Journal of Mathematical Chemistry 21 (1997) 83–101

A combinatorial approach to branched polymers' statistics

Thierry Huillet and Bernard Jeannet

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

Received 27 May 1996; revised 16 January 1997

At least ideally, a certain class of polymers presents itself as a collection (set) of connected components. Each of these components is a cycle of trees, that is branched polymers eventually rooted on a cycle. We derive (and study) an equilibrium statistical model that accounts for the main connectivity features of such structures, whose origin is to be found in combinatorial probability. Phase transition (gel–sol transition) is shown to occur when some internal control parameter crosses one (critical parameter). Various structural asymptotic results are shown to be available using singularity analysis.

1. Introduction

Branched polymers are giant macromolecules which are the repetition of a very large number of basic strings called monomers (typically several thousands). Their topology is either linear (bifunctional polymers which can simply react at their two ends) or strongly ramified (as a result of multifunctional monomers some end of which can be attached to several host monomers) [4,8]. Cycles (or loops) may be present, as a result of self-connection. Therefore, at least ideally, branched polymers present themselves as a collection (set) of connected components. Each of these connected components is a cycle of trees, that is branched polymers again (i.e., rooted labelled trees), eventually rooted on some vertex of a cycle.

The enumeration of such combinatorial structures is part of standard theory, using the notion of exponential generating function (EGF) which is recalled in section 2.1. These tools are adapted to our modeling purpose of polymers in section 2.2.

However, the point of view which is developed here is that all configurations are not allowed in the sense that there are combinatorial constraints to the modeling of such structures, depending on the particular "chemistry" of interest. This leads to the notion of restricted EGF for subconfigurations developed in section 2.3.

It is also reasonable to assume that all these configurations are not equally likely to occur during polymerization. For example, strongly ramified monomers, long cycles or multiple connected components are expected to be rare.

To this end, a randomizing procedure of these constraints is proposed, the average behavior of which is only taken into account. It also avoids entering into endless

"connectivity" details and considerably simplifies the problem. This leads to the notion of average generating functions (AEGF) discussed in section 2.4, which seems to be new.

We therefore are left with the study of an equilibrium statistical model that accounts for the main connectivity features of such structures, whose origin is to be found in combinatorial probability. This model is a three parameters construction that are $(0,1]$-valued probabilities (p_1, p_2, p_3). Their physical signification is as follows:

– p_1 reflects the multifunctionality of the monomers in the sense that p_1^m is the probability that m monomers append to some other monomer's end, in the "soup", thereby forming a degree-m node in some tree.

– p_2 accounts for the cycle-length in the sense that p_2^m is the probability that a cycle constituted with m monomers occurs.

– p_3 is a fragmentation parameter which measures the ability for a polymer to present several connected components. More precisely, p_3^m is the probability that a polymer is constituted of a collection of m connected subpolymers.

It should be noted that the probabilities of these events have been designed to tend to zero exponentially fast. More work is probably needed here in order to relate the analytic expressions of these probabilities to geometrical considerations such as self-obstructions.

Phase transition (gel–sol transition) is shown to occur when the internal control parameter p_2/p_1 crosses the unity from below (critical parameter equals one). Various structural asymptotic results are shown to be available using singularity analysis.

It should be underlined here that this "random-connection" model is very far removed from a "realistic" model of branched polymers. Indeed, "real" branched polymers are embedded in some physical subspace of $\mathbf{R}^3$ whereas abstract functional graphs' models proposed below are not embedded in any space, i.e., are "free". However, the randomizing procedure of functional graphs' models presented here, accounts for a rudimentary statistical form of "geometrical interactions" between monomers.

Another drawback of the model presented here is that it only allows a single cycle per connected component, which certainly is over-simplistic in practise. This restriction could be removed at the expense of more work on combinatorics.

As a result, this model can certainly be improved in various directions, our primary motivation in its development being its "solvability" and the fact that in this process, one may get insights into what may be happening in more realistic situations.

2. Counting the configurations

2.1. Exponential generating functions

Exponential generating functions are a very useful tool since they can account for quite deep combinatorial results in a straightforward way, simply by considering the composition of EGF's [3,9]. Let us first introduce these mathematical objects on which we shall focus all along this article.

Definition 1. We shall say that a function $Z(\theta)$ of the real variable θ is an exponential generating function (EGF), if

(i) $Z(0) = 1$,

(ii) it admits a convergent Taylor development, in some vicinity of the origin, that is: there is a sequence $(c_m)_{m\geqslant 1}$, and a non empty domain $D^+ = \{\theta \in \mathbf{R}^+ : Z(\theta) < \infty\}$, including zero, such that

$$Z(\theta) = 1 + \sum_{m\geqslant 1} \frac{c_m}{m!}\theta^m$$

is convergent for $\theta \in D^+$,

(iii) the sequence $(c_m)_{m\geqslant 1}$ has state-space $\mathbf{Z}^+$, the set of non-negative integers.

Definition 2. We let θ_0 denote the dominant (real positive) singularity of such EGF, which means that $Z(\theta)$ cannot be continued as an analytic function beyond θ_0.

This property is characteristic of EGF and more generally of Taylor power-series with non-negative Taylor coefficients.

The set of coefficients $(c_m)_{m\geqslant 1}$ is to be interpreted in the sequel as the configuration number with m labelled monomers; the EGF $Z(\theta)$ encodes these informations.

We shall also use the notation $\overline{Z}(\theta) \stackrel{def}{=} Z(\theta) - 1$, in the sequel.

Example 3. We list below a few examples of EGF that are remarkable and which shall be discussed in more detail in the sequel.

(1) $c_m = 1,\ m \geqslant 1$, and $Z(\theta) = \mathrm{e}^{\theta}$, $\theta_0 = \infty$ (set of isolated monomers);

(2) $c_m = (m-1)!,\ m \geqslant 1$, and $Z(\theta) = 1 - \log(1-\theta)$, $\theta_0 = 1$ (isolated cyclic polymer);

(3) $c_m = m!,\ m \geqslant 1$, and $Z(\theta) = 1/(1-\theta)$, $\theta_0 = 1$ (set of cyclic polymers);

(4) $c_m = m^{m-1},\ m \geqslant 1$, for which $\theta_0 = 1/\mathrm{e}$ (isolated acyclic branched polymer, tree);

(5) $c_m = (m+1)^{m-1},\ m \geqslant 1$, for which $\theta_0 = 1/\mathrm{e}$ (set of acyclic branched polymers, forest);

(6) $c_m = m^m,\ m \geqslant 1$, for which $\theta_0 = 1/\mathrm{e}$ (set of cyclic branched polymers);

(7) $c_m = 1$, for m even, 0 otherwise, and $Z(\theta) = \cosh\theta$, $\theta_0 = \infty$ (even);

(8) $c_m = 1$, for m odd, 0 otherwise, and $Z(\theta) = 1 + \sinh\theta$, $\theta_0 = \infty$ (odd);

(9) $c_m = 0$, for $m = 1$ and $m \geqslant 3, c_2 = 1$, and $Z(\theta) = 1 + \theta^2/2$, $\theta_0 = \infty$ (binary);

(10) $c_m = 0$, for $m \geqslant 2, c_1 = 1$, and $Z(\theta) = 1 + \theta$, $\theta_0 = \infty$.

