Master Informatique Montpellier 2
UMINR318 — Évaluation des performances

Anne-Elisabeth Baert, Alain Jean-Marie

Version 1.4 du 24 octobre 2006

1 Introduction

L'objet de ce cours est de passer en revue quelques modèles probabilistes utilisés dans le domaine de l'*évaluation quantitative de performance* pour l'analyse des systèmes à événements discrets (en abrégé : SED) en général, et celle des réseaux de communication en particulier. D'autres applications sont possibles dans le domaine de l'informatique et l'algorithmique comme l'analyse d'algorithmes probabilistes, séquentiels ou distribués.

La modélisation probabiliste des SED est un domaine vaste qui englobe de nombreuses techniques mathématiques et de nombreux résultats.

Ce texte met en valeur certains des aspects les plus en rapport avec :

- les objets combinatoires : graphes, chemins, mots, arbres,
- les méthodes analytiques similaires à celles mises en œuvre en analyse d'algorithmes : fonctions génératrices, analyse asymptotique.

Sources/Bibliographie Les principales sources de ce cours sont :

- Pour les fonctions génératrices en général : les livres de Sedgewick et Flajolet [15] et Wilf [19]. Pour les résultats en théorie des fonctions complexes, les livres de Titchmarsh [18] et Rudin [13].
- Pour les résultats sur les puissances de matrices, et le théorème de Perron-Frobenius, les livres de Seneta [16] et Minc [10].
- Pour les fondements en probabilités, les livres de Feller, surtout le premier volume [1].
- Pour les Chaînes de Markov en temps et espace discrets, les livres de Kemeny et Snell [8] et Kleinrock [9]. Pour les Chaînes de Markov en temps continu, les livres de Feller [1], Ross [12], ou Kleinrock [9]. Pour la réversibilité, le livre de Kelly [7].
- Pour les modèles de files d'attente, les livres de Takacs [17], Kleinrock [9] et Kelly [7]. Pour les applications aux réseaux, les livres de Pujolle et Fdida [11].
- Pour les processus de branchement, le livre de Harris [6].

2 Représentations combinatoires et fonctions génératrices

2.1 Combinatoire Analytique

Les fonctions génératrices sont des séries formelles. A l'intérieur de leur disque de convergence ce sont des séries entières. Elles sont aussi appelées *séries génératrices*. Ces fonctions offrent une manipulation très souple quand l'utilisateur arrive à bien distinguer les informations combinatoires qu'elles contiennent. Les fonctions génératrices sont utilisées pour trouver le nombre de composants dans une structure, une séquence ou un cycle. Elles s'appliquent à une grande variété d'objets dont ceux issus de l'analyse d'algorithmes ; elles sont l'un des outils privilégiés de cette thèse pour étudier les graphes aléatoires. Pour une description plus complète, on peut se référer au livre [15] de Sedgewick et Flajolet.

2.2 Fonctions génératrices

Définition 1 (- Classe combinatoire -). *Une classe combinatoire est un ensemble $\mathcal{C}$, fini ou dénombrable, muni d'une fonction taille $|.|$: $\mathcal{C} \mapsto \mathbb{N}$ tel que pour tout $n \in \mathbb{N}$, l'ensemble*

$$\mathbb{C}_n = \{x \in \mathcal{C}/|x| = n\} \tag{1}$$

soit fini.

Une classe est la représentation des objets combinatoires étudiés. On a, par exemple, la classe des permutations, des mots, des graphes. Le problème de dénombrement sur une classe $\mathcal{C}$ consiste alors à trouver la suite des $c_n = \text{card}(\mathbb{C}_n)$ où $\mathbb{C}_n$ représente l'ensemble des objets de $\mathcal{C}$ de taille n..

2.2.1 Fonctions génératrices ordinaires

Nous pouvons introduire les fonctions génératrices ordinaires (ordinary generating function (OGF)) qui représentent, d'une manière globale, l'ensemble de la suite des c_n.

Définition 2 (-Classe-). *La fonction génératrice ordinaire d'une classe combinatoire $\mathbb{C}$ est définie par :*

$$C(z) = \sum_{x \in \mathcal{C}} z^{|x|} = \sum_{n=0}^{\infty} c_n z^n \,, \tag{2}$$

avec $c_n = \text{card}(\mathbb{C}_n)$.

Le caractère formel de ces séries correspond au fait que nous ne nous préoccupons pas, à priori, de leur convergence.

Notons alors que cette définition dépend de la taille des objets considérés et généralise la notion de *polynôme.*

Par ailleurs, si $C(z) = \sum_n c_n z^n$, nous utilisons la notation

$$[z^n]\, C(z) = c_n$$

pour représenter le $n^{\text{ième}}$ coefficient de C.

2.2.2 Fonctions génératrices exponentielles

Un objet étiqueté voit sa structure enrichie par des étiquettes. Les fonctions génératrices exponentielles servent à énumérer les structures étiquetées. Elles différent des fonctions génératrices ordinaires par un facteur $1/n!$ au terme de degré n.

Définition 3 (Fonctions génératrices exponentielles). *: La fonction génératrice exponentielle d'une classe combinatoire $\mathbb{C}$ est définie par :*

$$\hat{C}(z) = \sum_{x \in \mathcal{C}} \frac{z^{|x|}}{|x|!} = \sum_{n=0}^{\infty} c_n \frac{z^n}{n!} \,, \tag{3}$$

avec $c_n = \text{card}(C_n)$.

2.2.3 Fonctions génératrices bivariées

Les *fonctions génératrices bivariées* (ou BGF) ordinaires ou exponentielles permettent de trouver le nombre de composants dans une structure composée : une séquence ou un cycle par exemple. Plus généralement, les fonctions génératrices bivariées permettent d'accéder à des paramètres additifs, définis par induction sur des objets combinatoires.

Définition 4. *: Soit $\mathcal{A}$ une classe de structures combinatoires, et soit ϵ un paramètre qui est une fonction de $\mathcal{A}$ dans $\mathbb{N}$. Soit $A_{n,k}$ le nombre d'objets dans A_n dont la valeur de ϵ est k, c'est à dire :*

$$A_{n,k} = \text{card}\,\{\alpha \in \mathbb{N} \textit{ tel que } |\alpha| = n \textit{ et } \epsilon(\alpha) = k\}.$$

La fonction génératrice bivariée ordinaire $A(u,z)$ et exponentielle $\hat{A}(u,z)$ de A avec la variable u marquant le paramètre ϵ est :

$$\begin{aligned} A(u,z) &= \sum_{\alpha \in A} u^{\epsilon(\alpha)}\, z^{|\alpha|} \\ &= \sum_{n,k} A_{n,k}\, u^k\, z^n\,; \end{aligned} \tag{4}$$

$$\begin{aligned} \hat{A}(u,z) &= \sum_{\alpha \in A} u^{\epsilon(\alpha)} \frac{z^{|\alpha|}}{|\alpha|!} \\ &= \sum_{n,k} A_{n,k} u^k \frac{z^n}{n!}\,. \end{aligned} \tag{5}$$

Les fonctions génératrices bivariées sont utilisées dans les graphes pour effectuer des énumérations. Ainsi, si $\mathcal{H}$ est l'ensemble des graphes ayant m arêtes et n sommets, nous lui associons la fonction génératrice bivariée suivante :

$$H(u,z) = \sum_{m,n} h_{m,k} u^m \frac{z^n}{n!}\,,$$

où $h_{m,n}$ est le nombre total de graphes ayant m arêtes pour n sommets.
Exemple : La fonction génératrice bivariée pour un triangle, c'est à dire, un graphe ayant 3 sommets et 3 arêtes est :

$$\frac{w^3 z^3}{3!}$$

La fonction génératrice bivariée pour 2 sommets isolés est :

$$\frac{z^2}{2!}$$

Par rapport aux techniques combinatoires classiques de traitement des suites et des récurrences, les séries génératrices apportent un caractère synthétique à la résolution. Elles permettent surtout, grâce aux techniques asymptotiques, d'étendre considérablement le nombre de problèmes que l'on sait résoudre.

2.3 Opérations sur les fonctions génératrices

Les descriptions combinatoires peuvent être traduites par des opérations sur les fonctions génératrices associées aux différentes structures combinatoires. La *méthode symbolique* des fonctions génératrices consiste à établir cette équivalence.

2.3.1 La méthode symbolique

On peut considérer les constructions combinatoires sur les classes comme des constructions ensemblistes classiques. Il faut toutefois préciser comment la fonction taille, dans l'ensemble ainsi produit, est obtenue à partir des tailles dans les ensembles de départ. Nous nous limiterons ici aux constructions usuelles. Pour une description plus complète, on peut se reporter au livre de Sedgewick et Flajolet [15]. De plus, un ensemble de constructions les plus intéressantes en combinatoire se trouvent dans le livre de Goulden et Jackson [4]. Pour les fonctions génératrices exponentielles les principaux opérateurs ont été introduits par Foata [3].

Nous énonçons ici deux théorèmes qui permettent de traduire les opérations sur les structures combinaires par des opérations sur les fonctions génératrices [19].

