been labeled $\{1,\ldots,j\}$. A labeled increasing tree is a rooted simply connected tree for which labels along any branch from the root are forced to go in increasing order. The enumeration of such trees has been undertaken in [2]. Suppose there are $(c_j)_{j\geqslant 1}$ such trees with j atoms. Introduce then this sequence's (exponential) generating function, say $\overline{t}(\theta)$, by

$$\overline{t}(\theta) \stackrel{\text{def}}{=} \sum_{j\geqslant 1} \frac{c_j}{j!}\theta^j.$$

The main result of these authors is that $\overline{t}(\theta)$ is the unique solution to the autonomous ordinary differential equation

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\overline{t}(\theta) \stackrel{\text{def}}{=} \overline{t}(\theta)' = g\big(\overline{t}(\theta)\big) \tag{9}$$

with initial condition $\overline{t}(0) = 0$, and for various g such that

$$g(\theta) = 1 + \sum_{k\geqslant 1} \frac{g_k}{k!}\theta^k,$$

where $(g_k)_{k\geqslant 1}$is any sequence of integral numbers, $g_k \leqslant k!$.

Function g is called the branch-generating function. It describes locally the variability of the edges pointing outwards any node of the tree.

The general unordered increasing tree (or non-plane tree) is, thus, recursively defined by appending an atom to a set of similar sub-trees. This amounts to choose $g(\theta) = \mathrm{e}^{\theta}$, for which $g_k = 1,\ k \geqslant 1$. Note that for such trees, there is no order distinction between the sub-trees dangling from their common root. Ordered trees, or plane trees, could be obtained in a similar way, when considering the modified functional equation (9) with $g(\theta) = 1/(1-\theta)$, for which $g_k = k!,\ k \geqslant 1$: for such trees, there are $k!$ ways to arrange k sub-trees, taking "chirality" into account. Therefore, modifying the "branch"-generating function g gives rise to a variety of tree structures implicitly defined by (9). For example, unordered binary trees are enumerated while using $g(\theta) = 1 + \theta^2/2$ in the above functional equation, whereas ordered binary trees can be obtained from $g(\theta) = 1 + \theta^2$. Linear increasing trees are enumerated while using $g(\theta) = 1 + \theta$ and constitute the simplest such structures.

If an explicit solution for the ordinary differential equation (9) exists, we shall call the model *solvable*. For example, choosing for g the following particular functions:

$$1+\theta, \quad (1+\theta)^{p+1} \text{ (with integer } p \geqslant 1), \quad 1+\theta^2, \quad 1+\theta^2/2!, \quad 1/(1-\theta), \quad \mathrm{e}^{\theta},$$

yields, respectively, for $\overline{t}(\theta)$,

$$\exp\theta - 1, \quad -1+[1-p\theta]^{-1/p}, \quad \tan\theta, \quad \sqrt{2}\tan(\theta/\sqrt{2}), \quad 1-\sqrt{1-2\theta}, \quad -\log(1-\theta).$$

Considering a forest of increasing trees leads to the generating function $\mathrm{e}^{\overline{t}(\theta)}$, the Taylor coefficient of which, say σ_n, counts the number of forests of increasing trees that one can form with n atoms, relaxing the connectedness condition for trees [9,12].

3.1. Discernible atoms and indiscernible clusters (model 1)

3.1.1. Partition function for the number of clusters distribution

Next consider the bivariate "marked" exponential generating function

$$Z^1(\theta,\gamma) \stackrel{\text{def}}{=} \mathrm{e}^{\gamma\overline{t}(\theta)}. \tag{10}$$

We shall also let

$$Z^1(\theta) \stackrel{\text{def}}{=} Z^1(\theta,1). \tag{11}$$

Developing (10),

$$Z^1(\theta,\gamma) \stackrel{\text{def}}{=} 1+\sum_{n\geqslant 1}\frac{\theta^n}{n!}Z_n^1(\gamma)$$

with

$$Z_n^1(\gamma) \stackrel{\text{def}}{=} \sum_{p=1}^{n}\sigma_n(p)\gamma^p.$$

In this interpretation, $\sigma_n(p)$ counts the number of forests made of p increasing trees that one can form with n atoms, and the Taylor coefficient of $Z^1(\theta)$, $\sigma_n = \sum_{p=1}^{n}\sigma_n(p)$, counts the total number of available configurations.

We now come to our cluster aggregation models.

Differentiating (10) with respect to θ gives, from (9),

$$\partial_\theta Z^1(\theta,\gamma) = \gamma\overline{t}(\theta)' Z^1(\theta,\gamma) = \gamma\sum_{k\geqslant 0}\frac{g_k}{k!}\big(\overline{t}(\theta)^k Z^1(\theta,\gamma)\big).$$

This leads to the following recurrences for the functions' sequence $(Z_n^1(\gamma))_{n\geqslant 1}$:

$$Z_{n+1}^1(\gamma) = \gamma\sum_{k\geqslant 0}\frac{g_k}{k!}\partial_\gamma^{(k)} Z_n^1(\gamma),$$

where $\partial_\gamma^{(k)}$ indicates derivation with respect to γ (k times).

In terms of the coefficients $\sigma_n(p)$ describing $Z_n^1(\gamma)$, this yields the recurrences

$$\sigma_{n+1}(p) = \sigma_n(p-1) + \sum_{k=1}^{n-(p-1)} a_{p+k-1,p}\sigma_n(p+k-1) \tag{12}$$

with

$$a_{p+k-1,p} = \frac{g_k}{k!}\prod_{j=0}^{k-1}(p+j) = g_k\binom{p+k-1}{k}. \tag{13}$$

In an alternative way, setting $q = p+k-1$,

$$\sigma_{n+1}(p) = \sigma_n(p-1) + \sum_{q=p}^{n} a_{q,p}\sigma_n(q), \quad p=1,\dots,n,\ n\geqslant 1, \tag{14}$$

with

$$a_{q,p} = g_{q-(p-1)}\binom{q}{p-1}.$$

Recurrences (12) or (14) constitute the announced generalization of (6).

We now interpret recurrences (12), (14) in terms of nucleation–aggregation phenomena, as described in section 2.

In such extended models, indeed, the additional atom in the transition $n \to n+1$, when it sees a clusters' situation with q groups, may connect to k groups simultaneously, $k = 1, \ldots, \min(K, q)$, where $K \stackrel{\text{def}}{=} \max(k \geqslant 1\colon\ g_k \neq 0)$ is the (possibly infinite) order of the branch generating function $g(\theta)$. This k-connection (fusion) occurs with transition probability

$$\mathbf{Q}_{n,q}(k) \stackrel{\text{def}}{=} \frac{a_{q,q-(k-1)}}{(1+\sum_{p=1}^{n} a_{q,p})}. \tag{15}$$

By doing so, the number of groups shifts from $q = p+k-1$ to p and decreases as soon as $k \geqslant 2$ (cluster–cluster aggregation).

Of course, nucleation occurs with probability

$$\mathbf{Q}_{n,q}(0) \stackrel{\text{def}}{=} \frac{1}{(1+\sum_{p=1}^{n} a_{q,p})} \tag{16}$$

and still remains possible.

In other words, equation (5) of section 2 has to be replaced by

$$P_{n+1} = (P_n + 1) \cdot \mathbf{1}_{K_{n+1}=0} + (P_n - K_{n+1} + 1) \cdot \mathbf{1}_{K_{n+1}>0}, \tag{17}$$

where K_{n+1} is the random variable giving the number of connections with transition distribution

$$\mathbf{Q}(K_{n+1} = k \mid P_n = q) = \mathbf{Q}_{n,q}(k), \quad k \geqslant 0. \tag{18}$$

The coefficients $a_{p+k-1,k}$ in (13) entering in the definition of these probabilities are now easy to interpret: they are the number of ways that the inserted atom will select k connection-groups out of $p+k-1$ possible, as soon as $g_k \neq 0$, (the term $\binom{p+k-1}{k}$), times g_k which is the number of ways to realize this k-fusion.

If g_k is "large", the connection probability will be large, so that nucleation is weak: in the asymptotic $n \to \infty$ there should be "quite few" clusters compared to a situation where g_k is "small" in which nucleation dominates. One of the purpose of the following is to quantify this mere observation.

Example 1. Let us give some examples that shows that models (a) and (b) actually are particular cases of this new interpretation.

(a) $g(\theta) = 1 + \theta$ for which $g_1 = 1,\ g_k = 0,\ k \geqslant 2$, leads to $\overline{t}(\theta) = \exp\theta - 1$ and $Z^1(\theta, \gamma) = \mathrm{e}^{\gamma(\mathrm{e}^{\theta}-1)}$, and from (12) we get (6(a))

$$\sigma_{n+1}(p) = \sigma_n(p-1) + p\sigma_n(p).$$

(b) $g(\theta) = \mathrm{e}^{\theta}$ for which $g_k = 1,\ k \geqslant 1$, leads to $\overline{t}(\theta) = -\log(1-\theta)$ and $Z^1(\theta, \gamma) = (1-\theta)^{-\gamma}$.

Recurrences (12) obtained while inserting $g_k = 1,\ k \geqslant 1$, in (13) constitute an alternative interpretation to 6(b). In this case, connection with any number of preexisting groups is allowed to take place.

(c) $g(\theta) = 1 + \theta^2$ for which $g_1 = 0,\ g_2 = 2,\ g_k = 0,\ k \geqslant 3$, leads to $\bar{t}(\theta) = \tan\theta$ and $Z^1(\theta, \gamma) = \mathrm{e}^{\gamma \tan\theta}$;

$$\sigma_{n+1}(p) = \sigma_n(p-1) + p(p+1)\sigma_n(p+1).$$

3.1.2. Partition function for fragment size distribution

Thus, the bivariate "marked" exponential generating function of (10)

$$Z^1(\theta, \gamma) \stackrel{\text{def}}{=} \mathrm{e}^{\gamma \bar{t}(\theta)},$$

where

$$\bar{t}(\theta) \stackrel{\text{def}}{=} \sum_{j \geqslant 1} \frac{c_j}{j!}\theta^j$$

is the unique solution to the autonomous ordinary differential equation

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\bar{t}(\theta) \stackrel{\text{def}}{=} \bar{t}(\theta)' = g\big(\bar{t}(\theta)\big),$$

appears crucial in the apprehension of nucleation–aggregation models.

Observing next that

$$Z^1(\theta, \gamma) = \prod_{j \geqslant 1}\left(1 + \sum_{l \geqslant 1}\frac{\gamma^l}{l!}\left(\frac{c_j\theta^j}{j!}\right)^l\right) = \prod_{j \geqslant 1}\mathrm{e}^{\gamma c_j \theta^j/j!}$$

we conclude that aggregates are obtained from the repetition l times of size-j clusters, each of which presenting a "variability" c_j. It follows that if one intends to understand the clusters size distributions, one should focus on the "marked" partition function

$$Z^1(\theta, \gamma, \gamma_1, \ldots, \gamma_j, \ldots) = \prod_{j \geqslant 1}\left(1 + \sum_{l \geqslant 1}\frac{(\gamma\gamma_j)^l}{l!}\left(\frac{c_j\theta^j}{j!}\right)^l\right) = \mathrm{e}^{\gamma\sum_{j\geqslant 1}\gamma_j c_j\theta^j/j!} \tag{19}$$

entering into additional details: here, $(\gamma_j)_{j\geqslant 1}$ "marks" the number N_j of size-j clusters, whereas γ "marks" the number of clusters variable, say P, summing up over j: $P = \sum_{j\geqslant 1} N_j$, which is in accordance with (2).

Developing,

$$Z^1(\theta, \gamma, \gamma_1, \ldots, \gamma_j, \ldots) \stackrel{\text{def}}{=} 1 + \sum_{n \geqslant 1}\frac{\theta^n}{n!}Z_n^1(\gamma, \gamma_1, \ldots, \gamma_n) \tag{20}$$

with

$$Z_n^1(\gamma, \gamma_1, \ldots, \gamma_n) \stackrel{\text{def}}{=} \sum_{p=1}^{n}\gamma^p \sum \Omega(n_1, \ldots, n_n)\prod_{j=1}^{n}\gamma_j^{n_j}. \tag{21}$$

In this formula, the "small" nested sum is to be performed over the integers $n_1, \ldots, n_n \geqslant 0$ under the constraints (2),

$$\sum_{j=1}^{n} n_j = p, \qquad \sum_{j=1}^{n} j n_j = n,$$

and

$$\Omega(n_1, \ldots, n_n) \stackrel{\text{def}}{=} n! \prod_{j=1}^{n} \frac{(c_j)^{n_j}}{n_j!(j!)^{n_j}} \tag{22}$$

is their Boltzmann degeneracy.