2.2. EGF arising from branched polymers

Let us now identify some fundamental EGF arising in the statistical physics' modeling of branched polymers. Polymers are macromolecules that can be seen as a collection of basic units called monomers. Each monomer has two distinct ends. At

one end (the source end), we have a certain amount of host sites, each of which serving as a potential host for other monomers. At the other end (the sink end), we have a single connection site which allows each monomer to (eventually) connect to any host site of another monomer. Such generated structures for polymers can therefore be identified with functional graphs or directed graphs with nodes and directed edges. The output degree at each node, that is the number of edges pointing outwards a node, cannot exceed one [3,9,10].

The nodes of these functional graphs will be identified with each monomer and the arrow on each edge points towards some host monomer. A monomer with a free connection site is called a root. A monomer with a free source end is called a leaf. We shall call a d-monomer or d-node, a monomer whose host capacity at the source end is d. A cycle is a monomers' structure that loops, a tree, or branched polymer, is a connected acyclic graph, a cycle tree a set of rooted labelled trees arranged in a cycle, a forest a set of trees...

Let us now translate such structures in the language of EGF.

Definition 4. Let $\overline{T}(\theta) \stackrel{\text{def}}{=} T(\theta) - 1$ be the (unique) solution of the functional equation $\overline{T}(\theta) = \theta \mathrm{e}^{\overline{T}(\theta)}$. Then $T(\theta)$ is an EGF with dominant singularity at $\theta_0 = 1/\mathrm{e}$. We shall call it the tree-EGF because $c_m = m^{m-1}$, $m \geqslant 1$, counts the number of rooted trees with m labelled monomers (nodes) [5].

This result dates back to A. Cayley [2]. An unordered rooted labelled tree (or non plane tree) is thus recursively defined by appending a monomer to a set of similar subtrees (example 3 (4)). Note that for such trees, there is no order distinction between the subtrees dangling from the common root. Ordered trees, or plane trees, could be obtained in a similar way, when considering the modified functional equation $\overline{T}(\theta) = \theta/(1-\overline{T}(\theta))$, for which there $m!$ ways to arrange m subtrees, taking "chirality" into account.

Definition 5. Let $\overline{C}(\theta) \stackrel{\text{def}}{=} C(\theta) - 1 = -\log(1-\theta)$, with dominant singularity $\theta_0 = 1$. Then $c_m = (m-1)!$, $m \geqslant 1$, counts the number of cyclic permutations, i.e., the number of ways one can assemble m labelled monomers on a cycle. We shall call it the cycle-EGF (example 3 (2)).

Definition 6. We shall call $E(\theta) = \mathrm{e}^{\theta}$ the set-EGF (example 3 (1)).

Considering now the composition of such simple EGF's can account for quite deep combinatorial results in a straightforward way. To illustrate this idea, let us consider the interplay of the three EGF introduced above.

Considering a set of cycles amounts to work with the EGF $E(\overline{C}(\theta)) = 1/(1-\theta)$, whose Taylor coefficients are $c_m = m!$, $m \geqslant 1$, counting the number of ways that m labelled monomers can present themselves as a collection (set) of cycles, as required (the number of permutations of m objects, see, e.g., example 3 (3)).

Rooting trees on a single cycle amounts to consider the EGF $1+\overline{C}(\overline{T}(\theta))$, whose Taylor coefficients grow like $c_m \sim \sqrt{\pi m/2}\, m^{m-1}$, for large m. We shall call it the cycle-trees-EGF. A cycle of trees is therefore a labelled cycle with rooted labelled trees eventually connected to each vertex of the cycle.

Considering a set of cycle-trees requires to consider the mapping-EGF $E(\overline{C}(\overline{T}(\theta)))$ $= 1/(1-\overline{T}(\theta))$, whose Taylor coefficients are $c_m = m^m$, $m \geqslant 1$, which counts the number of mappings from a set of cardinal m onto itself (example 3 (6)).

Considering a set of trees (a forest) amounts to consider the forest-EGF $E(\overline{T}(\theta))$, omitting the cycle-EGF, whose Taylor coefficients are $c_m = (m+1)^{m-1} \sim \mathrm{e} m^{m-1}$, as $m \to \infty$ (example 3 (5)).

These are the three fundamental EGF we shall play with. In fact, subconfigurations will be our main concern here which requires to introduce constraints in the enumeration problems alluded to previously. There are many ways to consider such constraints.

2.3. Restricted EGF of subconfigurations

One way to proceed is to let

$$g_1(\theta) \stackrel{\text{def}}{=} 1 + \sum_{m \geqslant 1} \frac{\varepsilon_m^1}{m!}\theta^m,$$

with $\varepsilon_m^1 \in \{0,1\}$, $m \geqslant 1$, any deterministic $\{0,1\}$-valued infinite sequence. Such an EGF may then be called a subexponential [1].

Considering the solution to the functional equation $\overline{t}_1(\theta) = \theta g_1(\overline{t}_1(\theta))$, amounts to count trees of labelled monomers when the number of incoming branches m (monomers with degree m) is forbidden as a result of $\varepsilon_m^1 = 0$, allowed otherwise. For example, "linear" trees are readily obtained from $g_1(\theta) = 1+\theta$ (example 3 (10)), binary trees from $g_1(\theta) = 1+\theta^2/2!$, i.e., example 3 (9), even trees (that is with an even number of incoming branches) from $g_1(\theta) = \cosh\theta$ (example 3 (7)) and odd trees from $g_1(\theta) = 1+\sinh\theta$ (example 3 (8)). We shall call $g_1(\theta)$ the branch-EGF.

In the same way, cycle lengths may be forbidden in the subconfigurations under concern, which is equivalent to consider the restricted cycle-EGF

$$\overline{c}_2(\theta) \stackrel{\text{def}}{=} \sum_{m \geqslant 1} \frac{\varepsilon_m^2}{m}\theta^m,$$

with $\varepsilon_m^2 \in \{0,1\}$, $m \geqslant 1$, any other deterministic $\{0,1\}$-valued infinite sequence.

