Une propriété caractéristique des processus de Poisson-Dirichlet

Thomas Lehéricy

Une propriété caractéristique des processus de Poisson-Dirichlet

Mémoire du M2 PMA (année 2014-2015) de l'université Paris 6, réalisé sous la direction de Bartłomiej Błaszczyszyn

Thomas Lehéricy

15 septembre 2015

Résumé

Les processus de Poisson-Dirichlet et les processus Griffiths-Engen-McCloskey (GEM) sont deux familles très liées de variables aléatoires à valeurs dans les suites de réels positifs de somme 1. Nous les étudions à l'aide de la théorie des processus ponctuels, en nous appuyant sur les travaux de Pitman, Yor, et Handa. Nous énonçons une propriété caractéristique des processus GEM, puis un Théorème de type Slivnyak-Mecke portant sur les mesures de Palm. A l'aide de ce dernier, nous redémontrons la formule de changement de mesure, qui permet de construire les processus de Poisson-Dirichlet au moyen de processus ponctuels de Poisson. Enfin, nous généralisons l'espace dans lequel les processus prennent leurs valeurs et explicitons les processus ponctuels qui vérifient la propriété caractéristique.

Table des matières

1 Introduction

Les processus de Poisson-Dirichlet et les GEM (Griffiths-Engen-McCloskey) sont deux familles à deux paramètres de probabilités aléatoires. Ils ont de très nombreuses applications en statistique bayésienne [5], en tant que loi a priori ; en théorie de la génétique des populations [3] ; en combinatoire [1, 7] ; et en informatique pour l'analyse d'image et la reconnaissance de langage [1] [2]. C'est un choix pertinent pour modéliser une population divisée en sous-populations de taille aléatoire, lorsqu'on s'intéresse aux proportions de chaque sous-population.

Ces processus sont assez bien connus depuis les travaux de référence de Pitman et Yor [11], qui les ont reliés à des subordinateurs et aux excursions de mouvement brownien et de processus de Bessel. On peut également les étudier à l'aide de la théorie des processus ponctuels, en voyant une probabilité aléatoire comme une somme de Diracs en ses poids. Les processus de Poisson-Dirichlet et les processus GEM s'y prêtent particulièrement bien [6], d'autant plus qu'un processus GEM et un processus de Poisson-Dirichlet de mêmes paramètres sont (en loi) des numérotations mesurables du même processus ponctuel : dans l'ordre décroissant pour Poisson-Dirichlet, et selon une numérotation randomisée invariante par permutation biaisée par la taille (IPBT) pour les GEM (nous verrons les numérotations IPBT dans la partie 2). Un résultat de Pitman [10] permet d'énoncer une propriété qui caractérise les processus GEM : ce sont les seuls modèles à allocation résiduelle (residual allocation model) invariants par permutation biaisée par la

1. language technology

taille (invariant by size-biased permutation) avec un nombre non p.s. fini d'atomes. C'est cette propriété qui est au coeur des résultats de ce mémoire. Nous la rappellerons dans le Théorème 1.

Nous commencerons dans la deuxième section par quelques définitions et caractérisations utiles, sans preuves. Nous verrons principalement la définition, deux constructions et quelques propriétés élémentaires des permutations biaisées par la taille, qui jouent un rôle central dans ce mémoire. Le lecteur intéressé trouvera en annexe quelques résultats utiles supplémentaires, avec leurs preuves, sur les permutations biaisées par la taille de probabilités aléatoires. Nous avons en particulier énoncé et démontré le Lemme 4, qui joue un rôle important dans la quatrième partie pour caractériser les marquages de processus de Poisson-Dirichlet satisfaisant une certaine propriété caractéristique, qu'on énoncera dans le Théorème 6. Il permet en effet d'expliciter la relation de dépendance entre les lois des différents poids qui est imposée par la propriété d'IPBT : on ne peut modifier un poids indépendamment des autres. Un résultat similaire est utilisé par Pitman dans [10] pour montrer le Théorème 1, mais nous n'avons pas retrouvé le résultat du Lemme 4 dans notre bibliographie.