Théorème 5 (-Opérations sur les fonctions génératrices ordinaires-)**.** *: Soient $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, trois classes combinatoires auxquelles nous associons les fonctions génératrices ordinaires $A(z)$, $B(z)$, $C(z)$. Nous avons les opérateurs associés suivants :*

Union disjointe	$\mathcal{C}=\mathcal{A}\cup\mathcal{B}$	$C(z)=A(z)+B(z)$
Produit cartésien	$\mathcal{C}=\mathcal{A}\times\mathcal{B}$	$C(z)=A(z).B(z)$
Suite finie	$\mathcal{C}=\mathcal{A}^*$	$C(z)=(1-A(z))^{-1}$
Ensemble fini	$\mathcal{C}=2^{\mathcal{A}}$	$C(z)=\exp(A(z))$

Théorème 6 (-Opérations sur les fonctions génératrices exponentielles-)**.** *: Soient $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, trois classes combinatoires auxquelles nous associons les fonctions génératrices exponentielles $\hat{A}(z)$, $\hat{B}(z)$, $\hat{C}(z)$. Nous avons les opérateurs associés suivants :*

Union disjointe	$\mathcal{C}=\mathcal{A}\cup\mathcal{B}$	$\hat{C}(z)=\hat{A}(z)+\hat{B}(z)$
Produit partitionnel	$\mathcal{C}=\mathcal{A}\star\mathcal{B}$	$\hat{C}(z)=\hat{A}(z)\,\hat{B}(z)$
Séquence	$\mathcal{C}=Seq\mathcal{A}$	$\hat{C}(z)=(1-\hat{A}(z))^{-1}$
Ensemble	$\mathcal{C}=Set\mathcal{A}$	$\hat{C}(z)=\exp(\hat{A}(z))$
Cycle	$\mathcal{C}=Cycle\mathcal{A}$	$\hat{C}(z)=\ln\left(\left(1-\hat{A}(z)\right)^{-1}\right)$

Exemple : La fonction génératrice d'un triangle est donnée par $\frac{w^3z^3}{3!}$. On obtient aisément celle de deux sommets isolés : $\frac{z^2}{2!}$. Si nous voulons connaître le nombre de graphes étiquetés formés d'un triangle et de deux sommets isolés :

$$\left(\frac{w^3z^3}{3!}\right)\left(\frac{z^2}{2!}\right) = 10w^3\frac{z^5}{5!}.$$

Il existe donc 10 graphes étiquetés différents formés d'un triangle et deux sommets isolés ayant 5 sommets étiquetés comme repésentés dans la figure 1.

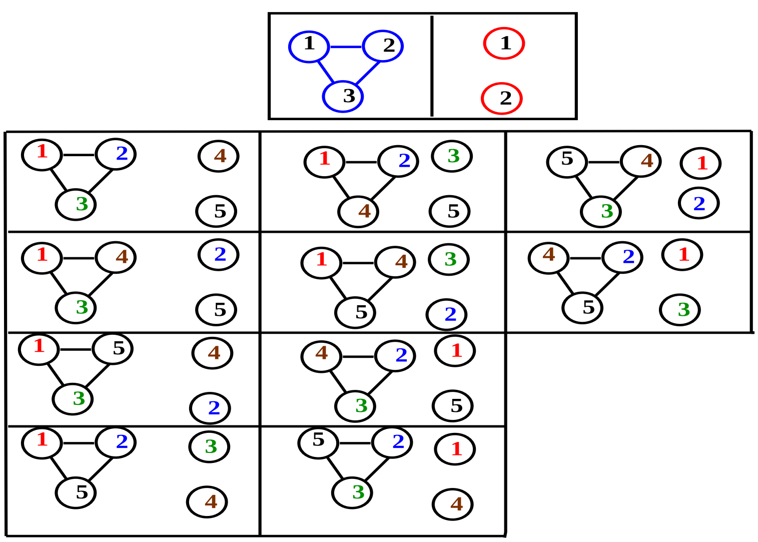

FIG. 1 – Nombre de graphes étiquetés formés d'un triangle et de deux sommets isolés

2.3.2 Opérateur de pointage

L'opérateur de pointage intervient fréquemment en analyse combinatoire de structures. Il consiste à *distinguer* une sous-structure dans un objet de taille n (voir figure 2).

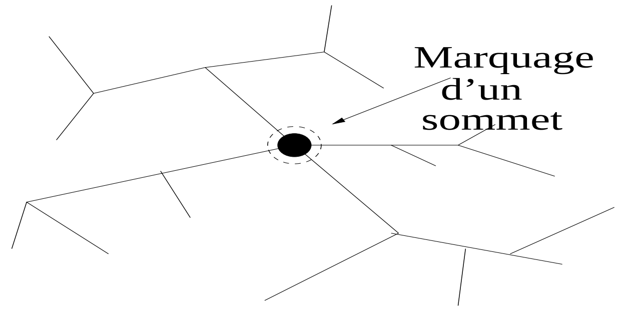

FIG. 2 – Marquage d'un sommet dans un graphe

Analytiquement, cette opération peut être définie comme :

$$\vartheta_x : F(x) \mapsto x\frac{\partial F(x)}{\partial x} \tag{6}$$

Donc, nous avons

$$[x^n]\,\vartheta_x F(x) = n\,[x^n]\,(F(x))\,.$$

2.4 Fonctions génératrices en calcul des probabilités

Les méthodes de l'évaluation de performances (systèmes à événements discrets stochastiques) font souvent appel aux *fonctions (ou : séries) génératrices*. Soit X une variable aléatoire discrète, à valeurs dans $\mathbb{N}$. On définit sa *fonction génératrice de probabilités* par :

$$F_X(z) \;=\; \mathbb{E}(z^X) \;=\; \sum_{n=0}^{\infty} \mathbb{P}(X=n)z^n\ . \tag{7}$$

Une fonction génératrice de probabilités a les propriétés suivantes :

- $F_X(z)$ existe sur le disque unité fermé : $\{z : |z| \leq 1\}$. En d'autres termes, le rayon de convergence de la série est au moins égal à 1.
- $F_X(z)$ est analytique dans le disque unité ouvert et continue sur le cercle unité. En particulier :

$$\lim_{z\to 1} F_X(z) \;=\; 1.$$

- En tant que fonction réelle, $F_X(x)$ est C^∞ (c'est-à-dire qu'on peut la dériver autant de fois qu'on veut) sur $]-1,1[$, croissante et convexe sur $[0,1]$.
- Les probabilités de X sont les coefficients du développement en série entière de F_X en $z=0$. Voir le paragraphe 2.5 sur la façon de calculer ces probabilités à partir de la fonction.
- Si le k-ième moment de X existe, on a :

$$\mathbb{E}(X(X-1)\dots(X-k)) \;=\; \left.\frac{d^n}{dz^n}F_X(z)\right|_{z=1}\ .$$

  Ce nombre s'appelle le k-ième *moment factoriel* de X. Les vrais moments s'en déduisent par combinaisons linéaires (voir : nombres de Stirling de première espèce).

– Si deux variables aléatoires X et Y sont indépendantes, alors :

$$F_{X+Y}(z) \;=\; \mathbb{E}(z^{X+Y}) \;=\; \mathbb{E}(z^X z^Y) \;=\; \mathbb{E}(z^X)\,\mathbb{E}(z^Y) \;=\; F_X(z)\,F_Y(z)\;.$$

– Plus généralement, si $X_1, X_2, \ldots, X_n$ sont n variables aléatoires de même distribution X, de fonction génératrice $F_X(z)$, et si on définit $S = \sum_{i=1}^n X_i$, alors

$$F_S(z) \;=\; \big(F_X(z)\big)^n\;.$$

– Si N est une variable aléatoire à valeurs dans $\mathbb{N}$, de fonction génératrice $F_N(z)$, si $\{X_1, X_2, \ldots, X_n, \ldots\}$ est une suite de variables aléatoires indépendantes et identiquement distribuées,[1] et on définit la *somme aléatoire*

$$S \;=\; \sum_{i=1}^{N} X_i\;, \tag{8}$$

alors on a le résultat :

$$F_S(z) \;=\; F_N\big(F_X(z)\big)\;.$$

Exemples Nous passons en revue les distributions de probabilités discrètes classiques et leurs fonctions génératrices.

La distribution de Bernoulli. La distribution de Bernoulli est donnée par :

$$\mathbb{P}(X=0) \;=\; p\;, \quad \mathbb{P}(X=1) \;=\; q \;:= 1-p\;.$$

La fonction génératrice est donc :

$$F_X(z) \;=\; p + qz\;.$$

La distribution Géométrique. La distribution géométrique sur $\mathbb{N}$ de paramètre p est donnée par :

$$\mathbb{P}(X=k) \;=\; (1-p)\,p^k\;, \quad k\in\mathbb{N}\;.$$

La fonction génératrice est donc (voir l'Identité Géométrique au paragraphe 2.6) :

$$F_X(z) \;=\; \frac{1-p}{1-pz}\;.$$

Il existe une variante : la distribution géométrique sur $\{1, 2, \ldots\}$, donnée par :

$$\mathbb{P}(X=k) \;=\; (1-p)\,p^{k-1}\;, \quad k\in\mathbb{N}\,, k\geq 1\;.$$

La fonction génératrice de cette variante est :

$$F_X(z) \;=\; \frac{(1-p)z}{1-pz}\;.$$

La distribution Binomiale. La distribution Binomiale de paramètres n et p est donnée par :

$$\left(\begin{matrix} \mathbb{P}(X=k) \;= \\ k, np^k q^{n-k}\;, \quad k\in\{0,1,\ldots,n\}\;, \end{matrix}\right)$$

avec $q := 1-p$. Sa fonction génératrice est donc :

$$F_X(z) \;=\; (p+qz)^n\;.$$

[1] On emploie l'abbréviation v.a.i.i.d. pour « variables aléatoires indépendantes et identiquement distribuées ».