Moreover, one may wish to forbid configurations with a certain amount of connected components; this can be done by considering the restricted set-EGF

$$e_3(\theta) \stackrel{\text{def}}{=} \sum_{m \geqslant 1} \frac{\varepsilon_m^3}{m!}\theta^m,$$

with $\varepsilon_m^3 \in \{0,1\}$, $m \geqslant 1$, any third deterministic $\{0,1\}$-valued infinite sequence.

Combining these three restricted EGF again leads to an infinite number of combinatorial problems, as a result of particular choices of the three $\{0,1\}$-valued infinite sequences.

2.4. Randomizing the subconfigurations

There is therefore a great variety of problems which can be dealt with when specifying the combinatorial constraints just mentioned, and which each strongly depend on the particular "chemistry" under study. We would like to design a model which avoids entering into too many details but which presents some illustrative "genericity". Particularization is then left to the interested reader. In order to inherit from the main statistical features of these constrained problems, and avoiding too much specification, we shall then treat the following three-parameters average model.

Suppose the three $\{0,1\}$-valued infinite sequences defined above are now random, in the sense that $(\varepsilon_m^i)_{m\geqslant 1}$, $i=1,2,3$, is an array of independent random variables. Suppose moreover that the probability that $(\varepsilon_m^i)_{m\geqslant 1}$, $i=1,2,3$, takes value one is given by $P(\varepsilon_m^i=1)=p_i^m$, with $p_i\in(0,1]$, $i=1,2,3$.

It is then assumed that the probability that a degree m-node occurs on a tree, in any configuration, decreases exponentially with m, according to p_1^m, due to the induced obstructions. Also, the probability that a cycle of length m (resp. a configuration with m connected components) presents itself, decreases exponentially with m, like p_2^m (resp. p_3^m).

We shall then rather work with the average branch-EGF (branch-AEGF)

$$g(\theta)\stackrel{\text{def}}{=}\mathbf{E}g_1(\theta)=\mathrm{e}^{p_1\theta},$$

where symbol $\mathbf{E}$ stands for mathematical expectation with respect to probability P just defined. The resulting average tree-EGF (tree-AEGF), solution to the functional equation

$$\overline{t}(\theta)=\theta g\big(\overline{t}(\theta)\big)$$

for this particular g, will then prove worth being considered. In a similar way, we shall need to introduce the average cycle-EGF (cycle-AEGF)

$$\overline{c}(\theta)\stackrel{\text{def}}{=}\mathbf{E}c_2(\theta)=-\log(1-p_2\theta)$$

and the average set-EGF (set-AEGF)

$$e(\theta)\stackrel{\text{def}}{=}\mathbf{E}e_3(\theta)=\mathrm{e}^{p_3\theta}$$

and the various combinations of these three AEGF, obtained by composition.

Remark 1. It should be noted that such AEGF are not stricto sensu EGF, as defined in subsection 2.1, definition 1. They belong in fact to a larger class of EGF (with very similar properties) for which condition (iii) in definition 2.1 should be relaxed to

(iii)′ The sequence $(c_m)_{m\geqslant 1}$ has state-space $\mathbf{R}^+$, the set of non-negative real numbers.

Apart from this extension, the algebra of AEGF appears quite similar to the one of ordinary EGF.

For example, if one is interested in the average number of overall possible configurations, given m monomers, requires the computation of $m![\theta^m]Z(\theta)$, with

$$Z(\theta) \stackrel{\text{def}}{=} e\big(\overline{c}\big(\overline{t}(\theta)\big)\big) = \frac{1}{\big(1-p_2\overline{t}(\theta)\big)^{p_3}}.$$

Here $[\theta^m]Z(\theta) \stackrel{\text{def}}{=} c_m/m!$ denotes, as usual, the coefficient of θ^m in the θ-expansion of $Z(\theta)$.

We shall see below how to extract the asymptotic, that is for large m, equivalent of such coefficient, using singularity analysis developed by [7].

Finally, we are left with the asymptotic study of a statistical three-parameter model of branched polymers whose physical signification is now clear.

3. Singularity analysis

As was noted just before, it would be helpful to extract directly the asymptotic form of coefficients of a complicated AEGF, such as $Z(\theta)$. Fortunately, these methods exist, and we shall now briefly indicate how they particularize to our situation.

We first recall a partial result of [7].

Singularity analysis result

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

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

where θ, $\theta_0 > \theta$, and η are positive real numbers. Assume that, with $\sigma(x) = x^\alpha \log^\beta x$, $\alpha \notin \{0,-1,-2,...\}$ and β any real number we have

$$Z(\theta) \sim \sigma\left(\frac{1}{1-\theta/\theta_0}\right) \quad \text{as } \theta \to \theta_0 \text{ in } D.$$

Then, the Taylor coefficients of $Z(\theta)$ satisfy

$$[\theta^m]Z(\theta) \sim \theta_0^{-m}\frac{\sigma(m)}{m\Gamma(\alpha)} \quad \text{as } m \to \infty,$$

where $\Gamma(\alpha)$ is the Euler function.

The main source of singularity in our problem stems from the tree-AEGF. The type of singularity is algebraic with parameter $\alpha = -1/2$ (branch point) [12], as a result of the implicit function theorem [5].

Singularity of the tree-AEGF

The tree AEGF function, $\overline{t}(\theta)$, defined above is analytic in the domain D formed by the complex plane slit along $(\theta_0 \stackrel{\text{def}}{=} 1/\mathrm{e}p_1, \infty)$. For θ tending to $1/\mathrm{e}p_1$ in D, $\overline{t}(\theta)$ admits the singular expansion

$$\overline{t}(\theta) \sim \frac{1}{p_1} - \frac{\sqrt{2}}{p_1}\left(\frac{1}{1-\theta/\theta_0}\right)^{-1/2}.$$

It results that

$$[\theta^m]\overline{t}(\theta) \sim \frac{1}{p_1} - \frac{\sqrt{2}}{p_1}\theta_0^{-m}\frac{m^{-1/2}}{m\Gamma(-1/2)} \quad \text{as } m \to \infty,$$

and that $m![\theta^m]\overline{t}(\theta) \sim (p_1 m)^{m-1}$ (from the Stirling formula) as soon as $p_1 > 1/\mathrm{e}$, whereas $m![\theta^m]\overline{t}(\theta) \sim 1/p_1$ otherwise.