La troisième section de ce document est consacrée à la caractérisation des processus de Poisson-Dirichlet en tant que MAR IPBT, et à quelques applications, notamment le calcul de ses mesures moyennes et de Palm et une caractérisation qui s'exprime en terme de processus ponctuels (proposition 6). Cette dernière caractérisation sera au cœur des autres parties.

Nous verrons dans la quatrième section la démonstration d'une construction connue des processus de Poisson-Dirichlet à l'aide de processus ponctuels de Poisson, appelée *change of measure formula* par Pitman et Yor [11] [2] à l'aide de cette caractérisation. Encore une fois, nous n'avons pas trouvé de démonstration semblable dans notre bibliographie. La propriété caractéristique énoncée dans la partie 3 permet une démonstration simple, en deux temps, à l'aide des probabilités de Palm : d'abord, on montre que les processus ainsi construits sont des processus de Poisson-Dirichlet ; puis on identifie leurs paramètres.

Enfin, la cinquième section explicite les marquages de processus de Poisson-Dirichlet qui vérifient la caractérisation de la section 3. On fera appel au Lemme 4.

2 Probabilité aléatoire et permutation biaisée par la taille

Définition d'une probabilité aléatoire. Une probabilité aléatoire est un processus $P = (P_n)_{n>0}$ tel que pour tout n, P_n est une variable aléatoire réelle positive, et p.s. $\sum_{n>0} P_n = 1$. On appelle P_n le n^{e} poids. On peut voir P comme un vecteur aléatoire de probabilité sur $\mathbb{N}^*$, qui représente une population divisée en sous-populations de tailles aléatoires. P_n est la proportion d'individus qui appartiennent à la n^{e} sous-population.

2. C'est la propriété 14 de [11], p.865. Il fait référence à l'article [9].

Rappels rapides de théorie des processus ponctuels. Il peut être intéressant de voir un tel processus comme un nuage de points sur la droite réelle, en associant à chaque variable P_n un point à l'abscisse P_n. Pour traiter le nuage associé, on fait appel à la théorie des processus ponctuels. Elle sera par exemple intensivement utilisée dans la section 4 pour prouver la construction des processus de Poisson-Dirichlet à l'aide de renormalisations de certains processus ponctuels de Poisson.

Un processus ponctuel à valeurs dans un espace polonais E muni de sa tribu borélienne $\mathcal{E}$ est une variable aléatoire à valeurs dans les mesures ponctuelles (à support discret) de $(E, \mathcal{E})$ de Radon (la mesure d'un compact est finie) à valeurs entières (et positives). Comme on est dans un polonais, on peut lui associer la suite de ses atomes, à renumérotation près. Si p.s. tous les atomes sont distincts, on dit que le processus ponctuel est simple.

A toute probabilité aléatoire P, on peut associer un processus ponctuel $\Phi = \sum_{n:P_n>0} \delta_{P_n}$. Il est à valeurs dans $]0, 1]$, et sa *somme* $\int_{]0,1]} x\Phi(\mathrm{d}x) = \sum_{X\in\Phi} X = \sum_{n>0} P_n$ vaut 1.[3]. Pour que ce processus associé soit bien un processus ponctuel, donc p.s. de Radon, on a besoin d'éliminer 0 de l'espace. En effet, si P contient une infinité de 0 avec probabilité non nulle, ou si 0 est un point d'accumulation du processus, alors on aura $\Phi([0, \epsilon[) = \infty$ pour tout $\epsilon > 0$, et avec probabilité non nulle Φ ne sera pas de Radon. C'est pour cela qu'on se restreint à]0,1]. Réciproquement, toute numérotation mesurable d'un tel processus, à laquelle on tolère l'ajout de 0 supplémentaires, est une probabilité aléatoire. Si un processus ponctuel vérifie ces propriétés, il existe donc au moins une probabilité aléatoire telle qu'il soit le processus ponctuel associé à cette probabilité aléatoire.