A particular partition function of interest in the sequel is from (19)

$$Z^1(\theta, \gamma, \gamma_m) = \mathrm{e}^{\gamma(\bar{t}(\theta)+(\gamma_m-1)c_m\theta^m/m!)} \stackrel{\text{def}}{=} \mathrm{e}^{\gamma \bar{t}(\theta,\gamma_m)} \tag{23}$$

with

$$\bar{t}(\theta, \gamma_m) \stackrel{\text{def}}{=} \left(\bar{t}(\theta) + (\gamma_m - 1)\frac{c_m\theta^m}{m!}\right), \tag{24}$$

which focuses on the size-m clusters only, setting $\gamma_j = 1$ for all indices j except $j = m$ in equation (19).

3.2. Discernible atoms and discernible clusters (model 2)

3.2.1. Partition function for the number of clusters distribution

We shall next consider the bivariate "marked" exponential generating function

$$Z^2(\theta, \gamma) \stackrel{\text{def}}{=} \frac{1}{1 - \gamma \bar{t}(\theta)} \tag{25}$$

with $Z^2(\theta) \stackrel{\text{def}}{=} Z^2(\theta, 1)$.

Developing,

$$Z^2(\theta, \gamma) \stackrel{\text{def}}{=} 1 + \sum_{n \geqslant 1} \frac{\theta^n}{n!} Z_n^2(\gamma)$$

with

$$Z_n^2(\gamma) \stackrel{\text{def}}{=} \sum_{p=1}^{n} \sigma_n'(p)\gamma^p.$$

In this interpretation,

$$\sigma_n'(p) \stackrel{\text{def}}{=} p!\sigma_n(p) \tag{26}$$

counts the number of forests made of p *discernible* increasing trees that one can form with n discernible atoms; let $\sigma'_n \stackrel{\text{def}}{=} \sum_{p=1}^{n} \sigma'_n(p)$. Note that, from (12), there is a recursive way to generate the sequence $\sigma'_n(p)$, namely,

$$\sigma'_{n+1}(p) = p\sigma'_n(p-1) + \sum_{k=1}^{n-(p-1)} a'_{p+k-1,p}\sigma'_n(p+k-1) \tag{27}$$

with $a'_{p+k-1,p} = pg_k/k!$ this time.

Partition function $Z^2(\theta,\gamma)$, therefore, concentrates on configuration numbers of aggregates when both constitutive atoms and clusters are distinguishable (e.g., labeled). One expects the conclusions to be highly sensitive to this clusters' distinguishability.

3.2.2. Partition function of fragment size distribution

In a similar way, if one intends to understand the clusters size distributions, one should consider the "marked" partition function

$$Z^2(\theta,\gamma,\gamma_1,\ldots,\gamma_j,\ldots) = \frac{1}{1-\gamma\sum_{j\geqslant 1}\gamma_j c_j\theta^j/j!}. \tag{28}$$

Developing,

$$Z^2(\theta,\gamma,\gamma_1,\ldots,\gamma_j,\ldots) \stackrel{\text{def}}{=} 1 + \sum_{n\geqslant 1}\frac{\theta^n}{n!}Z_n^2(\gamma,\gamma_1,\ldots,\gamma_n) \tag{29}$$

with

$$Z_n^2(\gamma,\gamma_1,\ldots,\gamma_n) \stackrel{\text{def}}{=} \sum_{p=1}^{n}\gamma^p\sum\Omega'(n_1,\ldots,n_n)\prod_{j=1}^{n}\gamma_j^{n_j} \tag{30}$$

and degeneracy system

$$\Omega'(n_1,\ldots,n_n) \stackrel{\text{def}}{=} p!n!\prod_{j=1}^{n}\frac{(c_j)^{n_j}}{n_j!(j!)^{n_j}}. \tag{31}$$

A particular partition function of interest is from (28)

$$Z^2(\theta,\gamma,\gamma_m) = \frac{1}{1-\gamma(\bar{t}(\theta)+(\gamma_m-1)c_m\theta^m/m!)} \stackrel{\text{def}}{=} \frac{1}{1-\gamma\bar{t}(\theta,\gamma_m)}, \tag{32}$$

which focuses on the size-m clusters only, setting $\gamma_j = 1$ for all indices j except $j = m$ in equation (28).

4. Random models for fragment size distributions

We now consider the averaging problem over the configurations.

Let

$$Z^1(\theta,\gamma,\gamma_1,\ldots,\gamma_j,\ldots)=1+\sum_{n\geqslant 1}\frac{\theta^n}{n!}Z_n^1(\gamma,\gamma_1,\ldots,\gamma_n)$$

be the first multivariate generating function introduced before.

Suppose one is able to extract $[\theta^n]Z^1(\theta,\gamma,\gamma_1,\ldots,\gamma_j,\ldots)$, or possibly its asymptotic equivalent, for large n.

Here,

$$[\theta^n]Z^1(\theta,\gamma,\gamma_1,\ldots,\gamma_j,\ldots)=\frac{1}{n!}Z_n^1(\gamma,\gamma_1,\ldots,\gamma_n)$$

stands for the coefficient of θ^n in the power-series expansion of $Z^1(\theta,\gamma,\gamma_1,\ldots,\gamma_j,\ldots)$.

Then, upon normalizing,

$$\mathbf{\Phi}_n^1(\gamma,\gamma_1,\ldots,\gamma_n)\stackrel{\text{def}}{=}\frac{[\theta^n]Z^1(\theta,\gamma,\gamma_1,\ldots,\gamma_j,\ldots)}{[\theta^n]Z^1(\theta)}=\frac{Z_n^1(\gamma,\gamma_1,\ldots,\gamma_n)}{Z_n^1(1,1,\ldots,1)} \tag{33}$$

will stand for the joint probability generating function of the number of clusters variable P, and clusters' size variables $(N_j)_{j=1}^n$, *given* the number of atoms is n, as a conditional probability generating function. In an alternative way,

$$\mathbf{P}^1(P=p,\ N_1=n_1,\ldots,N_n=n_n\mid N=n)=\frac{\mathbf{\Omega}(n_1,\ldots,n_n)}{\sigma_n}, \tag{34}$$

where $n_1,\ldots,n_n\geqslant 0$ are subject to the constraints $\sum_{j=1}^n n_j=p,\ \sum_{j=1}^n jn_j=n$.

Stated differently,

$$\mathbf{\Phi}_n^1(\gamma,\gamma_1,\ldots,\gamma_n)\stackrel{\text{def}}{=}\mathbf{E}^1\left[\gamma^P\gamma_1^{N_1}\ldots\gamma_n^{N_n}\,\middle|\,N=n\right]. \tag{35}$$

In a similar fashion, under model 2, we randomize the variables $N,\ P,\ (N_j)_{j=1}^n$ by

$$\mathbf{\Phi}_n^2(\gamma,\gamma_1,\ldots,\gamma_n)\stackrel{\text{def}}{=}\mathbf{E}^2\left[\gamma^P\gamma_1^{N_1}\ldots\gamma_n^{N_n}\,\middle|\,N=n\right]$$

with

$$\mathbf{\Phi}_n^2(\gamma,\gamma_1,\ldots,\gamma_n)\stackrel{\text{def}}{=}\frac{[\theta^n]Z^2(\theta,\gamma,\gamma_1,\ldots,\gamma_j,\ldots)}{[\theta^n]Z^2(\theta)}=\frac{Z_n^2(\gamma,\gamma_1,\ldots,\gamma_n)}{Z_n^2(1,1,\ldots,1)}. \tag{36}$$

4.1. Randomizing the number of atoms: from combinatorics to discrete probability

In general, the form of the fragment size distribution will change drastically as a function of some macroscopically observable variable, which is identified with a control parameter. The most natural such variable in our statistical approach is the number of atoms. Another refinement could add the number of fragments of any size produced in one aggregation; we shall not follow this path for the sake of simplicity and shall restrict ourselves to the sole first observable. It now proves useful to discuss the randomization of the number of atoms variable.

Suppose labeled (discernible) elementary objects called "atoms" are given the opportunity to be interconnected in a certain number of ways. By an interconnection of atoms (or network), we mean an oriented graph between these atoms. Specifying the admissible type of graph produces a combinatorial model of atoms. An interconnection model of atoms therefore requires to fix an integer-valued non-decreasing sequence $(s_n)_{n\geqslant 0}$ giving the configuration number of n atoms, that is, the number of ways these n labeled atoms can be interconnected.

These informations are advantageously encapsulated within the exponential partition function

$$Z(\theta) = 1 + \sum_{n\geqslant 1} \frac{s_n}{n!}\theta^n,$$

as a function of the real variable $\theta \geqslant 0$, varying in some definition domain

$$D^+ \stackrel{\text{def}}{=} \{\theta\colon\ 0 \leqslant \theta \leqslant \theta_0\} \quad \text{or} \quad D^+ \stackrel{\text{def}}{=} \{\theta\colon\ 0 \leqslant \theta < \theta_0\}$$

for some positive real number θ_0 (possibly infinite).

This number is the convergence radius of the series $Z(\theta)$ in the sense that Z admits a convergent power series expansion in D^+.

Identifying various exponential partition function $Z(\theta)$ belongs to the field of enumerative combinatorics, which is part of the field of graph theory. We shall limit ourselves here to the partition functions introduced in this paper, together with their physical meaning, namely,

$$Z^1(\theta) \stackrel{\text{def}}{=} \mathrm{e}^{\overline{t}(\theta)} = 1 + \sum_{n\geqslant 1} \frac{\sigma_n}{n!}\theta^n \quad (Z = Z^1,\ s_n = \sigma_n),$$

in the context of discernible atoms and indiscernible clusters, and

$$Z^2(\theta) \stackrel{\text{def}}{=} \frac{1}{1-\overline{t}(\theta)} = 1 + \sum_{n\geqslant 1} \frac{\sigma'_n}{n!}\theta^n \quad (Z = Z^2,\ s_n = \sigma'_n)$$

in the context of discernible atoms and discernible clusters, respectively.

For such Z, the convergence radius θ_0 will be shown to be finite in any case.

Discrete probability and combinatorics are now closely related in the following way: the sequence of coefficients $(s_n)_{n\geqslant 0}$ allows one to define the "prior" reference measure $\mathbf{R}$ of the event $N = n$, to "meet" n *discernible* atoms as

$$\mathbf{R}(N = n) \stackrel{\text{def}}{=} \mathbf{R}(n) = \frac{s_n}{n!}. \tag{37}$$

Note that this measure is not a probability measure, since it is not summable. It only is a positive "combinatorial" measure of this event. Assume now the exact number of such interconnected atoms, say N, is unknown to some observer so that N is assumed random. We shall then search for a "probability" measure of the event

$N = n$, say $\mathbf{P}(N = n) \stackrel{\text{def}}{=} \mathbf{P}(n)$, which minimizes the Kullback information between $\mathbf{P}$ and $\mathbf{R}$ [17,20]:

$$K(\mathbf{P} \parallel \mathbf{R}) \stackrel{\text{def}}{=} \sum_{n \geqslant 0} \mathbf{P}(n) \log \frac{\mathbf{P}(n)}{\mathbf{R}(n)} \tag{38}$$

under the constraints

$$\sum_{n \geqslant 0} \mathbf{P}(n) = 1 \quad \text{and} \quad \sum_{n \geqslant 0} n\mathbf{P}(n) = x \stackrel{\text{def}}{=} \langle N \rangle \geqslant 0, \tag{39}$$

fixing the non-negative average x of the probability distribution $\mathbf{P}$.

Performing this standard optimization program using Lagrange multipliers yields

$$\mathbf{P}(n) \stackrel{\text{def}}{=} \mathbf{P}_\theta(N = n) = \frac{s_n \theta^n}{Z(\theta) n!}, \quad n \geqslant 0,\ \theta \in D^+, \tag{40}$$

with θ and $x = \langle N \rangle$ related by

$$\theta = \mathrm{e}^{-\beta} \quad \text{and} \quad F'(\beta) = x, \tag{41}$$

where $F(\beta) \stackrel{\text{def}}{=} -\log Z(\mathrm{e}^{-\beta})$.