Singularity of the overall AEGF $Z(\theta)$

Recall

$$Z(\theta) = \frac{1}{\left(1 - p_2\overline{t}(\theta)\right)^{p_3}}.$$

The question is then to understand how the dominant singularity of $\overline{t}(\theta)$ "snowballs" into such a compound AEGF, for which there are two competing singularities depending on the value of the parameter p_2/p_1. Three cases then may arise.

- $p_2/p_1 < 1$ (subcritical):

$$Z(\theta) \sim (1 - p_2/p_1)^{-p_3}\left(1 - \frac{\sqrt{2}p_2p_3}{p_1 - p_2}\left(\frac{1}{1-\theta/\theta_0}\right)^{-1/2}\right).$$

- $p_2/p_1 = 1$ (critical):

$$Z(\theta) \sim \left(\frac{1}{2(1-\theta/\theta_0)}\right)^{p_3/2}.$$

- $p_2/p_1 > 1$ (supercritical): in this case the singularity is no longer at θ_0, but at $\theta_2 < \theta_0$ defined by $1 = p_2\overline{t}(\theta_2)$, hence $\theta_2 = \frac{1}{p_2}\mathrm{e}^{-p_1/p_2}$. It follows that

$$Z(\theta) \sim \left(\frac{1 - p_1/p_2}{1-\theta/\theta_2}\right)^{p_3},$$

around θ_2.

In each case, the asymptotics on the Taylor coefficients follow easily from the singularity analysis result.

- Subcritical case:

$$[\theta^m]Z(\theta) \sim (1-p_2/p_1)^{-p_3}\frac{p_2p_3}{p_1-p_2}\theta_0^{-m}\frac{1}{m^{3/2}\sqrt{2\pi}} \quad \text{as } m\to\infty,$$

  provided $p_1 > 1/\mathrm{e}$, $\sim (1-p_2/p_1)^{-p_3}$ otherwise.
- Critical case:

$$[\theta^m]Z(\theta) \sim \frac{1}{2^{p_3/2}}\theta_0^{-m}\frac{m^{p_3/2}}{m\Gamma(p_3/2)} \quad \text{as } m\to\infty,$$

  provided $p_1 > 1/\mathrm{e}$.
- Supercritical case:

$$[\theta^m]Z(\theta) \sim (1-p_1/p_2)^{p_3}\theta_2^{-m}\frac{m^{p_3}}{m\Gamma(p_3)} \quad \text{as } m\to\infty,$$

  provided $\theta_2 < 1$.

It should be noted here that such a distinction between critical regimes exclusively originates from the cycle-AEGF. No such critical behavior would take place if one considers the forest-AEGF $f(\theta) \stackrel{\text{def}}{=} \mathrm{e}^{p_3\bar{t}(\theta)}$, omitting cycling! Cycling is therefore a very essential feature for the asymptotic behavior of $[\theta^m]Z(\theta)$, since three distinct behaviors are then observed.

4. Structural parameters

4.1. "Marking" additive parameters of interest

So far, we have been mainly concerned by "counting" the average number of available configurations.

It is also of importance to derive the distribution of several structural parameters in the study of such average configurations such as the number of d-nodes on a tree, the number of nodes on a tree, the number of trees with d monomers, the number of trees, the number of cycles with d monomers, the number of cycles, the number of connected components with d monomers, the number of connected components...

All these can easily be obtained by considering the bivariate AEGF

$$Z(\gamma,\theta) = 1 + \sum_{m\geqslant 1}\theta^m\left(\sum_{p\geqslant 1}\frac{c_{p,m}}{m!}\gamma^p\right),$$

with a new variable $\gamma \in [0,1]$ marking the parameter under study [10]. Here $c_{p,m}$ is the number of configurations with m monomers whose parameter of interest takes value p. These global bivariate $Z(\gamma,\theta)$ can be obtained when inserting local marked AEGF into $Z(\theta)$. Here are a few examples of this mechanism.

1. $g_d(\gamma,\theta) \stackrel{\text{def}}{=} g(\theta) + (\gamma-1)([\theta^d]g(\theta))\theta^d$ is the branch-AEGF of a tree that will serve marking d-nodes ($d \geqslant 0$), including leaves for which $d = 0$.

2. $\gamma g(\theta)$ is the branch-AEGF of the tree-AEGF marking all nodes of a tree.

3. $\overline{t}(\theta) + (\gamma-1)([\theta^d]\overline{t}(\theta))\theta^d$ is the AEGF of the number of trees with d monomers.

4. $\gamma\overline{t}(\theta)$ is the local AEGF used for the number of trees' information.

5. $\overline{c}(\theta) + (\gamma-1)([\theta^d]\overline{c}(\theta))\theta^d$ is the AEGF of the number of cycles with d monomers.

6. $\gamma\overline{c}(\theta)$ is the AEGF that yields the number of cycles.

7. $\overline{c}(\overline{t}(\theta)) + (\gamma-1)([\theta^d]\overline{c}(\overline{t}(\theta)))\theta^d$ is the AEGF of the number of cycle-trees (connected components) with d monomers.

Inserting these marked AEGF into $Z(\theta)$ yield bivariate $Z(\gamma,\theta)$ coding for joint parameter and configuration number values. For example, if one wishes to study the average number of connected components, (example 6), one has to consider the bivariate AEGF

$$Z_1(\gamma,\theta) = e\big(\gamma\overline{c}\big(\overline{t}(\theta)\big)\big) = \mathrm{e}^{p_3\gamma\log\frac{1}{1-p_2\overline{t}(\theta)}}.$$

If one is interested in the average number of trees (example 4), one should focus on the bivariate AEGF

$$Z_2(\gamma,\theta) = \mathrm{e}^{p_3\log\frac{1}{1-p_2\gamma\overline{t}(\theta)}}.$$

If the number of d-monomers in a set of branched or cycle polymers is relevant, one has to study

$$e\big(\overline{c}\big(\overline{t}_d(\gamma,\theta)\big)\big) = \mathrm{e}^{p_3\log\frac{1}{1-p_2\overline{t}_d(\gamma,\theta)}},$$

with $\overline{t}_d(\gamma,\theta) \stackrel{\text{def}}{=} \theta g_d(\overline{t}_d(\gamma,\theta))$, as a result of example 1.

Observe that $Z(1,\theta) = Z(\theta)$ is the "counting" marginal, for all these examples; in other words $c_m = \sum_{p\geqslant 1} c_{p,m}$ is the number of configurations sequence.

Multivariate information is also obviously available (when one focuses on more than one parameter at the same time: variables $\gamma_1, \gamma_2, \ldots$).