La distribution de Poisson. La distribution de Poisson de paramètres λ est donnée par :

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

Sa fonction génératrice est donc (voir l'Identité Exponentielle au paragraphe 2.6) :

$$F_X(z) \; = \; e^{\lambda(z-1)} \; .$$

La distribution Uniforme. La distribution uniforme sur $\{a, a+1, \ldots, b-1, b\}$ est donnée par :

$$\mathbb{P}(X = k) \; = \; \frac{1}{b-a+1} \; , \quad k \in \{a, a+1, \ldots, b-1, b\} \; .$$

Sa fonction génératrice est :

$$F_X(z) \; = \; \frac{z^a}{b-a+1} \frac{1-z^{b-a+1}}{1-z} \; .$$

2.5 Inversion d'une fonction génératrice

Soit une fonction génératrice (de probabilités ou non) :

$$f(z) \; = \; \sum_{n=0}^{\infty} f_n z^n \; .$$

Le nombre f_n est le coefficient de z^n dans le développement en série de $f(z)$. On utilise la notation :

$$f_n \; = \; [z^n] f(z) \; .$$

Si on connaît une formule pour $f(z)$, il est possible en principe de calculer tous les f_n. On décrit ici deux méthodes. Une troisième est décrite dans l'Appendice A.

Identification des coefficients. L'idée est de calculer le développement en série de $f(z)$ à partir de développements connus.

Par exemple, s'il est su que $f(z) = 2/(2-z)(3-z)$, on a :

$$f_n \; = \; [z^n] f(z) \; = \; [z^n] \left(\frac{2}{2-z} - \frac{2}{3-z} \right) \; = \; [z^n] \frac{1}{1-z/2} - [z^n] \frac{2/3}{1-z/3} \; = \; 2^{-n} - \frac{2}{3} \, 3^{-n} \; .$$

On donne au paragraphe 2.6 quelques identités bien connues qui peuvent servir de base à un tel calcul.

Formule de Taylor. La *formule de Taylor* dit que :

$$f_n \; = \; \frac{1}{n!} \left. \frac{d^n}{dz^n} F_X(z) \right|_{z=0} \; .$$

Donc si on dispose de la fonction $f(z)$, il est possible d'obtenir les f_n par dérivées successives. Cette méthode ne sert en général que pour les premières valeurs.

2.6 Identités

On aura souvent affaire aux identités suivantes.

– Identité géométrique :

$$\frac{1}{1-az} \; = \; \sum_{k=0}^{\infty} a^k z^k \; . \tag{9}$$

– Identité exponentielle :
$$e^{az} \;=\; \sum_{k=0}^{\infty} \frac{a^k z^k}{k!} \;. \tag{10}$$
– Identité binomiale :
$$(az+b)^n \;=\; \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}\; z^k \;. \tag{11}$$
– Identité des nombres de Catalan :
$$\frac{1-\sqrt{1-4z}}{2z} \;=\; \sum_{k=0}^{n} \frac{1}{k+1} \binom{2k}{k}\; z^k \;. \tag{12}$$

3 Processus de Branchement

Les processus de branchement sont une classe particulière de chaînes de Markov. Ils sont utiles dans la modélisation de procesus naturels qui peuvent être vus comme des arbres aléatoires. Ces arbres modélisent également le déroulement de certains algorithmes ou de certains programmes parallèles.

Il y a de nombreuses variantes. Dans sa version la plus simple (processus de Galton-Watson), le processus se décrit de la manière suivante.

Soit X une distribution discrète à valeurs dans $\mathbb{N}$, de moyenne m, de variance σ^2 et de fonction génératrice ϕ. On considère que dans une population, un individu peut donner naissance à un nombre aléatoire de descendants directs, distribués selon la distribution de X. Les descendances (directes) d'individus distincts sont stochastiquement indépendantes.

3.1 Distribution du nombre de descendants par génération

On suppose qu'il y a un individu unique au départ, et on dénombre ses descendants par générations.

Soit $Z(n)$ le nombre de descendants à la génération n. Nous allons calculer la fonction génératrice de $Z(n)$, par récurrence. Puisque chacun des individus i de la génération n avoir X_i descendants, la population de la génération $n+1$ est donnée par :
$$Z(n+1) \;=\; \sum_{i=1}^{Z(n)} X_i \;,$$
où les v.a. X_i sont toutes indépendantes et identiquement distribuées, de distribution X. On reconnaît une somme aléatoire comme dans (8). Par conséquent, la fonctions génératrice de $Z(n+1)$ vaut :
$$F_{Z(n+1)}(z) \;=\; \;=\; F_{Z(n)}\big(F_X(z)\big) \;.$$
Par conséquent, par récurrence :
$$F_n(z) \;=\; F_{Z(0)}\left(\underbrace{F_X \circ F_X \circ \ldots \circ F_X}_{n \text{ fois}}(z)\right) \;=\; F_{Z(0)}\left(F_X^{(\circ n)}(z)\right) \;. \tag{13}$$
Dans le cas standard, il y a un seul individu initialement, donc $Z(0)=1$ et $F_{Z(0)}(x)=x$.

En général, on ne sait pas résoudre les équations fonctionnelles récurrentes obtenues. On peut cependant en extraire des informations sur les moments (ci-après) ou de nature asymptotique (§3.4).

En dérivant (13) par rapport à z, on déduit une récurrence qui conduit aux premiers moments :
$$\mathbb{E}(Z(n)) \;=\; m^n, \qquad \mathrm{Var}(Z(n)) \;=\; \sigma^2\, \frac{m^n(m^n-1)}{m(m-1)} \;. \tag{14}$$
Si $m=1$, il faut lire : $\mathrm{Var}(Z(n)) \;=\; n\sigma^2$.

3.2 Distribution du nombre total de descendants

Soit $T(n) = \sum_{m=0}^{n} Z(m)$ le nombre total de descendants dans l'arbre jusqu'à la génération n. On s'intéresse à la distribution limite de la variable aléatoire $T(n)$ quand $n \to \infty$. Cette limite existe car la suite $T(n)$ est stochastiquement croissante. Mais elle peut être "impropre", c'est-à-dire donner une probabilité non nulle à l'infini.

En utilisant (14), il est clair que :

$$\mathbb{E}T(n) \;=\; \frac{m^{n+1}-1}{m-1} \qquad (n+1 \text{ si } m=1)\ .$$

Par conséquent, si n croît, alors :
- si $m \geq 1$, le nombre moyen de descendants tend vers l'infini, ce qui suggère que l'arbre devient infini avec une probabilité non nulle, ou bien qu'il reste fini mais avec une taille d'espérance infinie ;
- si $m < 1$, le nombre moyen de descendants converge vers $1/(1-m)$, ce qui implique que l'arbre se termine avec probabilité 1.

Soit T la variable aléatoire qui est la taille de l'arbre quand n tend vers l'infini. Comme le processus de génération est markovien, chaque descendant de la première génération donne naissance à un arbre de même distribution que l'arbre complet. Les sous-arbres engendrés sont indépendants. On a donc, en distribution :

$$T \;=_d\; 1 \;+\; \sum_{i=1}^{X} T_i\ ,$$

où les T_i sont des v.a. indépendantes de même distribution que T. Par conséquent, en passant aux fonctions génératrices :

$$T^*(z) \;:=\; \mathbb{E}(z^T) \;=\; z\ \phi(T^*(z))\ . \tag{15}$$

3.3 Exemples

Bien que l'équation (13) soit difficile à résoudre en général, on peut parfois calculer exactement la distribution de $Z(n)$.

Le cas géométrique. Si X a une distribution géométrique : $\mathbb{P}(X=k) = \rho^k(1-\rho)$, de moyenne $m = \rho/(1-\rho)$. Comme :

$$\phi(z) \;=\; \frac{1-\rho}{1-\rho z}\ ,$$

on prouve par récurrence que

$$F_n(z) \;=\; (1-\rho)\ \frac{(1-\rho)^{n-1} + a_n(1-z)}{(1-\rho)^n + b_n(1-z)}\ .$$

En exercice, trouver les suites a_n et b_n, puis vérifier que $\mathbb{E}Z(n) = m^n$.

D'après (15), la fonction génératrice du nombre total de descendants est solution de :

$$T^*(z) \;=\; z\ \frac{1-\rho}{1-\rho T^*(z)}\ .$$

Si $m \leq 1$ (c'est-à-dire $\rho \leq 1/2$), il existe une unique solution de cette équation qui soit une fonction analytique dans le disque unité :

$$T^*(z) \;=\; \frac{1-\sqrt{1-4\rho(1-\rho)z}}{2\rho}\ .$$

En exercice : inverser cette fonction génératrice pour trouver $\mathbb{P}(T=k)$.

Le cas des chaînes. Si par exemple X vaut 1 avec proba m et 0 avec proba $1-m$, ce qui correspond à $\phi(z) = 1 - m + mz$, alors :

$$\begin{aligned} F_2(z) &= (1 - m + m(1 - m + mz)) = 1 - m^2 + m^2 z \\ F_3(z) &= (1 - m + m(1 - m^2 + m^2 z)) = 1 - m^3 + m^3 z \\ F_n(z) &= 1 - m^n + m^n z, \quad n \geq 0. \end{aligned}$$

Pour la taille totale de la chaîne, on trouve que :

$$T^*(z) = \frac{1-m}{1-mz} .$$

Dans ce cas, on retrouve que l'"arbre" est une chaîne de longueur géométriquement distribuée.