Thus, a "good" model for the probability to observe n atoms clearly is the "exponential" Gibbs family

$$\mathbf{P}_\theta(N = n) \stackrel{\text{def}}{=} \frac{s_n \theta^n}{Z(\theta) n!}, \quad n \geqslant 0,\ \theta \in D^+, \tag{42}$$

as a function of an "external" control parameter $\theta \in D^+$, related to the theoretical average x of the distribution as just mentioned. This actually is one of the postulates of statistical physics. Statistics is then concerned with the problem of identifying the value of θ which fits the best some observation sample (see section 4.1.1).

If $\Phi_\theta(u) \stackrel{\text{def}}{=} \sum_{n \geqslant 0} \mathbf{P}_\theta(N = n) u^n$ is now the associated probability generating function of this probability distribution, we have

$$\Phi_\theta(u) \stackrel{\text{def}}{=} \frac{Z(\theta u)}{Z(\theta)}, \tag{43}$$

relating partition function to probability generating function.

Specifying $Z = Z^1,\ s_n = \sigma_n$ and $Z = Z^2,\ s_n = \sigma'_n$ yields the definition of two probabilities for both models, namely,

$$\mathbf{P}^1_\theta(N = n) \stackrel{\text{def}}{=} \frac{\sigma_n \theta^n}{Z^1(\theta) n!} \quad \text{and} \quad \mathbf{P}^2_\theta(N = n) \stackrel{\text{def}}{=} \frac{\sigma'_n \theta^n}{Z^2(\theta) n!},$$

respectively.

This connects the worlds of combinatorics to the one of discrete probability.

Remark 1. Note that this randomization is only possible in the range $0 \leqslant \theta < \theta_0$ (possibly including $\theta = \theta_0$ if $Z(\theta_0)$ remains finite there) of the control parameter.

Otherwise, if $\theta > \theta_0$ the random variable N becomes degenerate, in the sense that $N = \infty$ with $\mathbf{P}_\theta$ probability one, as soon as $\theta \geqslant \theta_0$ (possibly $\theta > \theta_0$): N becomes unobservable.

Parameter θ_0 is, therefore, a *critical* parameter, in the sense that a phase transition takes places there. Both location of the singularity, θ_0, *and* behavior of the partition functions Z^1 and Z^2 at the singularity will, therefore, become essential for full understanding of this phase transition.

If $Z(\theta_0)$ remains finite, function Z will be said to be regular (and this will be shown to occur in our case studies). The randomization of the number of atoms still remains possible at the critical value $\theta = \theta_0$. This happens although the average number of atoms $\langle N \rangle$ diverges there, from (41), as a consequence of $Z'(\theta_0) = \infty$ (steepness of the derivative). The random variable, defined by its probability generating function

$$\Phi_{\theta_0}(u) \stackrel{\text{def}}{=} \frac{Z(\theta_0 u)}{Z(\theta_0)} \quad \text{at } \theta = \theta_0,$$

can easily be shown in our case to be discrete-stable of parameter $1/2$ in the sense of [23] for which $\Phi_{\theta_0}(u) = \exp(-\lambda(1-u)^{1/2})$ for some positive λ. For such distributions, only fractional moments of order strictly less than $1/2$ can be shown to exist as a result of the slow decay of the probability system $\mathbf{P}_{\theta_0}(N = n) \underset{n\to\infty}{\sim} n^{-3/2}$.

4.1.1. Identifying parameters from sampling

Let us now come to the classical question (in statistics) of identifying the value of θ which fits the best some observation sample in some sense. We shall recall how to construct a maximum likelihood estimator of θ.

Substituting parameter β to parameter θ as mentioned above, the distribution of the number of atoms N, under $\mathbf{P}_\theta$, takes the new form

$$\mathbf{P}_\beta(n) \stackrel{\text{def}}{=} \frac{s_n \mathrm{e}^{-\beta n}}{\alpha(\beta) n!}, \quad n \geqslant 0, \ \beta \in \Delta^+, \tag{44}$$

where $\alpha(\beta) = Z(\mathrm{e}^{-\beta})$ is the Laplace transform of the sequence $(s_n/n!)_{n\geqslant 0}$ and

$$\Delta^+ \stackrel{\text{def}}{=} \{\beta\colon\ \beta \geqslant \beta_0 \stackrel{\text{def}}{=} -\log\theta_0\} \quad \text{or} \quad \Delta^+ \stackrel{\text{def}}{=} \{\beta\colon\ \beta > \beta_0 \stackrel{\text{def}}{=} -\log\theta_0\}$$

is its definition domain obtained after an easy distortion of D^+. We shall also need its log-Laplace transform $F(\beta) \stackrel{\text{def}}{=} -\log\alpha(\beta)$. This function is negative and concave on the convex set Δ^+. Its Legendre transform

$$f(x) \stackrel{\text{def}}{=} \inf_{\beta\in\Delta^+} \big(\beta x - F(\beta)\big) \tag{45}$$

is well-defined, non-negative and concave on the convex set $x \geqslant 0$. Moreover, $f(x) = xf'(x) - F(f'(x))$, with $\mathbf{E}_\beta[N] = F'(\beta) = x$ and $\beta = f'(x)$. Symbol x is, thus,

identified with the average number of atoms $\mathbf{E}_\beta[N]$ under $\mathbf{P}_\beta$, and the control parameter β can be derived from x. Distribution $\mathbf{P}_\theta$ is well-parameterized by x, through

$$\mathbf{P}_x(n) \stackrel{\text{def}}{=} \frac{s_n \mathrm{e}^{-f'(x)n}}{\alpha(f'(x))n!}, \quad n \geqslant 0,\ x \geqslant 0. \tag{46}$$

Note, also, that $f(x) = -K(\mathbf{P}_x \parallel \mathbf{R})$ is the opposite of the value of the Kullback information [13] evaluated at $\mathbf{P} = \mathbf{P}_x$.

A maximum likelihood estimator of θ, say Θ^*, can, therefore, be derived from an estimator X^* of the average number of atoms by $\Theta^* = -\log f'(X^*)$.

Let now $(N_k)_{k=1}^K$be an independent K-sample of the random variable N. Introducing the likelihood $V_x(N_1, \ldots, N_K) = \prod_{k=1}^K \mathbf{P}_x(N_k)$ and searching for the value of x maximizing this likelihood, we get

$$X_K^* = \frac{1}{K}\sum_{k=1}^K N_k.$$

The experimental mean is thus an unbiased, efficient estimator of x, in the sense that the expectation and variance under $\mathbf{P}_x$ are $\mathbf{E}_x[X_K^*] = x$, and $\boldsymbol{\sigma}_x^2[X_K^*] = -1/(Kf''(x))$, with variance function $V(x) \stackrel{\text{def}}{=} -1/f''(x) > 0$. The quantity $-Kf''(x)$ is the Fisher information of the K-sample.

We also have the law of large numbers

$$X_K^* \underset{K\to\infty}{\longrightarrow} x \tag{47}$$

with $\mathbf{P}_x$ probability one, and the central limit theorem

$$\lim_{K\to\infty} \mathbf{P}_x\left(\frac{X_K^* - \mathbf{E}_x[X_K^*]}{\boldsymbol{\sigma}_x[X_K^*]} \leqslant y\right) = \mathrm{erf}(y), \tag{48}$$

together with its large deviation counterpart [1]

$$\frac{1}{K}\log \mathbf{P}_x\big(X_K^* \geqslant y\big) \underset{K\to\infty}{\longrightarrow} \int_x^y (y-z)f''(z)\,\mathrm{d}z < 0 \tag{49}$$

for $y > x$. The integral appearing in the right-hand side is easily seen to be the Legendre transform $f_x(y) \stackrel{\text{def}}{=} \inf_\eta(\eta y - F_x(\eta))$ with $F_x(\eta) = F(\eta + f'(x)) - F(f'(x))$. Here $F_x(\eta) = -\log \alpha_x(\eta)$ with $\alpha_x(\eta) = \alpha(f'(x)+\eta)/\alpha(f'(x))$ the Laplace transform of the probability $\mathbf{P}_x$ defined from (46). Moreover, it can be checked that $-f_x(y) = K(\mathbf{P}_y \parallel \mathbf{P}_x) > 0$ which is the positive Kullback information between $\mathbf{P}_y$ and $\mathbf{P}_x$.

If $0 < y < x$, we get in a similar way

$$\frac{1}{K}\log \mathbf{P}_x\big(X_K^* \leqslant y\big) \underset{K\to\infty}{\longrightarrow} \int_y^x (z-y)f''(z)\,\mathrm{d}z < 0.$$

Example 2. Consider $Z^2(\theta) = 1/(1-\overline{t}(\theta))$ with $\overline{t}(\theta) = 1 - \sqrt{1-2\theta}$ which can be derived, using (9), from the particular choice $g(\theta) = 1/(1-\theta)$. In this example,

$\alpha(\beta) \stackrel{\text{def}}{=} Z^2(\mathrm{e}^{-\beta}) = (1-2\mathrm{e}^{-\beta})^{-1/2}$ and $F(\beta) \stackrel{\text{def}}{=} -\log\alpha(\beta) = (1/2)\log(1-2\mathrm{e}^{-\beta})$. It follows that $F'(\beta) = 2\mathrm{e}^{-\beta}/(1-2\mathrm{e}^{-\beta}) = x$. Hence, upon inverting, $\beta = f'(x) = -\log(x/(2(1+x)))$. As a result, $f(x) \stackrel{\text{def}}{=} xf'(x) - F(f'(x)) = -x\log(x/2) + (x+1/2)\log(1+x) \geqslant 0$ for $x \geqslant 0$ and $f''(x) = -1/(x(1+x)) < 0$.

With these considerations at hand, it is now possible to discuss the joint distributions of the random vectors $(N, P, N_1, \ldots, N_j, \ldots)$ under $\mathbf{P}_\theta = \mathbf{P}^1_\theta$ for model 1 and $\mathbf{P}_\theta = \mathbf{P}^2_\theta$ for model 2.

4.2. Joint distributions in the sub-critical region ($\theta < \theta_0$)

4.2.1. Joint distributions for model 1

Recall from (33) that

$$\Phi^1_n(\gamma, \gamma_1, \ldots, \gamma_n) \stackrel{\text{def}}{=} \frac{Z^1_n(\gamma, \gamma_1, \ldots, \gamma_n)}{Z^1_n(1, 1, \ldots, 1)} = \frac{Z^1_n(\gamma, \gamma_1, \ldots, \gamma_n)}{\sigma_n}$$

has been interpreted as the conditional multivariate probability generating function

$$\Phi^1_n(\gamma, \gamma_1, \ldots, \gamma_n) \stackrel{\text{def}}{=} \mathbf{E}^1\left[\gamma^P \gamma_1^{N_1} \ldots \gamma_n^{N_n} \,\middle|\, N = n\right]$$

for the number of clusters variable P and size-j clusters contributions $(N_j)_{j=1}^n$. Multiplying by the probability

$$\mathbf{P}^1_\theta(N = n) \stackrel{\text{def}}{=} \frac{\sigma_n \theta^n}{Z^1(\theta) n!}, \quad n \geqslant 0,$$

that n atoms are being observed and summing up over $n \geqslant 0$ yields now the joint probability

$$\begin{aligned}\Phi^1_\theta(u, \gamma, \gamma_1, \ldots, \gamma_n) &\stackrel{\text{def}}{=} \mathbf{E}^1_\theta\left[u^N \gamma^P \gamma_1^{N_1} \ldots \gamma_n^{N_n}\right] \\ &= \sum_{n \geqslant 0} \mathbf{P}^1_\theta(N = n) u^n \Phi^1_n(\gamma, \gamma_1, \ldots, \gamma_n) \qquad (50)\end{aligned}$$

as a function of the control parameter θ.

Hence,

$$\begin{aligned}\Phi^1_\theta(u, \gamma, \gamma_1, \ldots, \gamma_j, \ldots) &\stackrel{\text{def}}{=} \frac{1}{Z^1(\theta)} \sum_{n \geqslant 0} \frac{1}{n!} (\theta u)^n Z^1_n(\gamma, \gamma_1, \ldots, \gamma_n) \\ &= \frac{Z^1(\theta u, \gamma, \gamma_1, \ldots, \gamma_j, \ldots)}{Z^1(\theta)} \qquad (51)\end{aligned}$$

from (20). In explicit form,

$$\begin{aligned}\Phi^1_\theta(u, \gamma, \gamma_1, \ldots, \gamma_j, \ldots) &= \mathrm{e}^{\gamma \sum_{j \geqslant 1} \gamma_j c_j (\theta u)^j / j!} \Big/ \mathrm{e}^{\sum_{j \geqslant 1} c_j \theta^j / j!} \\ &= \mathrm{e}^{\sum_{j \geqslant 1} (\gamma \gamma_j u^j - 1) c_j \theta^j / j!} \qquad (52)\end{aligned}$$

from (19).