"Numérotation mesurable" d'une probabilité aléatoire est ici pris au sens suivant : ϕ est une numérotation mesurable si pour tout $k, n > 0$, $\{\phi(k) = n\}$ est mesurable. Une numérotation mesurable d'un processus ponctuel Φ est une suite de variables aléatoires telles que le processus ponctuel associé à cette suite de variables aléatoires soit p.s. égal à Φ. Sous l'hypothèse que Φ a ses atomes dans]0,1] et est de somme finie, la numérotation de ses atomes dans l'ordre décroissant est une numérotation mesurable.

Proposition 1 (Équivalence entre probabilité aléatoire et nuage ponctuel)**.** *À toute probabilité aléatoire P, on peut associer un processus ponctuel Φ p.s. à atomes dans]0,1] et de somme 1, tel que P soit une numérotation mesurable de Φ.*

Réciproquement, si Φ est un processus ponctuel qui p.s. a ses atomes dans]0,1] et est de somme 1, alors il existe une probabilité aléatoire qui est une numérotation mesurable de Φ.

Suivant les cas, il est plus agréable de manipuler une probabilité aléatoire ou son processus ponctuel associé. Une fois qu'on a un processus ponctuel Φ associé à une probabilité aléatoire (à atomes dans]0,1] et de somme 1), il faut donc parfois choisir une numérotation mesurable judicieuse pour retomber sur une probabilité aléatoire. Une numérotation naturelle consiste à numéroter les atomes de Φ dans l'ordre décroissant. Une autre sont

3. Dans le cas où Φ est simple, on identifiera parfois abusivement Φ et l'ensemble de ses atomes, en écrivant par abus de notation $\Phi = \{P_n, n > 0\}$.

les permutations biaisées par la taille (PBT, *size-biaised permutation* ou SBP en anglais) de la probabilité aléatoire. Nous verrons que contrairement à la numérotation par ordre décroissant, il y a unicité en loi, mais pas p.s.. Quand, dans la suite, nous parlerons de "la" PBT, ou "du" tirage biaisé par la taille, l'unicité impliquée sera donc à comprendre comme l'unicité en loi.

Première construction de la PBT : Tirage avec remise. Donnons-nous une population répartie en sous-populations suivant la probabilité P. À chaque étape, on tire avec remise un individu uniformément dans la population, et on note la sous-population à laquelle il appartient, si elle n'a pas déjà été tirée, et la proportion de cette sous-population dans la population totale. On répète le procédé jusqu'à ce que toutes les sous-populations de taille non nulle aient été notées. La liste $\tilde{P}$ des proportions des sous-populations dans leur ordre d'apparition est une *permutation biaisée par la taille de* P. On appelle $\tilde{P}_n$ le n^e *tirage biaisé par la taille de* P.

De manière plus formelle, soient X_1, X_2, ... des variables aléatoires iid de loi P sachant P : $\mathbb{P}(X_1 = n|P) = P_n$. X_n représente la sous-population à laquelle appartient le n^{e} individu tiré. On pose $1 = N_1 < N_2 < ...$ les indices successifs où de nouvelles valeurs apparaissent dans la suite $(X_n)_{n>0}$, $\nu(n) = X_{N_n}$ la n^{e} sous-population à être tirée, et $\tilde{P}_n = P_{\nu(n)}$ la proportion de cette sous-population dans la population totale. Si la suite ne prend qu'un nombre fini m de valeurs, on pose $\tilde{P}_k = 0$ pour tout $k > m$.

Deuxième construction de la PBT : Tirage sans remise. La méthode de construction précédente est analogue à un tirage biaisé avec remise. Il existe une seconde méthode, sans remise, pour la construire : à chaque étape, on tire un individu uniformément parmi les sous-populations non encore tirées, jusqu'à les avoir toutes tirées. La loi de la permutation obtenue ainsi est la même que précédemment.