Two other bivariate AEGF will also be introduced now (without justification). They are:

8. $\overline{t}_c(\gamma,\theta) = \theta g(\overline{t}_c(\gamma,\gamma\theta))$ marking the cumulative "distance" (e.g., the number of monomers) from each node of a tree to its root.

9. $\overline{t}(\gamma,\theta) = \theta g(\gamma\overline{t}(\gamma,\theta))$ marking the sum of degrees parameter over every node of a tree. This parameter is important because it measures the *connectivity* of these structures, with a competition between leaves (dangling nodes of degree zero at the boundary of the tree) and the other nodes with positive output degree, nested "inside" the tree.

There associate bivariate AEGF, for a set of polymers, are respectively

$$e\big(\overline{c}\big(\overline{t}(\gamma,\theta)\big)\big) = \mathrm{e}^{p_3\log\frac{1}{1-p_2\overline{t}_c(\gamma,\theta)}} \quad \text{and} \quad Z_3(\gamma,\theta) = \mathrm{e}^{p_3\log\frac{1}{1-p_2\overline{t}(\gamma,\theta)}}.$$

4.2. Averaging over the available configurations

Let

$$Z(\gamma,\theta) = 1 + \sum_{m\geqslant 1} \theta^m \left(\sum_{p\geqslant 1} \frac{c_{p,m}}{m!} \gamma^p \right)$$

be any of the bivariate AEGF introduced before.

There are two different levels at which one can consider the averaging problem over the configurations.

4.2.1. The thermodynamic limit $m \to \infty$

The first level proceeds as follows: suppose one is able to extract $[\theta^m]Z(\gamma,\theta)$, or possibly its asymptotic equivalent, for large m.

Then,

$$\Phi_m(\gamma) \stackrel{\text{def}}{=} \frac{1}{[\theta^m]Z(\theta)} [\theta^m]Z(\gamma,\theta) = \sum_{p\geqslant 1} \frac{c_{p,m}}{c_m} \gamma^p$$

is the probability generating function (PGF) of the variable which has been marked by γ, as a function of the number m of monomers.

In particular,

$$\Phi_m^{(p)}(1) \stackrel{\text{def}}{=} \frac{1}{[\theta^m]Z(\theta)} [\theta^m]Z^{(p)}(1,\theta)$$

with

$$Z^{(p)}(1,\theta) \stackrel{\text{def}}{=} \frac{\partial^p}{\partial \gamma^p} Z(\gamma,\theta) \mid_{\gamma=1}$$

can be extracted if the moments of the distribution are helpful ($p = 1$ giving the mean...).

Depending on the three internal control parameters (p_1, p_2, p_3), however, $\Phi_m(\gamma)$ may converge or diverge, in the thermodynamic limit $m \to \infty$, as we shall now see. This happens if the parameter of interest is the number of connected components. This is the signature of a phase transition, known as percolation. The point separating these two behaviors will be called critical, as usual. At critical point it is of interest to study the fluctuations of the parameter versus its mean value, since these are expected to be comparatively large.

Number of connected components; the percolation result (gel–sol transition). We shall study now in some detail the behavior of the PGF, associated to $Z_1(\gamma,\theta)$, that is

$$\Phi_{1,m}(\gamma) \stackrel{\text{def}}{=} \frac{1}{[\theta^m]Z(\theta)} [\theta^m]Z_1(\gamma,\theta),$$

which focuses on the asymptotic number of connected components. We have the percolation result which follows directly from the asymptotic analysis on $Z(\theta) = Z_1(1,\theta)$.

- In the subcritical case, $p_2/p_1 < 1$, two cases arise.

  If $\theta_0 < 1$ (or $p_1 > 1/\mathrm{e}$), then $\Phi_{1,m}(\gamma) \sim \gamma \mathrm{e}^{-\lambda(1-\gamma)}$ as $m \to \infty$.
  If $\theta_0 \geqslant 1$ (or $p_1 \leqslant 1/\mathrm{e}$), then $\Phi_{1,m}(\gamma) \sim \mathrm{e}^{-\lambda(1-\gamma)}$ as $m \to \infty$.

- Therefore, $\Phi_{1,m}(\gamma)$ converges in this case to a Poisson distribution with intensity given by $\lambda = p_3 \log 1/(1 - p_2/p_1) > 0$, independent of m.
  We then have a "gel", for which the average number of connected components is finite in the thermodynamic limit $m \to \infty$, possibly equal to one on some surface in the internal parameter space defined by $\lambda = 1$.

- In the critical case, $p_2/p_1 = 1$,

$$\Phi_{1,m}(\gamma) \sim \mathrm{e}^{-(1-\gamma)p_3/2 \log(m/2)} \frac{\Gamma(p_3/2)}{\Gamma(\gamma p_3/2)} \quad \text{as } m \to \infty.$$

  Therefore, $\Phi_{1,m}(\gamma)$ converges in this case to a Poisson distribution with intensity given by $\lambda_m = (p_3/2)\log(m/2)$, providing the divergent (logarithmic) mean and variance of the asymptotic distribution for the number of connected components.
  Alternatively $\Phi_{1,m}^{1/\lambda_m}(\gamma) \sim \mathrm{e}^{-(1-\gamma)}$ as $m \to \infty$ (Poisson with intensity 1).

- In the supercritical case, $p_2/p_1 > 1$,

$$\Phi_{1,m}(\gamma) \sim (1 - p_1/p_2)^{(\gamma-1)p_3} \frac{\Gamma(p_3) m^{(\gamma-1)p_3}}{\Gamma(\gamma p_3)} \quad \text{as } m \to \infty.$$

  Alternatively $\Phi_{1,m}^{1/\lambda_m}(\gamma) \sim \mathrm{e}^{-(1-\gamma)}$ as $m \to \infty$ (Poisson with intensity 1), with $\lambda_m = p_3 \log((1 - p_1/p_2)m)$.
  We then have a "sol".

Number of trees. As was already underlined, this parameter is encoded in the PGF related to

$$Z_2(\gamma, \theta) = \mathrm{e}^{p_3 \log \frac{1}{1 - p_2 \gamma \bar{t}(\theta)}},$$

that is

$$\Phi_{2,m}(\gamma) \stackrel{\text{def}}{=} \frac{1}{[\theta^m] Z(\theta)} [\theta^m] Z_2(\gamma, \theta).$$

Proceeding in the same way, one can show the following result.