This formulation is full of interesting informations.

Zooming, for example, on size-m clusters yields, taking $\gamma_j = 1$, for all j but m,

$$\begin{aligned}\Phi^1_\theta(u,\gamma,\gamma_m) &\stackrel{\text{def}}{=} \mathbf{E}^1_\theta\big[u^N\gamma^P\gamma_m^{N_m}\big]\\ &= \exp\left(-\overline{t}(\theta)+\gamma\left\{\frac{c_m(\theta u)^m}{m!}(\gamma_m-1)+\overline{t}(u\theta)\right\}\right)\end{aligned} \tag{53}$$

after some easy computations.

In particular,

$$\Phi^1_\theta(1,1,\gamma_m) \stackrel{\text{def}}{=} \mathbf{E}^1_\theta\big[\gamma_m^{N_m}\big] = \exp\left((\gamma_m-1)\frac{c_m\theta^m}{m!}\right) \tag{54}$$

is the probability generating function of a Poisson variable, with intensity $\mathbf{E}^1_\theta[N_m] = c_m\theta^m/m!$ and standard deviation $\boldsymbol{\sigma}^1_\theta[N_m] = (c_m\theta^m/m!)^{1/2}$. Another marginal distribution is

$$\Phi^1_\theta(1,\gamma,1) \stackrel{\text{def}}{=} \mathbf{E}^1_\theta\big[\gamma^P\big] = \exp\big((\gamma-1)\overline{t}(\theta)\big), \tag{55}$$

which is the probability generating function of a Poisson variable, with intensity $\mathbf{E}^1_\theta[P] = \overline{t}(\theta) = \sum_{j\geqslant 1} c_j\theta^j/j!$, and

$$\Phi^1_\theta(u,1,1) \stackrel{\text{def}}{=} \mathbf{E}^1_\theta\big[u^N\big] = \exp\big(\overline{t}(u\theta)-\overline{t}(\theta)\big) \tag{56}$$

is a Gibbs variable with mean $\mathbf{E}^1_\theta[N] = \theta\overline{t}'(\theta)$ and standard deviation $\boldsymbol{\sigma}^1_\theta[N] = (\theta\overline{t}'(\theta)+\theta^2\overline{t}''(\theta))^{1/2}$.

Joint and conditional informations are also available, for example,

$$\begin{aligned}\Phi^1_\theta(1,\gamma,\gamma_m) &\stackrel{\text{def}}{=} \mathbf{E}^1_\theta\big[\gamma^P\gamma_m^{N_m}\big]\\ &= \exp\left(-\overline{t}(\theta)+\gamma\left\{(\gamma_m-1)\frac{c_m\theta^m}{m!}+\overline{t}(\theta)\right\}\right).\end{aligned} \tag{57}$$

As a result,

$$\begin{aligned}\mathbf{E}^1_\theta\big[\gamma^P\gamma_m^{N_m}\big] &= \sum_{p\geqslant 0}\frac{\overline{t}(\theta)^p\gamma^p\mathrm{e}^{-\overline{t}(\theta)}}{p!}\left\{(\gamma_m-1)\frac{c_m\theta^m}{m!\overline{t}(\theta)}+1\right\}^p\\ &= \sum_{p\geqslant 0}\mathbf{P}^1_\theta(P=p)\left\{(\gamma_m-1)\frac{c_m\theta^m}{m!\overline{t}(\theta)}+1\right\}^p.\end{aligned}$$

Hence,

$$\mathbf{E}^1_\theta\big[\gamma_m^{N_m}\,\big|\,P=p\big] = \left\{(\gamma_m-1)\frac{c_m\theta^m}{m!\overline{t}(\theta)}+1\right\}^p$$

is the probability generating function of a binomial variable whose mean is $\mathbf{E}^1_\theta[N_m \mid P=p] = p_\theta(m)p$, with $p_\theta(m) \stackrel{\text{def}}{=} \mathbf{E}^1_\theta[N_m]/\mathbf{E}^1_\theta[P] = c_m\theta^m/(m!\overline{t}(\theta))$.

4.2.2. Joint distributions for model 2

In a similar way, from (36),

$$\mathbf{\Phi}_n^2(\gamma,\gamma_1,\ldots,\gamma_n)\stackrel{\text{def}}{=}\frac{Z_n^2(\gamma,\gamma_1,\ldots,\gamma_n)}{\sigma'_n}=\mathbf{E}^2\left[\gamma^P\gamma_1^{N_1}\ldots\gamma_n^{N_n}\,\middle|\,N=n\right]$$

is the conditional multivariate probability generating function for the number of clusters variable P and the number of size-j clusters variables $(N_j)_{j=1}^n$ given $N=n$. Multiplying by the new probability

$$\mathbf{P}_\theta^2(N=n)\stackrel{\text{def}}{=}\frac{\sigma'_n\theta^n}{Z^2(\theta)n!},\quad n\geqslant 0,$$

that n atoms are being observed and summing up over $n\geqslant 0$ yields now the joint probability

$$\begin{aligned}\mathbf{\Phi}_\theta^2(u,\gamma,\gamma_1,\ldots,\gamma_n)&\stackrel{\text{def}}{=}\mathbf{E}_\theta^2\left[u^N\gamma^P\gamma_1^{N_1}\ldots\gamma_n^{N_n}\right]\\&=\sum_{n\geqslant 0}\mathbf{P}_\theta^2(N=n)u^n\mathbf{\Phi}_n^2(\gamma,\gamma_1,\ldots,\gamma_n)\end{aligned}\tag{58}$$

as a function of the control parameter θ.

Hence,

$$\mathbf{\Phi}_\theta^2(u,\gamma,\gamma_1,\ldots,\gamma_j,\ldots)=\frac{Z^2(\theta u,\gamma,\gamma_1,\ldots,\gamma_j,\ldots)}{Z^2(\theta)}\tag{59}$$

from (29) and, in explicit form,

$$\mathbf{\Phi}_\theta^2(u,\gamma,\gamma_1,\ldots,\gamma_j,\ldots)=\frac{1-\overline{t}(\theta)}{1-\gamma\sum_{j\geqslant 1}\gamma_j u^j c_j\theta^j/j!}\tag{60}$$

from (28).

Zooming on size-m clusters yields, taking $\gamma_j=1$, for all j but m,

$$\mathbf{\Phi}_\theta^2(u,\gamma,\gamma_m)\stackrel{\text{def}}{=}\mathbf{E}_\theta^2\left[u^N\gamma^P\gamma_m^{N_m}\right]=\frac{1-\overline{t}(\theta)}{1-\gamma\overline{t}(\theta u)-\gamma(c_m(\theta u)^m/m!)(\gamma_m-1)}\tag{61}$$

after some easy computations.

In particular,

$$\mathbf{\Phi}_\theta^2(1,1,\gamma_m)\stackrel{\text{def}}{=}\mathbf{E}_\theta^2\left[\gamma_m^{N_m}\right]=\frac{1-\overline{t}(\theta)}{1-\overline{t}(\theta)-(c_m\theta^m/m!)(\gamma_m-1)}\tag{62}$$

is the probability generating function of a geometric variable with mean $\mathbf{E}_\theta[N_m]=c_m\theta^m/(m!(1-\overline{t}(\theta)))$,

$$\mathbf{\Phi}_\theta^2(1,\gamma,1)\stackrel{\text{def}}{=}\mathbf{E}_\theta^2\left[\gamma^P\right]=\frac{1-\overline{t}(\theta)}{1-\gamma\overline{t}(\theta)}\tag{63}$$

is the probability generating function of a geometric variable with intensity $\mathbf{E}^2_\theta[P] = \overline{t}(\theta)/(1-\overline{t}(\theta))$, and

$$\Phi^2_\theta(u,1,1) \stackrel{\text{def}}{=} \mathbf{E}^2_\theta\big[u^N\big] = \frac{1-\overline{t}(\theta)}{1-\overline{t}(\theta u)} \tag{64}$$

is a Gibbs variable with mean $\mathbf{E}^2_\theta[N] = \theta\overline{t}(\theta)'/(1-\overline{t}(\theta))$.

4.3. *Conditional distributions in the thermodynamic limit $n \to \infty$: the super-critical region $\theta > \theta_0$*

As was noted before, the random variable N becomes degenerate, in the sense that $N = \infty$ with probability $\mathbf{P}_\theta$ one, as soon as $\theta > \theta_0$. The reasonable approach for the understanding of aggregates in this region of the control parameter is, thus, to extract and evaluate the asymptotic shape of the θ – Taylor coefficients of complicated generating functions, such as $Z(\theta,\gamma,\gamma_m)$ (with $Z = Z^1$ or $Z = Z^2$), where $Z^1(\theta,\gamma,\gamma_m) = \mathrm{e}^{\gamma\overline{t}(\theta,\gamma_m)}$ and $Z^2(\theta,\gamma,\gamma_m) = 1/(1-\gamma\overline{t}(\theta,\gamma_m))$ from (23) and (32) with $\overline{t}(\theta,\gamma_m) \stackrel{\text{def}}{=} \overline{t}(\theta) + (\gamma_m - 1)c_m\theta^m/m!$ from (24).

The number of groups, $P(n)$, and the number of size-m groups, $N_m(n)$, variables are now to be understood conditionally to $N = n$, with n becoming large. This amounts to search for an asymptotic equivalent to $\Phi_n(\gamma,\gamma_m)$ in the thermodynamic limit $n \to \infty$ upon normalizing. The techniques to be employed to perform this program derive from singularity analysis.

We first recall a partial result of [11] before discussing the way it particularizes to our situation.

Singularity analysis result

Let $Z(\theta)$ be any analytic function in the indented domain defined by

$$D = \big\{\theta\colon\ |\theta| \leqslant \theta_1,\ \big|\operatorname{Arg}(\theta-\theta_0)\big| > \pi/2 - \eta\big\},$$

where $\theta_0,\ \theta_1 > \theta_0$, and η are positive real numbers. Assume that, with $\sigma(x) = x^\alpha \times \log^\beta x,\ \alpha$ and β any real number (the singular exponents), we have

$$Z(\theta) \sim K_1 + K_2\sigma\left(\frac{1}{1-\theta/\theta_0}\right) \quad \text{as } \theta \to \theta_0 \text{ in } D, \tag{65}$$

for some real constants K_1 and K_2.

Then:

If $\alpha \notin \{0,-1,-2,\ldots\}$, the Taylor coefficients of $Z(\theta)$ satisfy

$$[\theta^n]Z(\theta) \sim K_1 + K_2\theta_0^{-n}\frac{\sigma(n)}{n}\frac{1}{\Gamma(\alpha)} \quad \text{as } n \to \infty, \tag{66}$$

where $\Gamma(\alpha)$ is the Euler function. $Z(\theta)$ presents an algebraic-logarithmic singularity at $\theta = \theta_0$.

If $\alpha \in \{0, -1, -2, \ldots\}$, the singularity $\theta = \theta_0$ is purely logarithmic and

$$[\theta^n]Z(\theta) \sim K_1 + K_2 \beta \theta_0^{-n} \frac{\sigma(n)}{n \log n} \left(\frac{1}{\Gamma}\right)'(\alpha) \quad \text{as } n \to \infty, \tag{67}$$

involving the derivative of the inverse of the Euler function at α.

Thus, for algebraic-logarithmic singularities, the asymptotic of the Taylor coefficients can be read from the singular behavior of the partition function under study.

Remark 2. If $Z(\theta)$ is now replaced by $Z(\theta, \gamma, \gamma_m)$ for $Z = Z^1$ or $Z = Z^2$, where (γ, γ_m) are considered as perturbation parameters, the above theorem applies with both the singularity location θ_0 possibly replaced by $\theta(\gamma, \gamma_m)$ and the singular exponents α (and β) possibly replaced by $\alpha(\gamma, \gamma_m)$ (and $\beta(\gamma, \gamma_m)$). It shall of course prove useful to understand how these parameters modify the singular behavior of $Z(\theta) = Z(\theta, 1, 1)$ in the two different models. This is the object of section 5.