Arbres binaires. Si enfin X vaut 2 ou 0 avec probas $m/2$ et $1 - m/2$, on a :

$$F_n(z) = 1 - \frac{m}{2} + \frac{m}{2} F_{n-1}(z)^2 .$$

Quand $m < 1$, la solution de (15) s'écrit ici :

$$\begin{aligned} T^*(z) &= \frac{1 - \sqrt{1 - m(2-m)z^2}}{mz} \\ &= \sum_{k=0}^{\infty} \frac{1}{k+1} \binom{2k}{k} \left(\frac{m}{2}\right)^k \left(1 - \frac{m}{2}\right)^{k+1} z^{2k+1} . \end{aligned}$$

Cette formule était à prévoir car il est clair que la probabilité que l'arbre ait k nœuds internes, et donc $k+1$ feuilles (soit en tout $2k+1$ nœuds) est la même quelle que soit sa forme, et vaut :

$$\mathbb{P}(k \text{ nœuds avec deux descendants, } k+1 \text{ sans descendants}) = \left(\frac{m}{2}\right)^k \left(1 - \frac{m}{2}\right)^{k+1} .$$

Or il est bien connu qu'il y a

$$\frac{1}{1+k} \binom{2k}{k}$$

tels arbres (nombre de Catalan).

Observer que quand $m = 1$, la distribution de T est propre et sa fonction génératrice est :

$$T^*(z) = \frac{1 - \sqrt{1 - z^2}}{z} .$$

Cette variable aléatoire a une espérance infinie.

3.4 Propriétés asymptotiques

La première question que l'on se pose sur un arbre aléatoire est celle de sa finitude.

La *probabilité d'extinction* de l'arbre, π_e est la probabilité qu'il existe un n tel que $Z(n) = 0$. Clairement, si $Z(n) = 0$ alors $Z(m) = 0$ pour tout $m \geq n$.

L'événement $\{\exists n | Z(n) = 0\}$ est égal à $\cup_n \{Z(n) = 0\}$. Comme les événements $\{Z(n) = 0\}$ sont inclus les uns dans les autres, on a :

$$\pi_e = \lim_{n \to \infty} \mathbb{P}(Z(n) = 0) .$$

La probabilité $\mathbb{P}(Z(n) = 0)$ vaut, d'après (13),

$$\pi_e = \lim_{n \to \infty} \phi^{(n)}(0) . \qquad (16)$$

On a alors le résultat :

Théorème 7. *Si $m \leq 1$, alors $\pi_e = 1$ et le processus de branchement s'éteint presque sûrement. Si $m > 1$, alors la probabilité d'extinction est la plus petite solution x de l'équation*

$$\phi(x) \; = \; x \; .$$

Démonstration. La fonction ϕ est continue, croissante et convexe sur l'intervalle $[0, 1]$, et $\phi(1) = 1$. Deux cas se présentent. Soit $\phi(0) = 0$. Dans ce cas, chaque individu a toujours au moins un descendant et l'arbre ne peut pas s'éteindre. Dans ce cas, on a forcément $m \geq 1$. Dans l'autre cas, $\phi(0) > 0$. Puisque ϕ est croissante, on a $\phi^{(2)}(0) > \phi(0)$ et la suite $\phi^{(n)}(0)$ est donc croissante et bornée par 1. Elle converge donc vers le plus petit point fixe de ϕ qui soit > 0. Si $m = \phi'(0) < 1$, alors $\phi(x) < x$ au voisinage de $x = 1$, donc il existe un point fixe de ϕ dans $]0, 1[$. Par contre, si $m \geq 1$, comme ϕ est convexe, le plus petit point fixe de ϕ est 1. □

Exemple : dans le cas où X vaut 2 ou 0 avec probas $m/2$ et $1-m/2$, on trouve que $\pi_e = 1-m/2+(m/2)\pi_e^2$, d'où :

$$\pi_e \; = \; \frac{2-m}{m} \; ,$$

dans le cas $m \geq 1$.

4 La marche aléatoire en dimension 1

Dans ce paragraphe, on analyse un système dynamique aléatoire central : la marche aléatoire (ou « promenade de l'ivrogne ») sur l'ensemble des entiers. La description du processus est la suivante : à chaque unité de temps :

- ou bien il diminue de 1 avec probabilité p,
- ou bien le processus augmente de 1, avec probabilité $q = 1 - p$.

Dans le cas ou le processus se trouve dans l'état 0, il ne peut plus diminuer. On supposera qu'à l'étape suivante, le processus « repart » dans un état aléatoire.

La définition plus formelle de la suite de variables aléatoires $\{X_0, X_1, \ldots, X_n, \ldots\}$ est donc :

- X_0 a une distribution donnée ;
- pour tout n, si $X_n \geq 1$, alors :
  - $X_{n+1} = X_n + 1$ avec probabilité q,
  - $X_{n+1} = X_n - 1$ avec probabilité p,
- et si $X_n = 0$, alors $X_{n+1} = V_n$, où $\{V_0, \ldots, V_n, \ldots\}$ est une suite de v.a.i.i.d. de distribution V donnée.

Ce processus appartient à la famille des *processus de naissance et de mort*, ainsi nommée parce que l'état ne change que par ajout de 1 (naissance) ou retrait de 1 (mort). Dans le cas où l'état atteint 0, la variable aléatoire V spécifie comment le processus se « régénère ».

Nous allons nous intéresser à diverses propriétés de ce processus.

4.1 Distribution de X_n

Nous montrons dans ce paragraphe comment aborder le calcul de la distribution de X_n par le moyen des fonctions génératrices.

L'objectif est de calculer les fonctions $F_{X_n}(z) = \mathbb{E}(z^{X_n})$ pour tout n. Nous procédons par récurrence. La fonction $F_{X_0}(z)$ est connue car la distribution de X_0 est une donnée.

Pour le calcul, il sera pratique de noter B_n une suite de v.a.i.i.d. telles que $\mathbb{P}(B_n = 1) = q$ et $\mathbb{P}(B_n = -1) = p$. Ainsi, si $X_n > 0$ on aura $X_{n+1} = X_n + B_n$.

$$
\begin{aligned}
F_{X_{n+1}}(z) \;=\; \mathbb{E}(z^{X_{n+1}}) &= \mathbb{E}(z^{X_{n+1}}|X_n = 0)\mathbb{P}(X_n = 0) \;+\; \mathbb{E}(z^{X_{n+1}}|X_n > 0)\mathbb{P}(X_n > 0) \\
&= \mathbb{E}(z^{V_n})\mathbb{P}(X_n = 0) \;+\; \mathbb{E}(z^{X_n + B_n})\mathbb{P}(X_n > 0) \\
&= F_V(z)\mathbb{P}(X_n = 0) \;+\; \mathbb{E}(z^{X_n} z^{B_n}|X_n > 0)\mathbb{P}(X_n > 0) \\
&= F_V(z)\mathbb{P}(X_n = 0) \;+\; \mathbb{E}(z^{B_n})\mathbb{E}(z^{X_n}|X_n > 0)\mathbb{P}(X_n > 0) \\
&= F_V(z)\mathbb{P}(X_n = 0) \;+\; F_B(z)\left(\mathbb{E}(z^{X_n}) - \mathbb{P}(X_n = 0)\right) \\
&= F_B(z)\; F_{X_n}(z) \;+\; \mathbb{P}(X_n = 0)\; (F_V(z) - F_B(z)) \; . \qquad (17)
\end{aligned}
$$

Cette relation donne un moyen algorithmique de calculer, par récurrence, les fonctions $F_{X_n}(z)$ successives. En effet, si on connaît F_{X_n}, on connaît en particulier $\mathbb{P}(X_n = 0) = F_{X_n}(0)$. On en déduit par (17) l'expression de $F_{X_{n+1}}(z)$.

4.2 Solution de la récurrence sur X_n

Pour aller plus loin et obtenir des expressions explicites pour les probabilités qui nous intéressent, nous résolvons cette récurrence de fonctions, à nouveau grâce aux fonctions génératrices.

Définissons la fonction génératrice (bidimensionnelle) :

$$
\mathcal{F}(u, z) \;=\; \sum_{n=0}^{\infty} u^n F_{X_n}(z) \; . \qquad (18)
$$

À partir de l'équation (17), qu'on multiplie par u^n puis qu'on somme pour n de 0 à l'infini, on obtient : pour le membre gauche,

$$
\begin{aligned}
\sum_{n=0}^{\infty} u^n F_{X_{n+1}}(z) &= \sum_{m=1}^{\infty} u^{m-1} F_{X_m}(z) \\
&= \frac{1}{u} \sum_{m=1}^{\infty} u^m F_{X_m}(z) \\
&= \frac{1}{u} \left(\sum_{m=0}^{\infty} u^m F_{X_m}(z) - F_{X_0}(z) \right) \\
&= \frac{1}{u} \left(\mathcal{F}(u, z) - F_{X_0}(z) \right) \; . \qquad (19)
\end{aligned}
$$

Pour le membre droit, on a :

$$
\begin{aligned}
&\sum_{n=0}^{\infty} u^n \left(F_B(z)\; F_{X_n}(z) \;+\; \mathbb{P}(X_n = 0)\; (F_V(z) - F_B(z)) \right) \\
&\quad = F_B(z) \sum_{n=0}^{\infty} u^n F_{X_n}(z) \;+\; (F_V(z) - F_B(z)) \sum_{n=0}^{\infty} u^n \mathbb{P}(X_n = 0) \\
&\quad = F_B(z)\; \mathcal{F}(u, z) \;+\; (F_V(z) - F_B(z)) \mathcal{F}(u, 0) \; . \qquad (20)
\end{aligned}
$$