- In the subcritical case, $p_2/p_1 < 1$, two cases arise.
  If $\theta_0 < 1$ (or $p_1 > 1/\mathrm{e}$), then

$$\Phi_{2,m}(\gamma) \sim \gamma \left(\frac{1 - p_2/p_1}{1 - \gamma p_2/p_1} \right)^{p_3+1} \quad \text{as } m \to \infty, \text{ for } \gamma \leqslant p_1/p_2.$$

  If $\theta_0 \geqslant 1$ (or $p_1 \leqslant 1/\mathrm{e}$), then

$$\Phi_{2,m}(\gamma) \sim \left(\frac{1 - p_2/p_1}{1 - \gamma p_2/p_1} \right)^{p_3} \quad \text{as } m \to \infty.$$

Therefore, $\Phi_{2,m}(\gamma)$ converges in this case to a negative binomial distribution.

- In the critical case, $p_2/p_1 = 1$,

$$\Phi_{2,m}(\gamma) \sim \sum_{p\geqslant 0} \frac{C_p}{p!}\big(\sqrt{m/2}(\gamma-1)\big)^p \quad \text{as } m\to\infty,$$

with

$$C_p = \frac{\Gamma(p_3/2)\Gamma(p_3+p)}{\Gamma(p_3)\Gamma((p_3+p)/2)}.$$

Therefore the mean number of trees grows like $C_1\sqrt{m/2}$, whereas the variance grows like $(C_2 - C_1^2)m/2$ (there are very large fluctuations here at critical point since the standard deviation and the mean are of the same order of magnitude).

- In the supercritical case, $p_2/p_1 > 1$,

$$\Phi_{2,m}(\gamma) \sim \left(\frac{\gamma p_2 - p_1}{\gamma p_2 - \gamma p_1}\right)^{p_3}\left(\gamma \mathrm{e}^{\frac{p_1}{\gamma p_2}-\frac{p_1}{p_2}}\right)^m \quad \text{as } m\to\infty, \text{ for } \gamma \geqslant p_1/p_2.$$

Therefore, the mean grows like $(1-p_1/p_2)m + p_1p_3/(p_2-p_1)$, whereas the variance is asymptotically equivalent to $(p_1/p_2)m - p_1p_2p_3/(p_2-p_1)^2$.

Sum of degrees parameter (connectivity). Concerning this parameter, one needs to study the PGF, associated to

$$Z_3(\gamma,\theta) = \mathrm{e}^{p_3 \log \frac{1}{1-p_2\bar{t}(\gamma,\theta)}},$$

with $\bar{t}(\gamma,\theta) = \theta \mathrm{e}^{p_1\gamma\bar{t}(\gamma,\theta)}$ that is

$$\Phi_{3,m}(\gamma) \stackrel{\text{def}}{=} \frac{1}{[\theta^m]Z(\theta)}[\theta^m]Z_3(\gamma,\theta).$$

Proceeding in the same way, one can prove the following results.

- In the subcritical case, $p_2/p_1 < 1$, if $\theta_0 < 1$,

$$\Phi_{3,m}(\gamma) \sim \left(\frac{1-p_2/p_1}{1-p_2/\gamma p_1}\right)^{p_3+1}\gamma^{m-1} \quad \text{as } m\to\infty,$$

provided $\gamma > p_2/p_1$.
Therefore, $\Phi_{3,m}^{1/m}(\gamma) \sim \gamma$ as $m\to\infty$. The mean connectivity grows like m.

- In the critical case, $p_2/p_1 = 1$, the mean connectivity grows like m/e, and the standard deviation like $\frac{\sqrt{3}}{2}m/\mathrm{e}$ (large fluctuations at critical point), with no dependence on p_3.
- In the supercritical case, $p_2/p_1 > 1$ and

$$\Phi_{3,m}(\gamma) \sim \left(\frac{1-\gamma p_1/p_2}{1-p_1/p_2}\right)^{p_3}\mathrm{e}^{-mp_1/p_2(1-\gamma)}.$$

Alternatively, $\Phi_{3,m}^{1/m}(\gamma)$ is the PGF of a Poisson variable of intensity p_1/p_2.

As far as the mean is concerned, this parameter therefore is quite "blind" to the phase transition which occurred for example on the connected components, or number of trees parameters. This is because, at the level of a single tree, the average connectivity can easily be shown to converge to one; there is an exact statistical balance between the (degree-zero) leaves at the boundary of the tree and the nodes in the bulk of the tree.

4.2.2. Randomizing the number of monomers

The other averaging level proceeds as follows: let the conditional configuration number $c_{p/m}$ be defined by $c_{p/m} \stackrel{\text{def}}{=} c_{p,m}/c_m$.

Then

$$Z(\gamma,\theta) = 1 + \sum_{m\geqslant 1} \frac{\theta^m c_m}{m!} \left(\sum_{p\geqslant 1} c_{p/m} \gamma^p \right).$$

Suppose the exact number of constitutive monomers is now *unknown*; it is then a classical model in statistical physics to assume that the probability to have m monomers at "fugacity" θ is

$$p_\theta(m) = \frac{1}{Z(1,\theta)} \frac{\theta^m c_m}{m!}.$$

The probability to have m monomers is therefore proportional to the number of configurations c_m, as required. Therefore the quantity of interest will now be the γ-probability generating function (PGF) $Z_\theta(\gamma) \stackrel{\text{def}}{=} Z(\gamma,\theta)/Z(1,\theta)$, as a function of θ. This approach sees parameter θ as an external parameter and is a standard way to proceed in statistical physics. The conclusions appear now quite different, for the parameters of interest. We illustrate this point of view on the number of components and number of trees parameters. Similar results can be obtained for other parameters.

Number of connected components. Concerning the PGF related to $Z_1(\gamma,\theta)$, i.e.,

$$Z_{1,\theta}(\gamma) \stackrel{\text{def}}{=} Z_1(\gamma,\theta)/Z_1(1,\theta),$$

we have $Z_{1,\theta}(\gamma) = \mathrm{e}^{-\lambda(\theta)(1-\gamma)}$, with $\lambda(\theta) = p_3 \log 1/(1 - p_2\overline{t}(\theta))$. This variable is Poisson distributed with intensity $\lambda(\theta)$. The behavior (definiteness) of the normalizing denominator of $Z_1(\gamma,\theta)$, which is $Z(\theta)$, is then crucial in this approach and we have the specific percolation result (gel–sol transition).

- $p_2/p_1 \leqslant 1$ (subcritical and critical): $Z_{1,\theta}(\gamma)$ is defined for $\theta < \theta_0 = 1/(\mathrm{e}p_1)$.
- $p_2/p_1 > 1$ (supercritical): in this case the θ-singularity is no longer at θ_0, but at $\theta_2 = \frac{1}{p_2}\mathrm{e}^{-p_1/p_2} < \theta_0$. It follows that $Z_{1,\theta}(\gamma)$ is defined only for $\theta < \theta_2$.