5. Asymptotic of the variables $N_m(n)$ and $P(n)$ in the thermodynamic limit

In order to answer the question, one first has to understand the singularity of the generating functions for trees $\overline{t}(\theta)$ and $\overline{t}(\theta, \gamma_m)$, and afterwards, by composition of singularity, the ones of $Z^1(\theta, \gamma, \gamma_m) \stackrel{\text{def}}{=} \mathrm{e}^{\gamma \overline{t}(\theta, \gamma_m)}$ and $Z^2(\theta, \gamma, \gamma_m) \stackrel{\text{def}}{=} 1/(1 - \gamma \overline{t}(\theta, \gamma_m))$ will follow. The main source of singularity in our problem mainly stems from the generating functions for trees $\overline{t}(\theta)$ and $\overline{t}(\theta, \gamma_m) \stackrel{\text{def}}{=} \overline{t}(\theta) + (\gamma_m - 1) c_m \theta^m / m!$. Two types of singularity are to be distinguished here:

Weakly connected aggregates

$\overline{t}(\theta)$ diverges at $\theta \stackrel{\text{def}}{=} \theta_1$. This will happen if the branch generating function g is an *entire* function: we shall discuss two examples, namely g polynomial (of maximal degree $d \geqslant 2$) and g dominated by an exponential, that is, $g \sim \theta^M \exp(-\theta)$, as $\theta \to \infty$, for some integral number $M \geqslant 1$. The singularity type of function $\overline{t}(\theta)$ will be shown to be algebraic with parameter $\alpha = 1/(d-1) > 0$ in the first case and purely logarithmic in the second case. In both cases, however, $\overline{t}(\theta)$ *diverges* at $\theta \stackrel{\text{def}}{=} \theta_1$.

In such aggregates, the sequence $(g_k)_{k \geqslant 1}$ grows "slowly" (i.e., not faster than k^M, for some integer M in our examples), which justifies the entry weakly connected aggregates from the combinatorial interpretation of these coefficients of section 3. Connection is weak, so that nucleation dominates: one, therefore, expects the number of groups to be large. This will be shown to occur in the sense that

$$\frac{1}{\Lambda_n} P(n) \to \tau'(0) > 0$$

with $\mathbf{P}^1$ and $\mathbf{P}^2$ probability one, from the law of large numbers. We shall indicate how to determine the divergence scales Λ_n ($\Lambda_n \to \infty$ as $n \to \infty$) in the different cases discussed above: Λ_n grows slower than n in model 1, so that no finite-size groups

are present asymptotically; on the contrary, Λ_n grows like n in model 2, so that a non-trivial finite-size group structure emerges. Also, large deviation results will be shown to hold generically, and we shall compute some large deviation rate functions.

Strongly connected aggregates

$\overline{t}(\theta)$ is defined and *finite* at $\theta = \theta_1$. This will happen if g itself has a singularity, say $t_0 \geqslant 1$, at *finite distance of the origin*. In this case the type of singularity of $\overline{t}(\theta)$ will always be algebraic with parameter $\alpha = -1/2 < 0$ (branch point) [18], as a result of the implicit function theorem. As a generic example, we shall treat the case $g(\theta) = 1/(1 - \overline{P}(\theta))$, where $\overline{P}(\theta)$ is a degree-$L \geqslant 2$ polynomial with non-negative Taylor coefficients, such that $\overline{P}(0) = \overline{P}'(0) = 0$ and $\overline{P}(1) < 1$. More precisely,

$$\overline{P}(\theta) = \sum_{l=2}^{L} \frac{\varepsilon_l}{l!}\theta^l \quad \text{with } \varepsilon_l \in \{0, 1\}.$$

In this case, the singularity of g is located at $t_0 \geqslant 1$, as a result of the Perron–Frobenius theorem for primitive matrices.

In such aggregates, the sequence $(g_k)_{k\geqslant 1}$ grows much faster (i.e., $g_k \sim k! t_0^{-k}$ as $k \to \infty$), which justifies the entry strongly connected aggregates from the combinatorial interpretation of these coefficients of section 3. Connection is strong, so that nucleation no longer dominates: one, therefore, expects the number of groups to be "smaller". Actually, it will be shown to occur in the sense that for model 1,

$$P(n) \to P(\infty) \quad \text{(in law)}$$

as $n \to \infty$, for some limit distribution $P(\infty)$, that will be shown to be Poisson-like. A similar behavior is observed for $N_m(n)$.

However, concerning the variable $P(n)$, in the context of model 2, it still continues to diverge in the sense that

$$\frac{1}{\Lambda_n} P(n) \to \tau'(0) > 0$$

with $\mathbf{P}^2$ probability one, for some divergent with n series Λ_n. In order that a limiting distribution $P(\infty)$ exists, such that

$$P(n) \underset{n\to\infty}{\longrightarrow} P(\infty)$$

holds in law, the condition that $t_0 \geqslant 1$ has to be violated and replaced by $t_0 < 1$, which means that the sequence $(g_k)_{k\geqslant 1}$ has to grow faster than $k!$: we shall call these aggregates *very strongly connected.*

5.1. Singularity analysis of the generating functions for trees

Let us first start by the θ-singularity analysis of the series $\overline{t}(\theta)$ and $\overline{t}(\theta, \gamma_m)$.

5.1.1. Weakly connected aggregates

(a) *The case g polynomial* [2]: Suppose $g(\theta) = 1 + \sum_{k=1}^{d} (g_k/k!)\theta^k,\ d \geqslant 2$. Rewrite (9) as

$$\theta = \int_0^{\overline{t}(\theta)} \frac{\mathrm{d}\theta'}{g(\theta')}. \tag{68}$$

Clearly, $\overline{t}(\theta)$ explodes (diverges) at finite distance θ_1 of the origin. Thus, the singularity of $\overline{t}(\theta)$ is located at

$$\theta_1 = \int_0^{\infty} \frac{\mathrm{d}\theta'}{g(\theta')}. \tag{69}$$

Hence,

$$\theta_1 - \theta = \int_{\overline{t}(\theta)}^{\infty} \frac{\mathrm{d}\theta'}{g(\theta')} \underset{\theta' \to \infty}{\sim} \frac{1}{g_d(d-1)} \overline{t}(\theta)^{1-d}$$

and

$$\overline{t}(\theta) \sim K_2 \left(\frac{1}{1 - \theta/\theta_1} \right)^{1/(d-1)} \quad \text{as } \theta \to \theta_1 \tag{70}$$

with $K_2 = (g_d \theta_1 (d-1))^{1/(1-d)}$.

The singularity of $\overline{t}(\theta)$ is purely algebraic of parameter $\alpha = 1/(d-1) > 0$. It follows from (66) that

$$[\theta^j]\overline{t}(\theta) \sim K_2 \theta_1^{-j} j^{1/(d-1)-1} \frac{1}{\Gamma(1/(d-1))} \quad \text{as } j \to \infty. \tag{71}$$

(b) *The case $g \underset{\theta \to \infty}{\sim} \theta^M \mathrm{e}^{\theta}$, for some integer $M \geqslant 1$*: This happens when $g(\theta) \overset{\text{def}}{=} g_M(\theta)$, recursively defined by $g_{m+1}(\theta) = 1 + \theta g'_m(\theta),\ m = 0, \ldots, M-1$, with $g_0(\theta) = \mathrm{e}^{\theta}$. In this case, $g(\theta) = 1 + \sum_{k \geqslant 1} (k^M/k!)\theta^k$, and $g_k = k^M,\ k \geqslant 1$, has a polynomial growth behavior.

Performing the same analysis as above,

$$\theta_1 - \theta = \int_{\overline{t}(\theta)}^{\infty} \frac{\mathrm{d}\theta'}{g(\theta')} \underset{\theta' \to \infty}{\sim} \overline{t}(\theta)^M \mathrm{e}^{-\overline{t}(\theta)},$$

so that

$$\overline{t}(\theta) \sim -\log(1 - \theta/\theta_1) \quad \text{as } \theta \to \theta_1 \tag{72}$$

with $\theta_1 \geqslant 1/\mathrm{e}$ again given by (69). It follows from (67)

$$[\theta^j]\overline{t}(\theta) \sim \theta_1^{-j} j^{-1} \quad \text{as } j \to \infty. \tag{73}$$

Remark 3. The generating function $\overline{t}(\theta,\gamma_m)=\overline{t}(\theta)+(\gamma_m-1)c_m\theta^m/m!$ is obtained by superimposing a degree-m monomial to function $\overline{t}(\theta)$. Thus, in the case just discussed where $\overline{t}(\theta)$ diverges at critical value $\theta=\theta_1$ the singular behavior of $\overline{t}(\theta,\gamma_m)$ matches the one of $\overline{t}(\theta)$: there is *no* influence of parameter γ_m neither on the location θ_1 nor on the singular exponents α and β of the singular expansion of $\overline{t}(\theta)$.

5.1.2. Strongly connected aggregates

(a) *The case g rational:* Suppose $g(\theta)=1/(1-\overline{P}(\theta))$, where $\overline{P}(\theta)=\sum_{l=2}^{L}(\varepsilon_l/l!)\theta^l$ with $\varepsilon_l\in\{0,1\}$, eventually aperiodic. This polynomial is the characteristic polynomial of its companion matrix. This matrix has non-negative entries, all bounded by one; it can easily be shown to be irreducible (actually, primitive if $\overline{P}(\theta)$ is aperiodic), so that from the Perron–Frobenius theorem, $\overline{P}(\theta)=1$ has a unique (algebraically simple) real solution $\theta=t_0$ (the inverse of the spectral radius of the companion matrix) such that all other roots are (strictly) outside the circle of radius $\theta=t_0$. From its particular shape, $\overline{P}(1)<1$, so that $\infty>t_0>1$. Moreover,

$$g(\theta)\sim\frac{1}{t_0\overline{P}'(t_0)}\left(\frac{1}{1-\theta/t_0}\right)\quad\text{as }\theta\to t_0,$$

so that

$$g_k\sim\frac{1}{t_0\overline{P}'(t_0)}k!t_0^{-(k+1)}\quad\text{as }k\to\infty,$$

from (66). Therefore, $g_k\sim Kk^{k+1/2}(\mathrm{e}t_0)^{-k}$ as $k\to\infty$, for some constant K and from the Stirling's formula. These coefficients, therefore, grow much faster than in any weakly connected aggregates: aggregation is "stronger".

In this situation, $\overline{t}(\theta)$ is undefined beyond the value θ_1, at finite distance of the origin, although it remains finite there: $\overline{t}(\theta_1)<\infty$. Thus, the singularity of $\overline{t}(\theta)$ is located at

$$\theta_1=\int_0^{t_0}\frac{\mathrm{d}\theta'}{g(\theta')}>1/\mathrm{e}. \tag{74}$$

Hence,

$$\theta_1-\theta=\int_{\overline{t}(\theta)}^{t_0}\frac{\mathrm{d}\theta'}{g(\theta')}.$$

Inverting this singularity, we get the "branch-point" behavior [18]:

$$\overline{t}(\theta)\sim t_0-\left(\frac{2\theta_1}{(-1/g)'(t_0)}\right)^{1/2}\left(\frac{1}{1-\theta/\theta_1}\right)^{-1/2}\quad\text{as }\theta\to\theta_1. \tag{75}$$

The singularity of $\overline{t}(\theta)$ is algebraic of parameter $\alpha = -1/2 < 0$. It follows from (66) that

$$[\theta^j]\overline{t}(\theta) \sim t_0 - \left(\frac{2\theta_1}{(-1/g)'(t_0)}\right)^{1/2} \theta_1^{-j} j^{-3/2} \frac{1}{\Gamma(-1/2)} \quad \text{as } j \to \infty \tag{76}$$

and $c_j \stackrel{\text{def}}{=} j![\theta^j]\overline{t}(\theta) \sim K j^{j-1} (\mathrm{e}\theta_1)^{-j}$ as $j \to \infty$, from Stirling's formula.