On s'est servi du fait que $F_{X_n}(0) = \mathbb{P}(X_n = 0)$ ce qui permet de voir que $\mathcal{F}(u, 0) = \sum_n u^n \mathbb{P}(X_n = 0)$. En regroupant (19) et (20), on obtient :

$$
\frac{1}{u} \left(\mathcal{F}(u, z) - F_{X_0}(z) \right) \;=\; F_B(z)\; \mathcal{F}(u, z) \;+\; (F_V(z) - F_B(z)) \mathcal{F}(u, 0) \; ,
$$

qui se ré-écrit en :

$$
\mathcal{F}(u, z)(1 - u F_B(z)) \;=\; F_{X_0}(z) \;+\; u (F_V(z) - F_B(z)) \mathcal{F}(u, 0) \; . \qquad (21)
$$

D'après la définition de la variable aléatoire B, on a : $F_B(z) = qz + p/z$. En remplaçant dans (21) on obtient que $\mathcal{F}$ vérifie l'équation fonctionnelle suivante :

$$\mathcal{F}(u,z)(z - pu - quz^2) \;=\; zF_{X_0}(z) \;+\; u(zF_V(z) - p - qz^2)\mathcal{F}(u,0) \;. \tag{22}$$

Dans cette équation, la fonction $F_{X_0}(z)$ est connue comme nous l'avons vu, mais $\mathcal{F}(u,0)$ est une *fonction inconnue*. Dans le membre droit, le polynôme :

$$K(u,z) \;=\; -quz^2 - pu + z \tag{23}$$

joue un rôle déterminant. Ce polynôme est appellé le *noyau* de l'équation. La méthode de résolution consiste à remarquer que si ce polynôme s'annule, alors le membre droit s'annule et le membre gauche doit être nul. Ceci donne une équation qui doit permettre de déterminer la fonction inconnue $\mathcal{F}(u,0)$.

On commence donc par résoudre $K(u,z) = 0$ en z. On obtient deux fonctions

$$z_1(u) \;=\; \frac{1}{2qu}\left(1 - \sqrt{1-4pqu^2}\right), \qquad z_2(u) \;=\; \frac{1}{2qu}\left(1 + \sqrt{1-4pqu^2}\right), \tag{24}$$

qui sont telles que $K(u, x_1(u)) = K(u, x_2(u)) = 0$ pour tout nombre complexe $u \neq 0$.

En remplaçant z par $z_1(u)$ dans (22), on obtient que :

$$0 \;=\; z_1(u)F_{X_0}(z_1(u)) \;+\; u(z_1(u)F_V(z_1(u)) - p - qz_1(u)^2)\mathcal{F}(u,0) \;.$$

Cette expression peut se simplifier parce que, par définition de $z_1(u)$, on a : $pu + quz_1(u)^2 = z_1(u)$. En fin de compte, il reste :

$$0 \;=\; z_1(u)F_{X_0}(z_1(u)) \;+\; z_1(u)(uF_V(z_1(u)) - 1)\mathcal{F}(u,0) \;,$$

d'où on peut extraire

$$\mathcal{F}(u,0) \;=\; \frac{F_{X_0}(z_1(u))}{1 - uF_V(z_1(u))} \;. \tag{25}$$

Il ne reste qu'à remplacer dans (22) pour trouver la valeur de $\mathcal{F}$:

$$\mathcal{F}(u,z) \;=\; \frac{1}{z - pu - quz^2}\left(zF_{X_0}(z) \;+\; F_{X_0}(z_1(u))u\frac{zF_V(z) - p - qz^2}{1 - uF_V(z_1(u))}\right) \;. \tag{26}$$

4.3 Probabilités stationnaires

Partie non donnée en cours, à étudier en s'inspirant du paragraphe 6.2.

On s'intéresse maintenant à ce qui se passe quand le nombre d'étapes n devient grand. Il est possible que la suite de probabilités

$$\mathbb{P}(X_0 = k),\ \mathbb{P}(X_1 = k),\ \ldots, \mathbb{P}(X_n = k), \ldots$$

converge vers une limite que nous noterons p_k. Si c'est bien le cas, alors, quand n tend vers l'infini, la distribution de X_n et la distribution de X_{n+1} sont la même. Cette même distribution est qualifiée de *distribution stationnaire*.

On cherche maintenant à calculer cette distribution. Comme d'habitude, notons $F_X(z)$ la fonction génératrice de la distribution stationnaire X. Si on considère à nouveau l'équation (17) et que l'on y fait $n \to \infty$, on obtient l'identité :

$$F_X(z) \;=\; F_B(z)\; F_X(z) \;+\; p_0\;(F_V(z) - F_B(z)) \;. \tag{27}$$

En résolvant cette équation, nous obtenons que la fonction recherchée est donnée par :

$$F_X(z) \;=\; p_0\,\frac{F_V(z) - F_B(z)}{1 - F_B(z)} \;. \tag{28}$$

Mais il reste une quantité inconnue dans le membre droit : p_0. Pour la déterminer, nous nous souvenons que la valeur d'une fonction génératrice de probabilités en la valeur 1 est toujours 1 : donc en principe, $F_X(1) = 1$. En remplaçant z par 1 dans (28), nous rencontrons un problème avec le membre droit : son numérateur et son dénominateur sont nuls quand $z = 1$. Nous sommes en présence d'une indétermination. Pour la lever, on utilise la « règle de l'Hôpital » qui s'énonce comme suit : si $f(x)$ et $g(x)$ sont deux fonctions dérivables qui sont nulles quand $x = x_0$, alors

$$\lim_{x \to x_0} \frac{f(x_0)}{g(x_0)} \; = \; \frac{f'(x_0)}{g'(x_0)} \; . \tag{29}$$

Dans le cas qui nous occupe, nous faisons tendre z vers 1 dans l'équation (28) et nous obtenons :

$$1 \; = \; p_0 \, \frac{(F_V)'(1) - (F_B)'(1)}{-(F_B)'(1)} \; .$$

Mais nous savons aussi que la dérivée d'une FGP en 1 vaut l'espérance de la variable aléatoire. Nous avons donc en fait la relation :

$$1 \; = \; p_0 \, \frac{\mathbb{E}B - \mathbb{E}V}{\mathbb{E}B} \; = \; p_0 \left(1 - \frac{\mathbb{E}V}{\mathbb{E}B}\right) \; ,$$

d'où il est facile de déduire p_0 :

$$p_0 \; = \; \frac{\mathbb{E}B}{\mathbb{E}B - \mathbb{E}V} \; . \tag{30}$$

Le calcul se conclut en remplaçant cette valeur dans l'équation (28), ce qui nous conduit à la valeur de la fonction génératrice de la distribution stationnaire :

$$F_X(z) \; = \; \frac{\mathbb{E}B}{\mathbb{E}B - \mathbb{E}V} \, \frac{F_V(z) - F_B(z)}{1 - F_B(z)} \; . \tag{31}$$

Discussion. Le calcul qui précède prédit dans l'équation (30) la valeur que doit avoir p_0, qui est la probabilité que X soit égal à 0. Cette valeur doit être positive et inférieure à 1. Nous savons que $\mathbb{E}V \geq 0$. D'autre part,

$$\mathbb{E}B \; = \; p \times (-1) \; + \; q \times 1 \; = \; q - p \; = \; 1 - 2p \; .$$

Cette valeur peut être positive ou négative, selon la valeur de p. Or pour que le membre droit de (30) soit positif, il faut que $\mathbb{E}V < \mathbb{E}B$. Ceci conduit à une discussion avec différents cas.

$\mathbb{E}B < 0$ Dans ce cas, la formule (30) donne bien une valeur qui est une probabilité. Le calcul est valide. La distribution stationnaire X existe et sa fonction génératrice est donnée par l'équation (31)

$\mathbb{E}B > 0$ Dans ce cas, la formule (30) donne une valeur qui est soit négative, soit plus grande que 1. Le calcul ne marche pas. C'est parce que la distribution stationnaire X *n'existe pas* ! En effet, dans ce cas la marche aléatoire a tendance à monter sans limite, puisque la probabilité de faire un pas vers le haut est supérieure à la probabilité de faire un pas vers le bas.

$\mathbb{E}B = 0$ Dans ce cas, le calcul ne marche pas : il prédit une valeur de p_0 nulle, et une fonction génératrice nulle aussi, ce qui est absurde. La distribution stationnaire n'existe pas non plus.

Exercice 1. Étudier la situation particulière où :

$$V = 0 \text{ avec probabilité } p, \qquad V = 1 \text{ avec probabilité } q.$$

Montrer que la distribution stationnaire est géométrique. Sous quelle condition sur p cette distribution existe-t-elle ?

4.4 Marche aléatoire et nombres de Catalan

Dans cet paragraphe, nous voyons comment faire pour obtenir les valeurs de probabilités à partir de leur fonction génératrice. Nous obtenons une formule où figurent les nombres de Catalan. Nous faisons alors un raisonnement combinatoire pour expliquer cette formule.