Crossing the critical value from below, in this approach, translates into a simple shift to the left in the external parameter space θ. In both cases, this approach predicts

the existence of a critical value θ_c (which could be either θ_0 or θ_2). Ordinary statistical mechanics appears to be undefined for $\theta > \theta_c$. The behavior near $\theta = \theta_c$ of the thermodynamical functions $Z_\theta(\gamma)$ is derivable. One may infer from this that θ_c may be some kind of limiting fugacity, although it is unclear whether $\theta > \theta_c$ may be reached following a phase transition whose nature still remains mysterious.

Number of trees. The PGF associated to $Z_2(\gamma, \theta)$ is

$$Z_{2,\theta}(\gamma) = \left(\frac{1 - p_2 \bar{t}(\theta)}{1 - \gamma p_2 \bar{t}(\theta)} \right)^{p_3}.$$

This parameter is distributed like a negative binomial, with the same phenomena as before.

5. Infinite divisibility of (A)EGF and related notions

We would like to end up this monograph by introducing the notion of infinite-divisibility which we think deserves interest in our context (infinitely divided matter). This notion and the related one of self-decomposability (resp. stability) derive their importance from the fact that they are the solution to a central limit problem: the set of self-decomposable distributions coincides with the set of limit laws of normalized sums of independent (resp. identically distributed) random variables.

Let us first connect the worlds of (A)EGF to the one of discrete probability. As was underlined just before, to each (A)EGF it is possible to associate the θ-family of discrete probability distributions

$$p_\theta(m) \stackrel{\text{def}}{=} \frac{c_m \theta^m}{Z(\theta) m!}, \quad m \geqslant 0, \ \theta \in D^+.$$

This defines the probability that the population is constituted with m monomers, as a function of external control parameter θ.

If

$$f_\theta(u) \stackrel{\text{def}}{=} \sum_{m \geqslant 0} p_\theta(m) u^m$$

is now the associated PGF of this probability distribution, we have

$$f_\theta(u) \stackrel{\text{def}}{=} \frac{Z(\theta u)}{Z(\theta)}$$

for $\theta \in D^+$, $u \in [0, 1]$. Obviously, $f_\theta(u)$ is a convex function of u for any θ in its definition domain.

Example 7. We give here some generating functions (PGF) associated to some simple EGF of section 2.1. An extension to AEGF is immediate.

(1) $f_\theta(u) = \mathrm{e}^{-\theta(1-u)}$ (Poisson);

(2) $f_\theta(u) = \frac{1-\log(1-\theta u)}{1-\log(1-\theta)}$ (logarithmic);
(3) $f_\theta(u) = \frac{1-\theta}{1-\theta u}$ (Pascal);
(10) $f_\theta(u) = \frac{1+\theta u}{1+\theta}$ (Bernoulli).

If $f_\theta(u)$ has been derived in such a way from some $Z(\theta)$, we shall say that the PGF derives from an (A)EGF. Note that this interpretation sheds some new light on the probabilistic meaning of these very standard distributions.

Let us now introduce the notion of infinite divisibility (ID).

Definition 8. An (A)EGF $Z(\theta)$ is said to be infinitely divisible if $Z(\theta)^{1/n}$ has non-negative Taylor coefficients for any $n > 0$.

It easily follows from the above definition that

Proposition 9. An (A)EGF $Z(\theta)$ is infinitely divisible iff $l(\theta) \stackrel{\text{def}}{=} 1 + \log Z(\theta)$ is itself an (A)EGF.

This results from the fact that a simple limit of (A)EGF remains an (A)EGF, if it exists.

Remark 2. The singularities of $l(\theta)$ are now to be searched within the (disjoint) union of the singularities of $Z(\theta)$ and zeros of $Z(\theta)$ (necessarily in the complex non-positive half plane). As was underlined before, the (A)EGF $Z(\theta)$ has a real positive dominant singularity θ_0. If θ_1 is now the (complex) dominant zero of $Z(\theta)$, we have: the dominant singularity of $l(\theta)$ is θ_1 if $|\theta_1| < \theta_0$, θ_0 otherwise.

This yields a constructive way to decide on infinite divisibility.

Proposition 10. If $|\theta_1| < \theta_0$, $Z(\theta)$ is not infinitely divisible.

Example 11. Examples 1–6 of section 2.1 are infinitely divisible EGF, whereas examples 7–10 are not. Concerning the tree-AEGF, $t(\theta)$, $\theta_0 = 1/\mathrm{e}p_1$ and $\theta_1 = -\mathrm{e}^{p_1}$. It follows that $t(\theta)$ is not ID if $p_1 < p_1^*$, with $1/\mathrm{e}p_1^* = \mathrm{e}^{p_1^*}$. Of course the overall AEGF

$$Z(\theta) = \frac{1}{\big(1 - p_2 \bar{t}(\theta)\big)^{p_3}}$$

is ID, by construction.

Let us now have a look to the probabilistic version of the infinite divisibility notion introduced above on (A)EGF. Let $\bar{l}(\theta) \stackrel{\text{def}}{=} \log Z(\theta)$, so that $Z(\theta) = \mathrm{e}^{\bar{l}(\theta)}$. The associated generating function is therefore

$$f_\theta(u) \stackrel{\text{def}}{=} \frac{Z(\theta u)}{Z(\theta)} = \mathrm{e}^{\overline{l}(\theta)(L_\theta(u)-1)}$$

with $L_\theta(u) \stackrel{\text{def}}{=} \overline{l}(\theta u)/\overline{l}(\theta)$. It is the one of a compound Poisson variable with intensity $\overline{l}(\theta)$, whose generating function in the jump (mark) space is $L_\theta(u)$, with $L_\theta(0) = 0$. Observe that $L_\theta(u)$ is a generating function of an atomic probability at the only condition that $\overline{l}(\theta)$ has non-negative Taylor coefficients which means that $Z(\theta)$ is infinitely divisible, as required.

Infinite divisibility of (A)EGF matches therefore with the standard definition of infinite divisibility of probability generating functions (the ones of compound Poisson variables) [6].