On peut la voir comme un algorithme de mélange d'un jeu de carte biaisé. On se donne un jeu de carte infini, avec un poids positif (éventuellement nul) inscrit sur chaque carte tel que la somme des poids inscrits vaille 1. Le poids inscrit sur la n^{e} carte est noté P_n. À l'instant initial, on dispose de deux tas : l'un est vide (tas de destination), l'autre est le tas de carte initial (tas source). À chaque étape, on tire une carte dans le tas source, avec une probabilité proportionnelle à son poids, et on l'ajoute au-dessus du tas de destination. La suite des poids, lue dans l'ordre du tas ainsi obtenu, est une PBT de P.

Pour le formaliser, on définit à chaque étape un échantillon biaisé par la taille, qui est l'indice du tirage biaisé par la taille. Si, à une étape k, un indice n n'a pas été tiré avant l'étape k, alors la probabilité que le k^{e} échantillon biaisé par la taille soit n est proportionnelle à P_n, et si n a déjà été tiré, la probabilité est nulle. On obtient donc la formule suivante, pour tout n_1, ..., n_k distincts deux à deux, sur l'évènement $\{P_{n_1} > 0, ..., P_{n_k} > 0\}$:

$$\mathbb{P}\left((N_1, ..., N_k) = (n_1, ..., n_k)|P\right) = P_{n_1} \frac{P_{n_2}}{1 - P_{n_1}} ... \frac{P_{n_k}}{1 - P_{n_1} - ... - P_{n_{k-1}}}. \tag{1}$$

Hors de ces évènements, certains échantillons ne sont pas bien définis. Les conséquences sont limitées, puisque les tirages associés sont alors nuls. La suite des tirages $\tilde{P}_n = P_{N_n}$, complétée par 0 là où les échantillons sont mal définis, est une permutation biaisée par la taille de P.

On peut montrer que ces deux constructions sont équivalentes.

A présent, donnons-nous une définition plus générale :

Définition 1. *On appelle PBT de P toute probabilité aléatoire qui a la même loi conditionnellement à P que les constructions précédentes.*

La définition à l'aide des indices permet de construire la PBT de P, mais elle est très peu pratique dès qu'il s'agit de vérifier qu'une probabilité aléatoire donnée est une PBT de P, puisqu'il faut reconstruire un jeu d'échantillons convenable. Heureusement, on dispose d'une caractérisation qui fait uniquement appel aux poids.

Dans le cas où le processus P est simple, il y a p.s. bijection entre les poids et les indices. Cette définition s'exprime alors de manière plus simple, et permet d'énoncer une version simplifiée de notre caractérisation :

Proposition 2 (Caractérisation d'une PBT, cas simple)**.** *Soit P une probabilité aléatoire telle que le processus ponctuel associé Φ soit simple. $\tilde{P}$ est une PBT de P si et seulement si p.s. pour tout $k > 0$, pour tout n_1, ..., n_k distincts deux à deux, sur l'évènement $\{P_{n_1} > 0, ..., P_{n_k} > 0\}$:*

$$\mathbb{P}\left((\tilde{P}_1, ..., \tilde{P}_k) = (P_{n_1}, ..., P_{n_k})|P\right) = P_{n_1}\frac{P_{n_2}}{1 - P_{n_1}}...\frac{P_{n_k}}{1 - P_{n_1} - ... - P_{n_{k-1}}}. \tag{2}$$

Dans le cas non simple, on n'a plus cette bijection : un même poids peut correspondre à plusieurs indices. On a cependant une formule très semblable, où on prend en compte la multiplicité des poids. Voici la caractérisation dans le cas le plus général :

Proposition 3 (Caractérisation d'une PBT, cas général)**.** *Soit P une probabilité aléatoire. $\tilde{P}$ est une PBT de P si et seulement si p.s., pour tout n_1, ..., n_k distincts deux à deux, si on note $A_j = Card\{n > 0 : P_n = P_{n_j}\} - Card\{0 < u < j : P_{n_u} = P_{n_j}\}$ pour tout $0 < j \leq k$ le nombre de poids égaux à P_{n_j} non encore comptés dans les P_{n_u}, $u < j$, alors sur l'évènement $\{P_{n_1} > 0, ..., P_{n_k} > 0\}$:*

$$\mathbb{P}\left((\tilde{P}_1, ..., \tilde{P}_k) = (P_{n_1}, ..., P_{n_k})|P\right) = A_1 P_{n_1}\frac{A_2 P_{n_2}}{1 - P_{n_1}}...\frac{A_k P_{n_k}}{1 - P_{n_1} - ... - P_{n_{k-1}}}. \tag{3}$$