Nous reprenons le modèle de marche aléatoire, dans le cas particulier où la valeur de V est toujours 0. Cela revient à dire que quand la marche aléatoire atteint la valeur 0, elle reste éternellement en 0. Supposons de plus que $X_0 = 1$. Traduit en fonctions génératrices, nous avons :

$$F_V(z) \ = \ 1 \ , \qquad F_{X_0}(z) \ = \ z \ .$$

Par conséquent, en remplaçant ces valeurs dans la formule (25), nous obtenons que :

$$\mathcal{F}(u,0) \ = \ \sum_{n=0}^{\infty} u^n \ \mathbb{P}(X_n = 0) \ = \ \frac{z_1(u)}{1-u} \ . \tag{32}$$

Nous allons développer cette dernière formule en série de la variable u, en utilisant les identités du paragraphe 2.6. Nous avons, en particulier :

$$\frac{1}{1-u} \ = \ \sum_{m=0}^{\infty} u^m \ , \tag{33}$$

et

$$\begin{aligned} z_1(u) \ &= \ \frac{1}{2qu}\left(1-\sqrt{1-4pqu^2}\right) \ = \ \frac{pu}{2pqu^2}\left(1-\sqrt{1-4pqu^2}\right) \ = \ \frac{pu}{2z} \sum_{k=0}^{\infty} \frac{1}{k+1}\binom{2k}{k} \left(pqu^2\right)^k \\ &= \ \sum_{k=0}^{\infty} \frac{1}{k+1}\binom{2k}{k} \ p^{k+1}q^k u^{2k+1} \ . \end{aligned} \tag{34}$$

En réunissant (33) et (34) dans (32), on obtient :

$$\begin{aligned} \sum_{n=0}^{\infty} u^n \ \mathbb{P}(X_n = 0) \ &= \ \sum_{m=0}^{\infty}\sum_{k=0}^{\infty} u^{m+2k+1} \ \frac{1}{k+1}\binom{2k}{k} \ p^{k+1}q^k \\ &= \ \sum_{\ell=0}^{\infty} \sum_{k=0}^{(\ell-1)/2} u^\ell \ \frac{1}{k+1}\binom{2k}{k} \ p^{k+1}q^k \ . \end{aligned}$$

En *identifiant* les coefficients de u^n dans le membre gauche et le membre droit, nous trouvons :

$$\mathbb{P}(X_n = 0) \ = \ \sum_{k=0}^{(n-1)/2} \frac{1}{k+1}\binom{2k}{k} \ p^{k+1}q^k \ . \tag{35}$$

Passons à l'interprétation combinatoire. Parmi les chemins qui conduisent de l'état initial $X_0 = 1$ à $X_n = 0$, nommons $\mathcal{C}_p$ l'ensemble de ceux qui prennent la valeur 0 pour la première fois à l'étape p. On voit facilement que p est forcément impair et donc il s'écrit $p = 2k+1$. Un chemin de l'ensemble $\mathcal{C}_{2k+1}$ a la probabilité $p^{k+1}q^k$ de se produire, car il est constitué de k pas « vers le haut » et $k+1$ « vers le bas ». Il se trouve que le nombre de tels chemins est exactement $\frac{1}{k+1}\binom{2k}{k}$. Par conséquent, on retrouve la formule (35) par le raisonnement suivant :

$$\begin{aligned}
\mathbb{P}(X_n = 0) &= \mathbb{P}(\text{le chemin va de } X_0 = 1 \text{ à } X_n = 0) \\
&= \sum_{k=0}^{(n-1)/2} \mathbb{P}(\text{le chemin est dans } \mathcal{C}_{2k+1}) \\
&= \sum_{k=0}^{(n-1)/2} p^{k+1} q^k \, \mathrm{card}(\mathcal{C}_{2k+1}) \\
&= \sum_{k=0}^{(n-1)/2} p^{k+1} q^k \, \frac{1}{k+1} \binom{2k}{k} .
\end{aligned}$$

5 Enumération exacte des graphes aléatoires

Nous nous intéressons aux méthodes d'énumérations des graphes aléatoires. Ces méthodes font intervenir les fonctions génératrices ainsi que les opérations combinatoires sur celles-ci. Nous nous intéressons dans un premier temps aux méthodes d'énumérations exactes des graphes aléatoires. Si on considère une composante connexe d'un graphe aléatoire, plus son excès est grand plus son étude devient difficile. Nous allons privilégier une étude basée sur l'excès des graphes aléatoires. Nous commençons ainsi par présenter les résultats attribués à Cayley sur le nombre d'arbres enracinés, puis les résultats de Rényi [14].

Dans cette section, nous allons étudier les graphes aléatoires à excès k et les énumérer de manière exacte. Nous allons voir que les graphes unicycliques et multicycliques peuvent s'exprimer assez simplement grâce à la fonction génératrice exponentielle des arbres enracinés dit de *Cayley*.

5.1 Les arbres

Les arbres sont des graphes connexes sans cycle. Ils sont aussi définis comme étant des *graphes acycliques* . Un arbre de n sommets contient $(n-1)$ arêtes ; son excès est par définition -1. Deux types d'arbres existent, les arbres enracinés non ordonnés appelé de *arbres de Cayley* et les arbres non enracinés. Pour les arbres enracinés, on distingue un sommet en le désignant comme étant la racine.

5.1.1 Les arbres enracinés de Cayley

Chaque arbre enraciné T d'étiquettes $< 1, \ldots, n >$ peut être vu comme une paire ordonnée $< a, b >$, où a est la racine de l'arbre et b est un ensemble d'arbres enracinés (les enfants de la racine) voir figure 3.

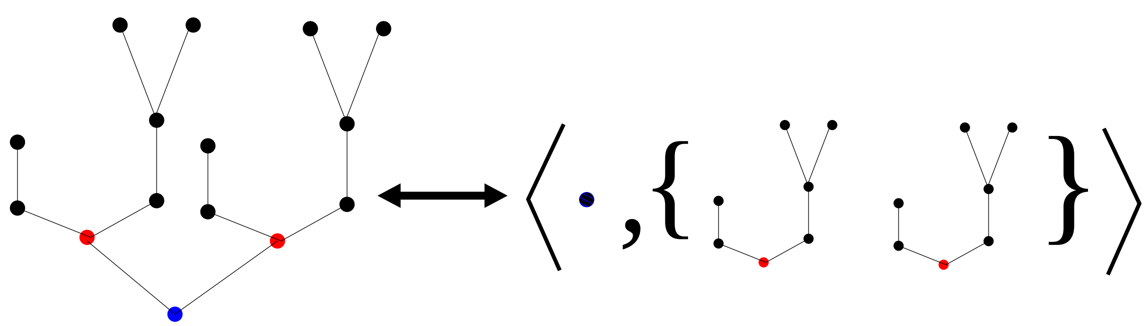

FIG. 3 – Représentation d'un arbre enraciné

Notons par $\mathcal{T}$ l'ensemble des arbres enracinés étiquetés et par $\mathcal{N}$ l'ensemble des sommets isolés étiquetés. L'équation combinatoire décrivant l'ensemble des arbres enracinés est alors :

$$\mathcal{T} = \mathcal{N} \star Set\{\mathcal{T}\}.$$

La fonction génératrice pour $\mathcal{N}$ est z, et la fonction génératrice pour $\mathcal{T}$ est $T(z)$. En appliquent la méthode symbolique (décrite dans le théorème 6) à cette description combinatoire, nous obtenons l'équation fonctionnelle suivante :

$$T(z) = z \exp T(z). \tag{36}$$

Le développement donné par l'équation (36) a été observé par Einseinstein.

Les équations implicites du type $F(z) = z\Phi(F(z))$ sont très importantes dans la théorie combinatoire et dans la théorie de l'analyse d'algorithmes [5]. Elles peuvent être résolues grâce aux théorèmes d'inversion de *Lagrange, Lagrange-Bürmann* .

Théorème 8 (Inversion de Lagrange, Lagrange-Bürmann)**.** *Soit $\Phi(u) = \sum_{j=0}^{\infty} \Phi_j u^j$ une fonction génératrice telle que $\Phi_0 \neq 0$ et soit $F(z)$ l'unique fonction génératrice solution de l'équation $F(z) = z\Phi(F(z))$. Nous avons alors :*

$$\begin{aligned} i) \qquad [z^n]\,F(z) &= \frac{1}{n}\left[u^{n-1}\right](\Phi(u))^n\,, \\ ii) \qquad [z^n]\,G(F(z)) &= \frac{1}{n}\left[u^{n-1}\right](\Phi(u))^n\,G'(u)\,, \end{aligned} \tag{37}$$

où G est une fonction génératrice arbitraire.

Démonstration. Par la formule intégrale de *Cauchy*, nous avons :

$$[z^n]\,F(z) = \frac{1}{2i\pi}\oint F(z)\frac{dz}{z^{n+1}}.$$

Posons $F(z) = y$, d'où $y = z\Phi(y)$ et $dz = \frac{\Phi(y) - y\Phi'(y)}{(\Phi(y))^2}dy$. Nous avons

$$\begin{aligned} [z^n]\,F(z) &= \frac{1}{2i\pi}\oint y\left(\frac{\Phi(y)}{y}\right)^{n+1}(\Phi(y) - yPhi'(y))\frac{dy}{(\Phi(y))^2} \\ &= \frac{1}{2i\pi}\oint (\Phi(y))^n\frac{dy}{y^n} - \frac{1}{2i\pi}\oint \Phi'(y)\frac{(\Phi(y))^{n-1}\,dy}{y^{n-1}}. \end{aligned}$$

Comme :

$$\begin{aligned} \frac{1}{2i\pi}\oint \Phi'(y)\frac{(\Phi(y))^{n-1}\,dy}{y^{n-1}} &= \left[y^{n-2}\right]\Phi'(y)\,(\Phi(y))^{n-1} \\ &= \frac{n-1}{n}\left[y^{n-1}\right](\Phi(y))^n \end{aligned}$$

Nous obtenons

$$\begin{aligned} [z^n]\,F(z) &= \left[y^{n-1}\right]\left((\Phi(y))^n - \frac{n-1}{n}(\Phi(y))^n\right) \\ &= \frac{1}{n}\left[y^{n-1}\right](\Phi(y))^n\,. \end{aligned}$$

La démonstration permettant de vérifie ii) utilise les mêmes arguments analytiques. □

En utilisant les théorèmes d'inversion de *Lagrange* et *Lagrange-Bürmann*, nous trouvons la valeur de T_n dans la série génératrice des arbres enracinés étiquetés :

$$T(z) = \sum_{n=1}^{\infty} T_n z^n \, .$$

Nous avons alors :

$$\begin{aligned} T_n &= n! \, [z^n] \, T(z) \\ &= n! \, \frac{1}{n} \left[u^{n-1}\right] (\exp(u))^n \\ &= n^{n-1} \, . \end{aligned} \tag{38}$$

La fonction génératrice des arbres non enracinés est donc :

$$T(z) = \sum_{n=1}^{\infty} n^{n-1} z^n \, . \tag{39}$$

Cette formule est communément attribuée à Cayley et le nom *d'arbres de Cayley* est donné aux arbres enracinés étiquetés et non ordonnés .