Remark 4. The generating function $\overline{t}(\theta, \gamma_m) = \overline{t}(\theta) + (\gamma_m - 1) c_m \theta^m / m!$ is obtained by superimposing a degree-m monomial to function $\overline{t}(\theta)$. Thus, in the case just discussed where $\overline{t}(\theta)$ converges at critical value $\theta = \theta_1$, the singular behavior of $\overline{t}(\theta, \gamma_m)$ now sensibly differs from the one of $\overline{t}(\theta)$ in the following sense: there still is *no* influence of parameter γ_m neither on the location θ_1 nor on the singular exponents $\alpha = -1/2$ of the singular expansion of $\overline{t}(\theta)$ (75), but the *value* $\overline{t}(\theta_1) = t_0$ at the singularity $\theta = \theta_1$ is now "polluted" by γ_m and should be replaced by

$$\overline{t}(\theta_1, \gamma_m) \stackrel{\text{def}}{=} t_0(\gamma_m) = t_0 + (\gamma_m - 1) \frac{c_m \theta_1^m}{m!}. \tag{77}$$

5.2. Singularity analysis of the partition function $Z^1(\theta, \gamma, \gamma_m)$

We shall now study, in these various cases, the type of singularity one is in right to expect, concerning the generating functions $Z^1(\theta, \gamma, \gamma_m)$ and $Z^2(\theta, \gamma, \gamma_m)$, simply by composition of singularities. This will allow us to derive asymptotic information on both the number of groups $P(n)$ and number of size-m groups $N_m(n)$ information encapsulated within (22) and (25), conditionally to n, and in the limit $n \to \infty$.

5.2.1. Weakly connected aggregates

(a) *The case g polynomial:* Recall from (70) that

$$\overline{t}(\theta) \sim K_2 \left(\frac{1}{1 - \theta/\theta_1}\right)^{1/(d-1)} \quad \text{as } \theta \to \theta_1;$$

the singularity of $\overline{t}(\theta)$ is purely algebraic of parameter $\alpha = 1/(d-1) > 0$. Moreover, $\overline{t}(\theta, \gamma_m)$ has the same singular expansion, with no pollution of parameter γ_m. Therefore, $Z^1(\theta, \gamma, \gamma_m) \stackrel{\text{def}}{=} \mathrm{e}^{\gamma \overline{t}(\theta, \gamma_m)}$ presents an *essential* singularity at $\theta = \theta_1$, with no influence of γ_m. When $d = 2$, it follows, from a saddle point analysis [15] that

$$[\theta^n] Z^1(\theta, \gamma, \gamma_m) \underset{n \to \infty}{\simeq} \theta_1^{-n} \frac{\exp(K_2 \gamma / 2)}{2\sqrt{\pi} (K_2 \gamma)^{-1/4}} \exp\left(2\sqrt{K_2 \gamma n}\right). \tag{78}$$

Of course,

$$[\theta^n] Z^1(\theta) \underset{n \to \infty}{\simeq} \theta_1^{-n} \frac{\exp(K_2/2)}{2\sqrt{\pi} (K_2)^{-1/4}} \exp\left(2\sqrt{K_2 n}\right), \tag{79}$$

so that, after normalization,

$$\Phi_n^1(\gamma,\gamma_m) \stackrel{\text{def}}{=} \frac{[\theta^n]Z^1(\theta,\gamma,\gamma_m)}{[\theta^n]Z^1(\theta)}$$
$$\underset{n\to\infty}{\longrightarrow} \frac{\exp(K_2(\gamma-1)/2)}{\gamma^{-1/4}}\exp\bigl(2\sqrt{K_2 n}(\sqrt{\gamma}-1)\bigr), \tag{80}$$

which is independent of γ_m.

As a first conclusion, taking $\gamma=1$ in (80) yields $\lim_{n\to\infty}\Phi_n^1(1,\gamma_m)=1$. Thus, given $N=n$, for any *finite* group size m held fixed

$$N_m(n) \underset{n\to\infty}{\longrightarrow} 0 \tag{81}$$

with $\mathbf{P}^1$ probability one.

Concerning the number of groups variable, obtained while setting $\gamma_m=1$,

$$\Phi_n^1(\gamma,1)^{1/\Lambda_n} \underset{n\to\infty}{\longrightarrow} \rho^1(\gamma,1) \stackrel{\text{def}}{=} \rho^1(\gamma) \tag{82}$$

with

$$\rho^1(\gamma)=\exp\bigl(2\sqrt{K_2}(\sqrt{\gamma}-1)\bigr). \tag{83}$$

Here $\Lambda_n=\sqrt{n}$. The mean and variance of the number of clusters random variable diverge algebraically. (Actually, for any d, the divergence rate can be shown to be the power-law $\Lambda_n=n^{1/d}$.) As a result, the number of groups, given $N=n$, say $P(n)$, grows like $\Lambda_n=n^{1/d}$. The number of atoms per group, that is, the group-size variable $n/P(n)$, tends to infinity, which is consistent with the result stated in (81): there is *no* finite-size groups asymptotically!

(b) *The case* $g \underset{\theta'\to\infty}{\sim} \theta^M \mathrm{e}^{\theta}$, *for some integer* $M\geqslant 1$: Note from (72) that $\overline{t}(\theta)\sim -\log(1-\theta/\theta_1)$ as $\theta\to\theta_1$; the singularity of $\overline{t}(\theta)$ (and $\overline{t}(\theta,\gamma_m)$) is purely logarithmic. Therefore, $Z^1(\theta,\gamma,\gamma_m)\stackrel{\text{def}}{=}\mathrm{e}^{\gamma\overline{t}(\theta,\gamma_m)}$ presents an exp–log type singularity at $\theta=\theta_1$. It follows that

$$Z^1(\theta,\gamma,\gamma_m) \underset{\theta=\theta_1}{\sim} \left(\frac{1}{1-\theta/\theta_1}\right)^{\gamma},$$

independent of γ_m. The singularity of $Z^1(\theta,\gamma,\gamma_m)$ is algebraic, with parameter γ. It follows from the singularity analysis result (67) that

$$[\theta^n]Z^1(\theta,\gamma,\gamma_m) \underset{n\to\infty}{\sim} \theta_1^{-n}\frac{n^{\gamma}}{n}\frac{1}{\Gamma(\gamma)}. \tag{84}$$

Of course,

$$[\theta^n]Z^1(\theta) \underset{n\to\infty}{\sim} \theta_1^{-n}, \tag{85}$$

so that

$$\Phi_n^1(\gamma,\gamma_m) \stackrel{\text{def}}{=} \frac{[\theta^n]Z^1(\theta,\gamma,\gamma_m)}{[\theta^n]Z^1(\theta)} \underset{n\to\infty}{\sim} \frac{1}{\Gamma(\gamma)}\exp\bigl(-\log n(1-\gamma)\bigr), \tag{86}$$

independent of γ_m.

Concerning $N_m(n)$, similar conclusions as equation (81) hold.

Concerning the number of groups variable,

$$\Phi_n^1(\gamma,1)^{1/\Lambda_n} \underset{n\to\infty}{\longrightarrow} \rho^1(\gamma,1) \stackrel{\text{def}}{=} \rho^1(\gamma) \tag{87}$$

with

$$\rho^1(\gamma) = \exp\bigl(-(1-\gamma)\bigr). \tag{88}$$

This is the probability generating function of a standard Poisson variable.

Here $\Lambda_n = \log n$. The mean and variance of the number of clusters random variable diverge in a logarithmic way.

Thus, an *infinite cluster* regime is observed for weakly connected aggregates under $\mathbf{P}^1$: there are no finite-size groups and the number of groups diverges with the number n of atoms but slower than n (like $n^{1/d}$ or like $\log n$).

5.2.2. Strongly connected aggregates

(a) *The case g rational:* Note from (75) that

$$\overline{t}(\theta,\gamma_m) \sim t_0(\gamma_m) - \left(\frac{2\theta_1}{(-1/g)'(t_0)}\right)^{1/2}\left(\frac{1}{1-\theta/\theta_1}\right)^{-1/2} \quad \text{as } \theta\to\theta_1,$$

with $t_0(\gamma_m) = t_0 + (\gamma_m - 1)c_m\theta_1^m/m!$; the singularity of $\overline{t}(\theta,\gamma_m)$ is purely algebraic of parameter $\alpha = -1/2 < 0$. Therefore, $Z^1(\theta,\gamma,\gamma_m) \stackrel{\text{def}}{=} \mathrm{e}^{\gamma\overline{t}(\theta,\gamma_m)}$ also presents an *algebraic* type singularity (with parameter $\alpha = -1/2$) at $\theta = \theta_1$.

Indeed,

$$Z^1(\theta,\gamma,\gamma_m) \underset{\theta=\theta_1}{\sim} \mathrm{e}^{\gamma t_0(\gamma_m)}\left(1-\gamma\left(\frac{2\theta_1}{(-1/g)'(t_0)}\right)^{1/2}\left(\frac{1}{1-\theta/\theta_1}\right)^{-1/2}\right).$$

It follows from the singularity analysis result (66) that

$$[\theta^n]Z^1(\theta,\gamma,\gamma_m) \underset{n\to\infty}{\sim} -\gamma\mathrm{e}^{\gamma t_0(\gamma_m)}\left(\frac{2\theta_1}{(-1/g)'(t_0)}\right)^{1/2}\frac{\theta_1^{-n}n^{-3/2}}{\Gamma(-1/2)}. \tag{89}$$

As a result,

$$\Phi_n^1(\gamma,\gamma_m) \stackrel{\text{def}}{=} \frac{[\theta^n]Z^1(\theta,\gamma,\gamma_m)}{[\theta^n]Z^1(\theta)} \underset{n\to\infty}{\sim} \gamma\exp\bigl(\gamma t_0(\gamma_m) - t_0\bigr). \tag{90}$$

Hence,

$$\Phi_n^1(\gamma,\gamma_m) \underset{n\to\infty}{\longrightarrow} \rho_\infty^1(\gamma,\gamma_m) \stackrel{\text{def}}{=} \Phi_\infty^1(\gamma,\gamma_m) \tag{91}$$

with

$$\mathbf{\Phi}^1_\infty(\gamma,\gamma_m)=\gamma\exp\left((\gamma-1)t_0+\gamma(\gamma_m-1)\frac{c_m\theta_1^m}{m!}\right) \tag{92}$$

from (77).

The random variables $P(n)$ and $N_m(n)$ converge in law this time.

Setting $\gamma_m=1$, $\mathbf{\Phi}^1_\infty(\gamma,1)=\gamma\exp(\gamma-1)t_0$ appears to be the probability generating function of a (shifted) Poisson variable for $P(n)$, with mean and variance $1+t_0$. If $\gamma=1$,

$$\mathbf{\Phi}^1_\infty(1,\gamma_m)=\exp\left((\gamma_m-1)\frac{c_m\theta_1^m}{m!}\right)$$

yields a Poisson probability generating function for $N_m(n)$, with intensity $c_m\theta_1^m/m!$ (compare with (54)).

5.3. *Singularity analysis of the partition function* $Z^2(\theta,\gamma,\gamma_m)$

We shall now show that the conclusions to be drawn are significantly different under probability $\mathbf{P}^2$.

5.3.1. *Weakly connected aggregates*

(a) *The case g polynomial:* Note from (70) that

$$\overline{t}(\theta)\sim K_2\left(\frac{1}{1-\theta/\theta_1}\right)^{1/(d-1)}\quad\text{as }\theta\to\theta_1;$$

the singularity of $\overline{t}(\theta)$ is purely algebraic of parameter $\alpha=1/(d-1)>0$. The function $\overline{t}(\theta)$ diverges as $\theta\to\theta_1$.

The singularity of $Z^2(\theta,\gamma,\gamma_m)\stackrel{\text{def}}{=}1/(1-\gamma\overline{t}(\theta,\gamma_m))$ differs now sensibly from the one of $\overline{t}(\theta)$ (and thus $\overline{t}(\theta,\gamma_m)$) in the sense that both location and singular exponent are corrupted by the variables (γ,γ_m) this time.