Quelques propriétés remarquables des PBT. Une probabilité aléatoire est dite invariante par permutation biaisée par la taille (IPBT, *invariant by size-biased permutation* ou ISBP en anglais) si sa PBT a la même loi. Remarquons que la propriété d'IPBT est une propriété intrinsèque, inconditionnelle de la probabilité aléatoire, alors qu'être la PBT d'une autre probabilité aléatoire nécessite la connaissance d'une autre probabilité aléatoire, et de la loi conditionnelle de la première par rapport à celle-ci.

Toute PBT d'une probabilité aléatoire est toujours une probabilité aléatoire, et le processus ponctuel $\tilde{\Phi}$ associé est p.s. égal au processus ponctuel Φ associé à P (dans le tirage avec remise, p.s. tous les poids non nuls apparaissent une et une seule fois). Ceci assure que toute PBT d'une probabilité aléatoire peut être vue comme une numérotation particulière associée au même nuage ponctuel.

La proposition assez subtile suivante assure qu'une probabilité aléatoire qui a la même loi qu'une PBT est elle-même une PBT, autrement dit elle assure l'existence d'une probabilité aléatoire telle que la loi conditionnelle de la première par rapport à celle-ci soit celle d'une PBT. Cela permet de passer d'une propriété sur la loi inconditionnelle à une propriété sur les lois conditionnelles. On s'en servira dans la démonstration du Lemme 4.

Proposition 4 (Une PBT en loi est une PBT)**.** *Toute probabilité aléatoire P qui a même loi qu'une PBT est une PBT de la probabilité aléatoire $(P_{(n)})_{n>0}$, obtenue en renumérotant ses atomes par ordre décroissant.*

Démonstration. Supposons qu'il existe Q une probabilité aléatoire et R une PBT de Q tel que P ait même loi que R. Notons $\bar{P} = (P_{(n)})_{n>0}$, $\bar{Q} = (Q_{(n)})_{n>0}$ et $\bar{R} = (R_{(n)})_{n>0}$. La caractérisation des PBT appliquée à Q et R s'écrit, avec $A_j^Q = \text{Card}\{n > 0 : Q_n = Q_{n_j}\} - \text{Card}\{0 < u < j : Q_{n_u} = Q_{n_j}\}$:

$$\mathbb{P}\left((R_1, ..., R_k) = (Q_{n_1}, ..., Q_{n_k})|Q\right) = A_1^Q Q_{n_1} \frac{A_2^Q Q_{n_2}}{1 - Q_{n_1}} ... \frac{A_k^Q Q_{n_k}}{1 - Q_{n_1} - ... - Q_{n_{k-1}}}. \tag{4}$$

Cette équation reste vraie en remplaçant Q par $\bar{Q}$, qui est une fonctionnelle de Q. Or, p.s. les processus ponctuels associés à P, Q et R sont égaux. Comme $\bar{P}$, $\bar{Q}$ et $\bar{R}$ sont des fonctionnelles de ces processus ponctuels, elles sont p.s. égales. On peut donc conditionner par rapport à $\bar{R}$.

$$\mathbb{P}\Big((R_1, ..., R_k) = (\bar{R}_{n_1}, ..., \bar{R}_{n_k})|\bar{R}\Big) = A_1^{\bar{R}} \bar{R}_{n_1} \frac{A_2^{\bar{R}} \bar{R}_{n_2}}{1 - \bar{R}_{n_1}} ... \frac{A_k^{\bar{R}} \bar{R}_{n_k}}{1 - \bar{R}_{n_1} - ... - \bar{R}_{n_{k-1}}}. \tag{5}$$

Comme P a même loi que R, cette équation reste valable en remplaçant R par P. D'où P est une PBT de $\bar{P}$. □

Cette proposition a trois conséquences immédiates, et dont on se servira bien plus souvent.