Un arbre de n sommets possède $n-1$ arêtes, nous pouvons par conséquent obtenir la fonction génératrice exponentielle bivariée des arbres de Cayley, en marquant par w les arêtes (z marque les sommets) :

$$T(w,z) = \sum_{n=1}^{\infty} (wn)^{n-1} \, \frac{z^n}{n!} \, . \tag{40}$$

5.1.2 Les arbres non enracinés

Intéressons nous, maintenant, aux arbres non enracinés. Notons par $U(z)$, la fonction génératrice exponentielle qui énumère les arbres étiquetés non enracinés. Un arbre T étiqueté de 1 à n peut être représenté soit comme un arbre non enraciné U si 1 est la racine de T, soit comme une paire ordonnée $< a, b >$ où a et b sont deux arbres enracinés, sinon [2]. Nous obtenons alors :

$$U(z) = T(z) - \frac{1}{2} \left(T(z)\right)^2 \, . \tag{41}$$

Nous pouvons aussi aboutir à ce résultat en utilisant l'opérateur de pointage défini en 2.3.2. Si nous *distinguons* un sommet dans un arbre non enraciné, nous obtenons un arbre enraciné :

$$\vartheta_z U(z) = T(z) \, .$$

Un calcul intégral nous donne :

$$U(z) = \int_0^z \frac{T(z)}{t} dt \, .$$

La fonction génératrice exponentielle est donc :

$$U(z) = z + \sum_{n=2}^{\infty} (n)^{n-2} \, \frac{z^n}{n!} \, .$$

Il existe, par conséquent, n^{n-2} arbres de Cayley non enracinés.

En outre, en tenant compte des arêtes, la fonction génératrice exponentielle *bivariée* $U(w,z)$ des arbres non-enracinés, s'exprime en fonction de w et de $T(wz)$. Elle vérifie :

$$U(w,z) = \frac{U(wz)}{w} = \frac{T(wz)}{w} - \frac{\left(T(wz)\right)^2}{2w} \, . \tag{42}$$

5.2 Les graphes unicycliques

Par définition, les graphes unicycliques ont n sommets et n arêtes.

Lemme 9 ([14])**.** *Soit $n \geq 3$, un (n,n)-graphe connexe contient exactement un cycle comportant au moins trois sommets.*

Démonstration. Un graphe d'excès -1 ayant n sommets possède exactement $(n-1)$ arêtes. Pour former un cycle dans un graphe connexe, il faut ajouter une arête à un graphe d'excès -1. Cette arête ne doit ni former une *boucle*, ni former une *arête multiple* ; le plus petit cycle formé sera alors de longueur supérieure ou égale à trois. □

Une composante unicyclique peut être considérée comme un cycle de k arbres enracinés pour $k \geq 3$ voir figure 4.

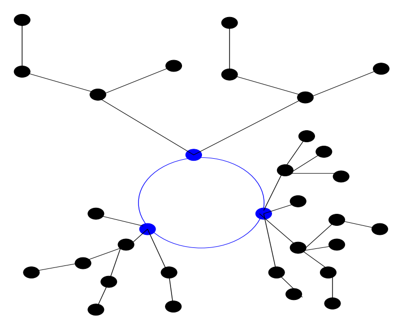

FIG. 4 – Représentation d'un graphe unicyclique comme un cycle de trois arbres

En utilisant le théorème 6, et en notant $U(z)$ la fonction génératrice des graphes unicycliques connexes, nous avons :

$$V(z) = \frac{1}{2}\ln\left(\frac{1}{1-T(z)}\right) - \frac{T(z)}{2} - \frac{T(z)^2}{4}\,. \tag{43}$$

La fonction génératrice univariée pour un graphe unicyclique de longueur k est

$$\frac{T(z)^k}{2k}\,,$$

le terme $2k$ compte l'ordre du cycle et le changement d'orientation du cycle. Nous pouvons de nouveau obtenir l'équation 43 en sommant sur $k \geq 3$ l'équation précédente [2].

La fonction génératrice bivariée des graphes unicycliques est alors donnée par :

$$V(w,z) = \frac{1}{2}\ln\left(\frac{1}{1-T(wz)}\right) - \frac{T(wz)}{2} - \frac{T(wz)^2}{4} \tag{44}$$

où $T(z)$ est la FGE des arbres enracinés de Cayley et $T(wz) = wT(w,z)$).

6 Files d'attente

Considérons une file d'attente en temps discret, dont le fonctionnement est le suivant :

- Les clients ont une durée de service égale à une unité de temps ; à chaque unité, un des clients présents est donc servi.
- Les clients arrivent en groupes à chaque unité de temps. La taille du groupe à l'étape n est une variable aléatoire A_n.

– La taille de la salle d'attente est illimitée : on peut y stocker une nombre arbitrairement grand de clients.

Nous supposerons que la suite $\{A_0, A_1, \ldots, A_n, \ldots\}$ est i.i.d. et que la distribution commune est donnée par :

$$\mathbb{P}(A = k) \;=\; a_k \;.$$

Pour étudier l'évolution de la longueur de la file d'attente, définissons Q_n comme le nombre de clients à l'étape n, *juste après* le départ (éventuel) du client servi mais *juste avant* l'arrivée des nouveaux clients.

En faisant le bilan des clients qui partent et qui arrivent, on établit les règles d'évolution de la suite $\{Q_n; n \in \mathbb{N}\}$:

– si $Q_n > 0$, alors $Q_{n+1} = Q_n + A_n - 1$,

– et si $Q_n = 0$, alors $Q_{n+1} = \max(A_n - 1, 0)$.

Pour raccourcir certaines formules, nous utiliserons la notation $\tilde{A}_n := \max(A_n - 1, 0)$.

6.1 Récurrence sur les fonctions génératrices

Appliquons la méthode mise en œuvre pour la marche aléatoire, et cherchons une récurrence sur les fonctions génératrices F_{Q_n}. On a :

$$\begin{aligned}
F_{Q_{n+1}}(z) \;=\; \mathbb{E}(z^{Q_{n+1}}) \;&=\; \mathbb{E}(z^{Q_{n+1}}|Q_n = 0)\mathbb{P}(Q_n = 0) \;+\; \mathbb{E}(z^{Q_{n+1}}|Q_n > 0)\mathbb{P}(Q_n > 0) \\
&=\; \mathbb{E}(z^{\tilde{A}_n})\mathbb{P}(Q_n = 0) \;+\; \mathbb{E}(z^{Q_n + A_n - 1})\mathbb{P}(Q_n > 0) \\
&=\; F_{\tilde{A}}(z)\mathbb{P}(Q_n = 0) \;+\; \frac{F_A(z)}{z}\mathbb{E}(z^{Q_n}|Q_n > 0)\mathbb{P}(Q_n > 0) \\
&=\; F_{\tilde{A}}(z)\mathbb{P}(Q_n = 0) \;+\; \frac{F_A(z)}{z}\left(\mathbb{E}(z^{Q_n}) - \mathbb{P}(Q_n = 0)\right) \\
&=\; \frac{F_A(z)}{z}\,F_{Q_n}(z) \;+\; \mathbb{P}(Q_n = 0)\,\frac{1}{z}(zF_{\tilde{A}}(z) - F_A(z)) \;.
\end{aligned}$$

En fait, le calcul montre que $zF_{\tilde{A}}(z) - F_A(z) = a_0(z-1)$. Cette expression se simplifie donc en :

$$F_{Q_{n+1}}(z) \;=\; \frac{F_A(z)}{z}\,F_{Q_n}(z) \;+\; \mathbb{P}(Q_n = 0)\,\frac{a_0}{z}\,(z-1) \;. \tag{45}$$

6.2 Distribution stationnaire

Plutôt que de continuer en résolvant cette récurrence comme au paragraphe 4.2, nous nous intéressons maintenant au *cas stationnaire.*

La suite de variables aléatoires $\{Q_0, Q_1, \ldots, Q_n, \ldots\}$ est dite stationnaire si toutes les distributions des Q_n sont identiques. En particulier, la distribution de Q_n et celle de Q_{n+1} sont identiques. On l'appelle la *distribution stationnaire.*

Il n'est pas obligatoire qu'une distribution stationnaire existe. Mais supposons que c'est le cas, et examinons les conséquences. Les fonctions génératrices de Q_n et Q_{n+1} sont identiques, donc : $F_{Q_n}(z) = F_{Q_{n+1}}(z) = F_Q(z)$. En remplaçant dans (45), on trouve que :

$$F_Q(z) \;=\; \frac{F_A(z)}{z}\,F_Q(z) \;+\; \mathbb{P}(Q = 0)\,\frac{1}{z}(zF_{\tilde{A}}(z) - F_A(z)) \;.$$

Donc la fonction $F_Q(z)$ est solution de l'équation :

$$F_Q(z) \;=\; \mathbb{P}(Q = 0)\;a_0\;\frac{1-z}{A(z) - z} \;. \tag{46}$$

Pour trouver la valeur inconnue $\mathbb{P}(Q = 0)$, on peut remplacer z par 1 dans cette équation. Comme cela conduit à une indétermination (de type « 0/0 ») dans le membre droit, l'approche correcte est de faire tendre z vers 1 dans l'Équation (46). En appliquant la règle de l'Hôpital (voir Équation (29)), on trouve :

$$1 \;=\; \mathbb{P}(Q=0)\; a_0 \;\frac{-1}{A'(1)-1}\;, \qquad \text{soit donc} \qquad \mathbb{P}(Q=0) \;=\; \frac{1-\mathbb{E}A}{a_0}\;. \tag{47}$$

Cette valeur est une probabilité si et seulement si :

$$\mathbb{E}A \;\leq\; 1\;.$$

Mais le cas où $\mathbb{E}A = 1$ conduit à une valeur $\mathbb{P}(Q=0) = 0$, et donc, si on applique (46), à une fonction génératrice nulle. Ceci n'est pas possible.