The PGF of an ID distribution admits also the following useful representation, adapting a result of [13],

$$f_\theta(u) = \mathrm{e}^{\int_1^u R_\theta(v)\,\mathrm{d}v}.$$

Here, $R_\theta(u) \stackrel{\text{def}}{=} \overline{l}(\theta)L'_\theta(u)$, $u \in [0,1]$, is called the θ-canonical measure (with "prime" indicating partial derivation of $L_\theta(u)$ with respect to u). We shall let

$$R_\theta(u) \stackrel{\text{def}}{=} \sum_{n\geqslant 0} r_\theta(n)u^n, \quad r_\theta(n) \geqslant 0,$$

define the power series expansion of $R_\theta(u)$.

This representation will prove helpful to distinguish between subclasses of ID distributions. Among these, discrete self-decomposable (A)EGF play a remarkable role. Discrete self-decomposability should not be confused with its standard continuous counterpart [11].

Definition 12. An (A)EGF $Z(\theta)$ is self-decomposable if $\forall \theta, \theta' \in D^+, \theta' \leqslant \theta$, $\forall p \in (0,1)$, there is an (A)EGF $Z_p(\theta)$ such that

$$Z(\theta')Z_p(\theta) = Z\big((1-p)\theta + p\theta'\big)Z_p(\theta').$$

Changing indeed $\theta' = \theta u$, $u \in [0,1]$, in the above equation characterizing self-decomposability for (A)EGF, yields the following definition.

Definition 13. $f_\theta(u)$ is self-decomposable if $\forall p \in (0,1)$, there is a (probability) generating function $f_{p,\theta}(u)$ such that $f_\theta(u) = f_\theta(1-p(1-u))f_{p,\theta}(u)$.

This is the standard (discrete) version of self-decomposability of probability distributions on the integers. We then have the characterization property

Proposition 14. An (A)EGF $Z(\theta)$ is self-decomposable iff

$$h_\theta(u) \stackrel{\text{def}}{=} 1-(1-u)\frac{L'_\theta(u)}{L'_\theta(0)}$$

defines a u-PGF such that $h_\theta(0)=0$. As a consequence

$$f_\theta(u) \stackrel{\text{def}}{=} \frac{Z(\theta u)}{Z(\theta)} = \mathrm{e}^{\int_1^u R_\theta(v)\,\mathrm{d}v}$$

with canonical measure

$$R_\theta(u) = \overline{l}(\theta)L'_\theta(0)\frac{1-h_\theta(u)}{1-u}.$$

In an equivalent manner, this means that the Taylor coefficients of $R_\theta(u)$, $(r_\theta(n))_{n\geqslant 1}$, is a non-increasing sequence of n. As a result, the associated $p_\theta(n)$, $n\geqslant 0$, is unimodal, with mode at the origin iff $r_\theta(0)=p_\theta(1)/p_\theta(0)\leqslant 1$ [13]. The self-decomposable subclass of ID distributions therefore focuses on unimodal distributions, with mode possibly at the origin.

Example 15. 1. Let $Z(\theta)=1/(1-\theta)$, $\theta\leqslant 1$, (example 3 (3)) which is ID. Then

$$f_\theta(u)=\frac{1-\theta}{1-\theta u}=\mathrm{e}^{\log\frac{1}{(1-\theta)}(\frac{\log(1-\theta u)}{\log(1-\theta)}-1)},$$

identifying $\overline{l}(\theta)$ and $L_\theta(u)$. It follows that

$$h_\theta(u)\stackrel{\text{def}}{=}1-(1-u)\frac{L'_\theta(u)}{L'_\theta(0)}=\frac{(1-\theta)u}{1-\theta u},$$

which is a PGF. $Z(\theta)$ therefore is self-decomposable and $f_\theta(u)$ is the PGF of a unimodal distribution. Moreover, $R_\theta(u)=\theta/(1-\theta u)$ and $r_\theta(0)=\theta\leqslant 1$ in the whole parameter range. Mode is always at the origin for any admissible value of the parameter.

2. Let $Z(\theta)=\mathrm{e}^\theta$, $\theta\geqslant 0$, (example 3 (1)), which clearly is ID, and for which $f_\theta(u)=\mathrm{e}^{\theta(u-1)}$ (Poisson) so that $\overline{l}(\theta)=\theta$ and $L_\theta(u)=u$, independent of θ. It follows that $h_\theta(u)=u$ ($Z(\theta)$ is self-decomposable) and that $R_\theta(u)=\theta$. Thus mode is at the origin iff $r_\theta(0)=\theta\leqslant 1$, otherwise at $[\theta]$ (integer part of θ), as is well-known.

Actually, Poisson distributions belong to a remarkable subclass of self-decomposable distributions, namely the discrete α-stable distributions whose generating function is defined as follows.

Definition 16. $f_\theta(u)$ is α-stable, $\alpha\in(0,1]$, if $\forall p\in(0,1)$, it satisfies the functional equation

$$f_\theta(u)=f_\theta\big(1-p^{1/\alpha}(1-u)\big)f_\theta\big(1-(1-p)^{1/\alpha}(1-u)\big),$$

whose solution is $f_\theta(u)=\mathrm{e}^{-\lambda(\theta)(1-u)^\alpha}$, $\lambda(\theta)>0$, including Poisson ($\alpha=1$).

References

[1] E.T. Bell, Exponential polynomials, Ann. Math. 35 (1934) 258–277.
[2] A. Cayley, A theorem on trees, Quart. J. Math. 23 (1889) 376–378.
[3] L. Comtet, *Analyse Combinatoire* (P.U.F., Paris, 1970).
[4] M. Daoud, Polymers and fragmentation, in: *Proceedings of the Workshop Fragmentation Phenomena*, eds. D. Beysens, X. Campi and E. Pefferkorn (World Scientific, Les Houches, 1993).
[5] J. Dieudonné, *Calcul Infinitésimal* (Hermann, Paris, 1980).
[6] W. Feller, *An Introduction to Probability Theory and its Applications*, Vols. 1,2 (Wiley, New York, 1968).
[7] P. Flajolet and A. Odlyzko, Singularity analysis of generating functions, SIAM J. Disc. Math. 3(2) (1990) 216–240.
[8] P.G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, 1979).
[9] I.P. Goulden and D.M. Jackson, *Combinatorial Enumeration* (Wiley, New York, 1983).
[10] V.F. Kolchin, *Random Mappings* (Optimization Software Inc., New York, 1986).
[11] E. Lukacs, *Developments in Characteristic Function Theory* (Griffin, London, 1983).
[12] A. Meir and J.W. Moon, On the altitude of nodes in random trees, Can. J. Math. 30 (1978) 997–1015.
[13] F.W. Steutel and K. van Harn, Discrete analogues of self-decomposability and stability, Ann. Probab. 7 (1979) 893–899.