More precisely, $Z^2(\theta,\gamma,\gamma_m)\stackrel{\text{def}}{=}1/(1-\gamma\overline{t}(\theta,\gamma_m))$ presents an *algebraic* singularity (of parameter $\alpha=1$) at $\theta\stackrel{\text{def}}{=}\theta(\gamma,\gamma_m)<\theta_1$: *the singularity has been shifted to the left.*

Here $\theta(\gamma,\gamma_m)$ is well-defined implicitly by

$$\overline{t}\big(\theta(\gamma,\gamma_m),\gamma_m\big)\stackrel{\text{def}}{=}1/\gamma. \tag{93}$$

Near this new singularity $\theta=\theta(\gamma,\gamma_m)$,

$$Z^2(\theta,\gamma,\gamma_m)\sim\frac{1}{\gamma\theta(\gamma,\gamma_m)\overline{t}'(\theta(\gamma,\gamma_m),\gamma_m)}\frac{1}{(1-\theta/\theta(\gamma,\gamma_m))}; \tag{94}$$

hence, with $\overline{t}'(\theta(\gamma,\gamma_m),\gamma_m)$ standing for the partial derivative w.r. to θ of $\overline{t}(\theta,\gamma_m)$ at $\theta=\theta(\gamma,\gamma_m)$,

$$[\theta^n]Z^2(\theta,\gamma,\gamma_m)\underset{n\to\infty}{\sim}\frac{1}{\gamma\theta(\gamma,\gamma_m)\overline{t}'(\theta(\gamma,\gamma_m),\gamma_m)}\theta(\gamma,\gamma_m)^{-n}, \tag{95}$$

so that, upon normalizing,

$$\begin{aligned}\Phi_n^2(\gamma,\gamma_m) &\overset{\text{def}}{=} \frac{[\theta^n]Z^2(\theta,\gamma,\gamma_m)}{[\theta^n]Z^2(\theta)}\\ &\underset{n\to\infty}{\sim}\frac{\theta(1,1)\overline{t}'(\theta(1,1),1)}{\gamma\theta(\gamma,\gamma_m)\overline{t}'(\theta(\gamma,\gamma_m),\gamma_m)}\left(\frac{\theta(1,1)}{\theta(\gamma,\gamma_m)}\right)^n.\end{aligned} \tag{96}$$

Stated differently,

$$\Phi_n^2(\gamma,\gamma_m)^{1/\Lambda_n}\underset{n\to\infty}{\longrightarrow}\rho^2(\gamma,\gamma_m) \tag{97}$$

with

$$\rho^2(\gamma,\gamma_m)\overset{\text{def}}{=}\frac{\theta(1,1)}{\theta(\gamma,\gamma_m)}. \tag{98}$$

Here $\Lambda_n=n$. The mean and variance of the number of clusters random variables diverge linearly with n.

Setting $\gamma_m=1$, it should be observed, due to (9) and (93), that $\theta(\gamma)\overset{\text{def}}{=}\theta(\gamma,1)$ is defined *explicitly* from the branch generating function g by

$$\theta(\gamma)=\int_0^{1/\gamma}\frac{\mathrm{d}\theta'}{g(\theta')}. \tag{99}$$

Function $\rho^2(\gamma,1)$ can, therefore, be quite intricate, in general not available in explicit form, although this may happen as the following example show.

Example 3. If $g(\theta)=1+\theta^2$, $\overline{t}(\theta)=\tan\theta$, $Z^2(\theta,\gamma,1)=1/(1-\gamma\tan\theta)$ and $\theta(\gamma,1)=\arctan(1/\gamma)$, function $\rho^2(\gamma,1)$ is explicitly given by $\rho^2(\gamma,1)=\pi/(4\arctan(1/\gamma))$.

(b) *The case* $g\underset{\theta\to\infty}{\sim}\theta^M\mathrm{e}^\theta$, *for some integer* $M\geqslant 1$: Recall from (72) that $\overline{t}(\theta)\sim-\log(1-\theta/\theta_1)$ as $\theta\to\theta_1$; $\overline{t}(\theta)$ also diverges at θ_1 in a logarithmic way this time. The above results (93)–(99) are still valid, the only shape of $\rho^2(\gamma,\gamma_m)$ changing in a drastic way.

Example 4. If $g(\theta)=\exp\theta$, $\overline{t}(\theta)=-\log(1-\theta)$, $Z^2(\theta,\gamma,1)=1/(1+\gamma\log(1-\theta))$ and $\theta(\gamma,1)=1-\exp(-1/\gamma)$, function $\rho^2(\gamma,1)$ is explicitly given by

$$\rho^2(\gamma,1)=\big(1-\exp(-1)\big)\big/\big(1-\exp(-1/\gamma)\big).$$

5.3.2. Strongly connected aggregates

(a) *The case g rational:* Recall from (75) that

$$\overline{t}(\theta,\gamma_m) \sim t_0(\gamma_m) - \left(\frac{2\theta_1}{(-1/g)'(t_0)}\right)^{1/2}\left(\frac{1}{1-\theta/\theta_1}\right)^{-1/2} \quad \text{as } \theta \to \theta_1;$$

the singularity of $\overline{t}(\theta,\gamma_m)$ is purely algebraic of parameter $\alpha = -1/2 < 0$. In the singularity analysis of function $Z^2(\theta,\gamma,\gamma_m) \stackrel{\text{def}}{=} 1/(1-\gamma\overline{t}(\theta,\gamma_m))$ the value of $t_0(\gamma_m)$ appears critical, in particular its relative position with respect to one. Since $t_0 > 1$ for strongly connected aggregates, the conclusions are quite similar to the ones of weakly connected aggregates. Indeed, $Z^2(\theta,\gamma,\gamma_m)$ still presents an *algebraic* singularity (of parameter $\alpha = 1$) at $\theta = \theta(\gamma,\gamma_m) < \theta_1$, which is defined in some open neighborhood of $\gamma = \gamma_m = 1$: the singularity is again to be shifted to the left. Function $\theta(\gamma,\gamma_m)$ is still defined by (93) but condition $t_0(\gamma_m) \geqslant 1/\gamma$ defines an admissible sub-domain of $[0,1]^2$ in the parameter space (γ,γ_m), *including* the point $(1,1)$. This results from the fact that $\overline{t}(\theta,\gamma_m)$ is finite at $\theta = \theta_1$ this time.

Thus, the range of (γ,γ_m) is restricted to this sub-domain, which is the most notable difference with the previous situation.

The above results (93)–(99) are still valid, the only shape of $\rho^2(\gamma,\gamma_m)$ changing in a drastic way. In particular,

$$\Phi_n^2(\gamma,\gamma_m)^{1/\Lambda_n} \underset{n\to\infty}{\longrightarrow} \rho^2(\gamma,\gamma_m) \tag{100}$$

with

$$\rho^2(\gamma,\gamma_m) = \frac{\theta(1,1)}{\theta(\gamma,\gamma_m)} \tag{101}$$

with $\theta(\gamma,\gamma_m)$ still defined by (93), but for (γ,γ_m) now restricted to satisfy $t_0(\gamma_m) \geqslant 1/\gamma$.

Here $\Lambda_n = n$. The mean and variance of the number of clusters random variable continue to diverge linearly: the convergence in law phenomenon observed for strongly connected aggregates in section 5.2.2 no longer holds!

Example 5. If $g(\theta) = 1/(1-\theta^2/2)$, one has

$$\overline{t}(\theta) = \left(-3\theta + \left(9\theta^2-8\right)^{1/2}\right)^{1/3} + \left(-3\theta - \left(9\theta^2-8\right)^{1/2}\right)^{1/3},$$

with $t_0 = \sqrt{2}$ and $\theta_1 = (2/3)\sqrt{2}$. Hence, $\theta(\gamma,1) = 1/\gamma - 1/(6\gamma^3)$. Function $\rho^2(\gamma,1)$ is the function explicitly given by $\rho^2(\gamma,1) = 5\gamma^3/(6\gamma^2-1)$, which should be considered on the sub-domain $\gamma \in [1/\sqrt{2},1]$ only.

5.3.3. Very strongly connected aggregates: $t_0 < 1$

Suppose the sequence $(g_k)_{k\geqslant 1}$ grows faster than $k!$. In this situation, $t_0 < 1$. If $t_0 < 1$, there is no solution $\theta(\gamma,\gamma_m)$ to (93), for (γ,γ_m) defined in an open neighborhood of the point $(1,1)$.

The singularity of function $Z^2(\theta,\gamma,\gamma_m) \stackrel{\text{def}}{=} 1/(1-\gamma\overline{t}(\theta,\gamma_m))$ remains, therefore, located at the value $\theta=\theta_1$. Indeed,

$$Z^2(\theta,\gamma,\gamma_m) \underset{\theta=\theta_1}{\sim} \frac{1}{1-\gamma t_0(\gamma_m)}\left(1-\frac{\gamma(2\theta_1/(-1/g)'(t_0))^{1/2}}{1-\gamma t_0(\gamma_m)}\left(\frac{1}{1-\theta/\theta_1}\right)^{-1/2}\right).$$

The singularity of $Z^2(\theta,\gamma,\gamma_m)$ is algebraic, with parameter $\alpha=-1/2$. It follows from the singularity analysis result (66) that

$$[\theta^n]Z^2(\theta,\gamma,\gamma_m) \underset{n\to\infty}{\sim} -\frac{\gamma}{(1-\gamma t_0(\gamma_m))^2}\left(\frac{2\theta_1}{(-1/g)'(t_0)}\right)^{1/2}\frac{\theta_1^{-n}n^{-3/2}}{\Gamma(-1/2)}. \tag{102}$$

As a result, upon normalizing,

$$\Phi_n^2(\gamma,\gamma_m) \stackrel{\text{def}}{=} \frac{[\theta^n]Z^2(\theta,\gamma,\gamma_m)}{[\theta^n]Z^2(\theta)} \underset{n\to\infty}{\sim} \gamma\left(\frac{1-t_0}{1-\gamma t_0(\gamma_m)}\right)^2. \tag{103}$$

Hence

$$\Phi_n^2(\gamma) \underset{n\to\infty}{\longrightarrow} \rho_\infty^2(\gamma,\gamma_m) \stackrel{\text{def}}{=} \Phi_\infty^2(\gamma,\gamma_m) \tag{104}$$

with

$$\rho_\infty^2(\gamma,\gamma_m) = \gamma\left(\frac{1-t_0}{1-\gamma(t_0+(\gamma_m-1)c_m\theta_1^m/m!)}\right)^2. \tag{105}$$

As $\gamma_m=1$, this is the probability generating function of a (shifted) squared geometric variable, with mean $1+2t_0/(1-t_0)^2$. The number of clusters random variable, $P(n)$, converges in law this time. As $\gamma=1$, this is the probability generating function of a squared geometric variable, $N_m(\infty)$, with mean $2c_m\theta_1^m/(m!(1-t_0))$ (compare with (62)).

5.4. Summary

Let us put altogether the results of this section, concerning the variable $P(n)$ and $N_m(n)$, given $N=n$, in the thermodynamic limit $n\to\infty$.

Model 1: discernible atoms, indiscernible clusters

Weakly connected aggregates (g entire): In any case, given $N=n$, $N_m(n) \underset{n\to\infty}{\longrightarrow} 0$ with $\mathbf{P}^1$ probability one, for any *finite* group-size m held fixed.

- g degree-d polynomial ($g_k=0,\ k>d\geqslant 2$), $P(n)$ grows like $n^{1/d}$.
- $g\sim\theta^M\exp\theta$ ($g_k\sim k^M$), $P(n)$ grows like $\log n$.

Strongly connected aggregates (g has a singularity $t_0\geqslant 1$ at finite distance):

- $g(\theta)=1/(1-\overline{P}(\theta))$ ($g_k\sim k^k t_0^{-k}$), the random vector $(P(n),N_m(n))$ converges in law to a (Poisson) distribution given by (92).

Model 2: discernible atoms, discernible clusters

Weakly connected aggregates (g entire):

– g degree-d polynomial ($g_k = 0,\ k > d$), the random vector $(P(n), N_m(n))$ grows like n.

– $g \sim \theta^M \exp\theta$ ($g_k \sim k^M$), the random vector $(P(n), N_m(n))$ grow like n.

There is a non-trivial size-m group structure this time (see, e.g., (93), (97), (98)).

Strongly connected aggregates (g has a singularity $t_0 > 1$ at finite distance):

– $g(\theta) = 1/(1 - \overline{P}(\theta))$ ($g_k \sim k^k t_0^{-k}$), the random vector $(P(n), N_m(n))$ still grows like n. Conclusions are similar to the previous ones.

Very strongly connected aggregates (g has a singularity $t_0 < 1$ at finite distance):

– $g(\theta) = 1/(1 - \overline{P}(\theta))$ ($g_k \sim k^k t_0^{-k}$), the random vector $(P(n), N_m(n))$ converges in law to a (squared geometric) distribution given by (105).