On a donc les différents cas :

- Soit $\mathbb{E}A < 1$, et alors il existe effectivement une distribution stationnaire pour la suite $\{Q_n\}$, dont la fonction génératrice est :

$$P(z) \;=\; (1-\mathbb{E}A)\;\frac{1-z}{A(z)-z}\;. \tag{48}$$

- Soit $\mathbb{E}A \geq 1$ et il n'existe pas de distribution stationnaire. En effet, dans ce cas il y a trop d'arrivées de clients, et le serveur ne peut pas supporter la charge. Le nombre de clients en attente a donc tendance à s'accumuler indéfiniment.

Exercice 2. Montrer que les deux premiers moments de la distribution stationnaire Q sont donnés par :

$$\mathbb{E}Q \;=\; \frac{\mathbb{E}A^2 - \mathbb{E}A}{2(1-\mathbb{E}A)} \qquad \mathbb{E}Q^2 \;=\; \frac{3(\mathbb{E}A^2)^2 - 9\mathbb{E}A^2\mathbb{E}A + 6(\mathbb{E}A)^2 + 2\mathbb{E}A^3 - 2\mathbb{E}A^3\mathbb{E}A + 3\mathbb{E}A^2 - 3\mathbb{E}A}{6(1-\mathbb{E}A)^2}\;.$$

Exercice 3. Supposons que la taille des groupes de clients soit : 2 avec probabilité α, et 0 avec probabilité $1-\alpha$. Calculer la distribution stationnaire du nombre de clients.

Exercice 4. Supposons que la distribution de A soit géométrique de paramètre $\rho = \mathbb{P}A > 0$. Calculer la distribution stationnaire de Q, et les deux premiers moments de Q.

A Propriétés analytiques des fonctions génératrices de probabilités

Cet appendice est consacré à un complément sur les propriétés des fonctions génératices, vues comme *fonctions de la variable complexe.*

Une fonction $f(z)$, où $z \in \mathbb{C}$, est dite analytique (on dit aussi « holomorphe ») au point z_0 si elle y est dérivable. Les points où f n'est pas dérivable sont nommés *singularités.*

Le premier résultat permet de voir que certaines fonctions génératrices sont analytiques au moins dans un certain cercle.

Lemme 10. *Soit* $\{a_n, n \in \mathbb{N}\}$ *une suite de nombres complexes bornés. Alors la fonction*

$$f(z) \;=\; \sum_{n=0}^{\infty} a_n z^n$$

est analytique (on dit aussi « holomorphe ») dans le cercle unité : $\{z \in \mathbb{C} \mid |z| < 1\}$.

C'est le cas en particulier des FGP.

Le second résultat permet de trouver la singularité la plus proche de 0 pour une fonction génératrice.

Lemme 11 (Théorème de Pringsheim)**.** *Soit* $\{a_n, n \in \mathbb{N}\}$ *une suite de nombres positifs, et* $f(z) = \sum_{n=0}^{\infty} a_n z^n$. *Si le rayon de convergence de cette série est* $R < \infty$, *alors* $z = R$ *est une singularité (dite : singularité dominante).*

Le troisième résultat donne une façon de calculer les coefficients d'une série génératrice, utile dans certaines situations.

Lemme 12 (Formule de Cauchy)**.**

$$f_n \;=\; \frac{1}{2i\pi} \int_{\mathcal{C}} f(z)\, \frac{dz}{z^{n+1}} \;,$$

où $\mathcal{C}$ *est un contour simple quelconque entourant le point* $z = 0$ *dans le sens trigonométrique, et inclus dans le domaine d'analyticité de* f.

B Correction des Exercices

Corrigé de l'exercice 1

La fonction génératrice de V est alors : $F_V(z) = p + qz$. Nous savons déjà que $F_B(z) = p/z + qz$. En remplaçant dans l'équation (31), et en simplifiant, nous trouvons :

$$F_X(z) \;=\; \frac{1 - q/p}{1 - qz/p} \;.$$

À partir de l'identité (9), on obtient :

$$F_X(z) \;=\; \frac{1 - q/p}{1 - qz/p} \;=\; (1 - q/p) \sum_{k=0}^{\infty} \left(\frac{qz}{p}\right)^k \;,$$

ce qui permet de trouver :

$$\mathbb{P}(X = k) \;=\; \pi_k \;=\; \left(\frac{q}{p}\right)^k \;(1 - \frac{q}{p}) \;. \tag{49}$$

Il s'agit d'une *distribution géométrique.*

Comme nous l'avons vu lors de la discussion ci-dessous, ce calcul n'est valide que quand $q < p$, c'est-à-dire quand $p < 1/2$.

Corrigé de l'exercice 3

Dans ce cas, on a :

$$A(z) \;=\; 1-\alpha(1-z^2)\;, \qquad \mathbb{E}A \;=\; 2\alpha\;.$$

Pour que $\mathbb{E}A$ soit inférieur à 1, il est nécessaire que $\alpha \in (0, 1/2)$. La FGP du nombre de clients est donnée, d'après (48), par :

$$F_Q(z) \;=\; (1-2\alpha)\,\frac{1}{1-\alpha(1-z)}\;.$$

L'inversion de cette fonction génératrice donne les probabilités stationnaires :

$$\mathbb{P}(Q=k) \;=\; (1-2\alpha)\,\frac{\alpha^k}{(1-\alpha)^{k+1}}\;, \quad k \geq 1, \qquad \pi_0 \;=\; \frac{1-2\alpha}{1-\alpha}\;.$$

Corrigé de l'exercice 3

La FGP de A est dans ce cas (voir le paragraphe 2.4) : $A(z) = (1-\rho)/(1-\rho z)$. En remplaçant dans l'Équation (48), il vient :

$$F_Q(z) \;=\; \frac{1-2\rho}{1-\rho}\,\frac{1-z}{1-\rho(1+z)}\;.$$

Les premiers moments de la distribution stationnaire sont :

$$\mathbb{E}Q \;=\; \frac{\rho^2}{(1-\rho)(1-2\rho)} \qquad \mathbb{E}Q^2 \;=\; \frac{\rho^2}{(1-\rho)(1-2\rho)^2}\;.$$

L'inversion de la FGP donne les valeurs des probabilités stationnaires :

$$\pi_k \;=\; (1-2\rho)\,\frac{\rho^{k+1}}{(1-\rho)^{k+2}}\;, \quad k \geq 1, \qquad \pi_0 \;=\; \frac{1-2\rho}{(1-\rho)^2}\;.$$

Table des matières

Bibliographie

[1] W. Feller. *An introduction to probability theory and its applications*, volume I. J. Wiley & Sons, New York, troisième edition, 1968.

[2] P. Flajolet, D. E. Knuth, and B Pittel. The first cycles in an evolving graph. *Discrete Math*, 75 :167–215, 1989.

[3] D. Foata. La série exponentielle dans les problèmes d'énumération. *S.M.S Montreal University-Press*, 1974.

[4] I. P. Goulden and D. M. Jackson. *Analytic Combinatorics.* Combinatorial Enumeration, wiley, new york edition, 1983.

[5] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete Mathematics.* Addison Wesley, 1989.

[6] T.E. Harris. *The theory of branching processes.* Dover, 1989.

[7] F. Kelly. *Reversibility and Stochastic Networks.* J. Wiley & Sons, Chichester, 1979. Disponible sur le site web de l'auteur.

[8] J.G. Kemeny and Snell J.L. *Finite Markov Chains*, volume 76. Springer-Verlag, 1976.

[9] L. Kleinrock. *Queueing Systems*, volume I, Theory. J. Wiley & Sons, New York, 1975.

[10] H. Minc. *Nonnegative Matrices.* J. Wiley & Sons, 1988.

[11] G. Pujolle and S. Fdida. *Modèles de Systèmes et de Réseaux – Tome I : Performance.* Eyrolles, 1989.

[12] S. Ross. *Stochastic Processes.* Wiley & Sons, 1996.

[13] W. Rudin. *Real and Complex Analysis.* McGraw-Hill, 1974.

[14] A. Rényi. On connected graphs. *Mat. Kut. INT. Közl.*, 4 :385–388, 1959.

[15] R. Sedgewick and Ph. Flajolet. *Analysis of Algorithms.* Addison-Wesley, 1996. Voir les sites web des auteurs.

[16] E. Seneta. *Non-negative Matrices and Markov chains.* Springer-Verlag, second edition, 1981.

[17] L. Takács. *An introduction to the theory of queues.* Oxford University Press, New York, 1962.

[18] E.C. Titchmarsh. *The Theory of Functions.* Oxford University Press, second edition, 1939.

[19] H.S. Wilf. *Generatingfunctionology.* Academic Press, second edition, 1994. Disponible sur le site web de l'auteur.