In other words, the ratio $n/P(n)$, giving the "average" number of atoms per cluster can present a great variability, as $n \to \infty$, from infinite to finite, depending on the pattern formation process. Its behavior strongly influences the fragment size distribution.

5.5. Large deviation results

Concerning the number of groups' and the number of size-m groups' asymptotic, two regimes can thus be distinguished:

(1) *Convergence:* this happens if $\Phi_n(\gamma, \gamma_m) \underset{n\to\infty}{\longrightarrow} \rho_\infty(\gamma, \gamma_m) \overset{\text{def}}{=} \Phi_\infty(\gamma, \gamma_m)$ (see, e.g., (92), (105)). In these situations, the random vector $(P(n), N_m(n))$ converges in law to $(P(\infty), N_m(\infty))$.

(2) *Divergence:* this happens if $\Phi_n(\gamma, \gamma_m)^{1/\Lambda_n} \underset{n\to\infty}{\longrightarrow} \rho(\gamma, \gamma_m)$. Two sub-cases arise:

(a) If $\rho(\gamma, \gamma_m)$ is defined in a neighborhood of $(\gamma, \gamma_m) = (1, 1)$. In this situation, we have large deviation results, strongly reminiscent of multifractal theory, that we now give.

Let $\lambda \overset{\text{def}}{=} -\log\gamma,\ \lambda_m \overset{\text{def}}{=} -\log\gamma_m,\ \boldsymbol{\lambda} \overset{\text{def}}{=} (\lambda, \lambda_m)$.

Next, define

$$\varsigma(\boldsymbol{\lambda}) \overset{\text{def}}{=} \varsigma(\lambda, \lambda_m) \overset{\text{def}}{=} \rho\left(\mathrm{e}^{-\lambda}, \mathrm{e}^{-\lambda_m}\right) \tag{106}$$

and the "free energy" function

$$\tau(\boldsymbol{\lambda}) = \tau(\lambda, \lambda_m) \overset{\text{def}}{=} -\log\varsigma(\lambda, \lambda_m). \tag{107}$$

This function is well-known to be concave on its convex definition domain, $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}$, from the Hölder inequality. Domain $\boldsymbol{\Lambda}$ can be derived from the definition domain of (γ, γ_m) and contains the origin $\boldsymbol{\lambda} = \mathbf{0} = (0, 0)$.

Define then its concave Legendre transform

$$s(\boldsymbol{\alpha}) \stackrel{\text{def}}{=} s(\alpha, \alpha_m) \stackrel{\text{def}}{=} \inf_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}} \big(\boldsymbol{\alpha}\boldsymbol{\lambda} - \tau(\boldsymbol{\lambda})\big). \tag{108}$$

Clearly, $s(\boldsymbol{\alpha}) = \boldsymbol{\alpha}\nabla s(\boldsymbol{\alpha}) - \tau(\nabla s(\boldsymbol{\alpha}))$ with $\nabla\tau(\nabla s(\boldsymbol{\alpha})) = \boldsymbol{\alpha}$ (here ∇ is the gradient symbol). Moreover, $s(\boldsymbol{\alpha}) \leqslant 0$, as $\boldsymbol{\alpha} \in \boldsymbol{\Xi}$, the induced definition domain on $\boldsymbol{\alpha}$. Define $\boldsymbol{\alpha}_0$ by $\nabla s(\boldsymbol{\alpha}_0) = 0$. Function $s(\boldsymbol{\alpha})$ attains its maximum, zero, at $\boldsymbol{\alpha}_0$.

In all divergence situations described above, one can show the following local limit theorem for large deviation result [8,15]:

$$\mathbf{P}\left(\frac{1}{\Lambda_n}\mathbf{X}(n) \to \boldsymbol{\alpha}\right)^{1/\Lambda_n} \underset{n\to\infty}{\longrightarrow} \exp s(\boldsymbol{\alpha}) \tag{109}$$

with $s(\boldsymbol{\alpha})$ the large deviation rate function and $\mathbf{X}(n)$ the random vector $\mathbf{X}(n) \stackrel{\text{def}}{=} (P(n), N_m(n))$.

In particular, we have

$$\frac{1}{\Lambda_n}\mathbf{X}(n) \to \boldsymbol{\alpha}_0 \stackrel{\text{def}}{=} \nabla\tau(\mathbf{0}) \tag{110}$$

with $\mathbf{P}$ probability one.

Moreover, if $\mathbf{E}[\mathbf{X}(n)] = \Lambda_n\nabla\tau(\mathbf{0})$ and $[\mathbf{X}(n)] = \Lambda_n\Delta\tau(\mathbf{0})$ denote, respectively, the mean and variance–covariance matrix of $\mathbf{X}(n)$, we have the central limit theorem

$$\mathbf{P}\Big(\big[\mathbf{X}(n)\big]^{-1/2}\big(\mathbf{X}(n) - \mathbf{E}\big[\mathbf{X}(n)\big]\big) < \mathbf{x}\Big) \underset{n\to\infty}{\longrightarrow} \operatorname{erf}\,\mathbf{x} \tag{111}$$

for $\mathbf{x} = \mathrm{O}(\mathbf{1})$.

(b) If $\rho(\gamma, \gamma_m)$ is independent of γ_m (section 5.2.1). The above analysis only holds for the one-dimensional variable $P(n)$. Let us illustrate these points in one-dimensional situations.

Example 6. In section 5.2.1, we had $\rho^1(\gamma) = \exp(\gamma - 1)$. Therefore, $\tau(\lambda) = 1 - \exp(-\lambda)$ for $\lambda \in \mathbf{R}$, and $s(\alpha) = \alpha - 1 - \alpha\log\alpha \leqslant 0$ for any $\alpha \in \mathbf{R}^+$. Here $s(\alpha_0) = 0$ at $\alpha_0 = 1$.

In the example of section 5.3.2, we had $\rho^2(\gamma, 1) = \rho^2(\gamma) = 5\gamma^3/(6\gamma^2 - 1)$. Therefore,

$$\tau(\lambda) \stackrel{\text{def}}{=} -\log\rho^2\big(\mathrm{e}^{-\lambda}\big) = -\log\left(\frac{5\mathrm{e}^{-3\lambda}}{6\mathrm{e}^{-2\lambda} - 1}\right)$$

for $\lambda \leqslant \lambda_0 = (1/2)\log 2$ and its Legendre transform $s(\alpha)$ (which can be explicitly computed) is *only* defined in the range $\alpha \in [0, 1]$. Moreover, $s(\alpha_0) = 0$ at $\alpha_0 = 3/5$. Function $s(\alpha)$ tends to $-\infty$ as $\alpha \to 1^-$ and $s(0) = -\lambda_0$ with a positive finite slope there.

6. Concluding remarks

This paper presents a statistical physics' approach to the modeling of nucleation–aggregation phenomena of atoms. This work was motivated by a need to understand the fundamentals involved in clusters' formation processes. In this context, real experimental data have been related to one of the models discussed here (model (a) of section 2) by Cohen [5–7], in an attempt to understand processes such as coagulation of fine particles or coalescence of droplets, by considering the most probable group size distributions that have not been introduced here (see, also, [4,14]). These vital informations can be reached while maximizing a Boltzmann type entropy which turn out to be strongly dependent on the average number of atoms per cluster ratio $n/P(n)$, whose variability has been shown to be large.

Two different models have been first developed within this statistical framework, showing that small causes can produce large effects. These models were designed to represent simple aggregation by single particle adjunction (monomer addition).

It has also been shown, however, how to include in aggregation models the possibility for an atom to connect simultaneously to more than one group at the same time. In this situation, indeed, the number of groups in the transition $n \to n+1$ may decrease because clusters themselves can coalesce.

These informations are nicely represented by two partition functions of unordered (and ordered) forests of increasing trees.

In order now to understand these aggregation phenomena, we first derived the fragment size distributions (that is, the number P of fragments, or clusters, and the number N_m of size-m fragments with m constitutive atoms), as a function of the *control* parameter which is chosen here to be the average number of atoms $\langle N \rangle$. As $\langle N \rangle$ approaches infinity, we derived the study of these variables in the thermodynamic limit $n \to \infty$.

This formulation has shown that the nucleation and aggregation processes were in competition, and that macroscopic effects on the clusters number distributions could emerge. This intuition has received a rigorous positive answer, using singularity analysis techniques of the partition functions involved. This allowed for the distinction between two regimes: the ones of weakly and strongly connected aggregates. In the first regime, nucleation dominates aggregation which results in the divergence of the number of clusters variable, at rates depending on the connection politics of the additional atoms. In the second regime, on the contrary, this variable attains a macroscopically observable statistical equilibrium. As the number of atoms n tends to infinity, the number $P(n)$ of fragments, or clusters, and the number $N_m(n)$ of size-m fragments have thus been shown to present a great statistical variability under the two different models introduced.

References

[1] R. Azencott, *Grandes Déviations et Applications, Cours de Saint-Flour*, Lecture Notes in Math., Vol. 774 (Springer, Heidelberg, 1978).

[2] F. Bergeron, P. Flajolet and B. Salvy, Varieties of increasing trees, in: *CAAP's 92*, ed. J.C. Raoult, Lecture Notes in Comput. Sci., Vol. 581 (1992) pp. 24–48.

[3] R. Botet and Ploszajczak, Fragmentation–inactivation binary model – A new model of kinetic sequential fragmentation, in: *Proceedings of the Les Houches Workshop: Fragmentation Phenomena*, eds. D. Beysens, X. Campi and E. Pefferkorn (World Scientific, 1995) pp. 300–311.

[4] X. Campi and H. Krivine, Observables in fragmentation, in: *Proceedings of the Les Houches Workshop: Fragmentation Phenomena*, eds. D. Beysens, X. Campi and E. Pefferkorn (World Scientific, 1995) pp. 312–321.

[5] R.D. Cohen, Steady-state cluster size distribution in stirred suspensions, J. Chem. Soc. Faraday Trans. 86(12) (1990) 2133–2138.

[6] R.D. Cohen, Evolution of the cluster size distribution in stirred suspensions, J. Chem. Soc. Faraday Trans. 87(8) (1991) 1163–1168.

[7] R.D. Cohen, Development of the cluster-size distribution in flowing suspensions, AICHE J. 38(7) (1992) 1129–1134.

[8] R.S. Ellis, *Entropy, Large Deviations and Statistical Mechanics*, Grundlehren Math. Wiss. (Springer, New York, 1985).

[9] L. Comtet, *Advanced Combinatorics* (Reidel, Dordrecht, 1974).

[10] I.M. Elminyawi, S. Gangopadhyay and C.M. Sorensen, Numerical solutions to the Schmoluchowski aggregation–fragmentation equations, J. Colloid Interf. Sci. 144 (1991) 315.

[11] P. Flajolet and A. Odlyzko, Singularity analysis of generating functions, SIAM J. Discrete Math. 3(2) (1990) 216–240.

[12] I.P. Goulden and D.M. Jackson, *Combinatorial Enumeration* (Wiley, New York, 1983).

[13] P. Hammad, *Information, Systèmes et Distributions* (Editions Cujas, Paris, 1987).

[14] T. Huillet, Statistical physics' models of aggregation phenomena, J. Phys. A 30(6) (1997) 1849–1863.

[15] H.-K. Hwang, Théorèmes limites pour les structures combinatoires et les fonctions arithmétiques, Ph.D. thesis, Ecole Polytechnique (December 1994).

[16] V.F. Kolchin, *Random Mappings* (Optimization Software Inc., New York, 1986).

[17] R.D. Levine and M. Tribus, eds., *The Maximum Entropy Formalism* (MIT Press, Cambridge, MA, 1979).

[18] A. Meir and J.W. Moon, On the altitude of nodes in random trees, Canad. J. Math. 30 (1978) 997–1015.

[19] J. Riordan, *An Introduction to Combinatorial Analysis* (Wiley, New York, 1958).

[20] R.D. Rosenkrantz, in: *Papers on Probability, Statistics and Statistical Physics*, Vol. 158, ed. E.T. Jaynes (Reidel, Dordrecht, 1983).

[21] M.R. Schroeder, *Number Theory in Science and Communication* (Springer, Berlin, 1989).

[22] T. Sintes, R. Toral and A. Chakrabarti, Reversible aggregation in self-associating polymer systems, Phys. Rev. E 50 (1994) 2967–2976.

[23] F.W. Steutel and K. Van Harn, Discrete analogues of self-decomposability and stability, Ann. Probab. 7 (1979) 893–899.