- *Deux PBT associées au même processus ponctuel Φ ont même loi, qu'on appelle la loi de la PBT de Φ.* En effet, ce sont toutes les deux des PBT de P, la numérotation des atomes de Φ dans l'ordre décroissant. Elles ont donc la même loi.

  Cette propriété d'unicité permet de voir la PBT d'un processus ponctuel comme une numérotation particulière, ayant des propriétés agréables. On peut notamment directement définir *la* numérotation IPBT, unique en loi, d'un processus ponctuel à valeurs sur $]0,1]$ de somme 1 ; elle est obtenue en prenant la PBT d'une numérotation quelconque de ses points.

- *Toute PBT d'une probabilité aléatoire est IPBT.* En effet, cette probabilité aléatoire a même loi que sa PBT, puisqu'elles sont associées au même processus ponctuel.
- *Toute probabilité aléatoire IPBT est une PBT.* C'est une conséquence directe de la proposition 4, et une réciproque au point précédent.

On en déduit l'unicité en loi de la numérotation IPBT d'un processus ponctuel. Enfin, une probabilité aléatoire IPBT aura p.s. tous ses éléments non nuls au début, les zéros n'apparaissant que si elle a un nombre fini de poids non nuls, et après tous les poids non nuls :

Proposition 5 (Support d'une probabilité aléatoire IPBT)**.** *Si P est une probabilité aléatoire IPBT, il existe une variable aléatoire K $\sigma(P)$-mesurable à valeurs dans $\mathbb{N}^* \cup \{\infty\}$ telle que*

$$\{n : P_n > 0\} = \{1, ..., K\} \cap \mathbb{N}^*. \tag{6}$$

Idée de démonstration. La démonstration se fait à partir de la construction de la PBT. Conditionnellement à P, p.s., la PBT $\tilde{P}$ de P vérifie

$$\{n : P_n > 0\} = \{1, ..., K\} \cap \mathbb{N}^* \tag{7}$$

où $K = \mathbb{E}\left[\mathrm{Card}\{n : P_n > 0\} | P\right]$ est une variable aléatoire. D'où le résultat. $\square$

On intersecte avec $\mathbb{N}^*$ pour être sûr d'exclure l'infini. K est le nombre de poids non nuls. Il peut être infini ; c'est le cas par exemple pour les processus de Poisson-Dirichlet.

Cette variable est invariante p.s. par PBT, car c'est une fonctionnelle du nuage ponctuel associé à la probabilité aléatoire.

En conditionnant une probabilité aléatoire P à $K = k$, pour un k tel que cet évènement est de probabilité non nulle, on obtient toujours une probabilité aléatoire. Notons-la P^k. Alors P est IPBT si et seulement si pour tout k tel que $\mathbb{P}(K = k) > 0$, P^k est IPBT : on peut toujours se restreindre à K constant p.s.. En pratique, pour les cas qui nous intéresse (les MAR IPBT), K sera p.s. constant, et on pourra se passer d'un tel conditionnement. C'est pour cela que la plupart des propriétés énoncées en annexe supposent K constant p.s..

Les permutations IPBT peuvent être vues comme point fixe d'un certain nombre de dynamiques de mélange[4] (*cf* annexe), dont certaines donneront des résultats que nous exploiterons dans la suite (Lemme 4).

Cette propriété d'IPBT porte sur une probabilité aléatoire, donc sur une *numérotation* mesurable d'un processus ponctuel ; il est donc nécessaire, pour pouvoir en tirer des informations, de disposer de contraintes supplémentaires sur cette même probabilité aléatoire IPBT. La propriété d'être un modèle à allocation résiduelle en est une qui mène à des résultats intéressants.

4. Terminologie non canonique.

3 MAR IPBT

Dans le cadre de la seconde construction apparaissent naturellement les grandeurs $V_n = \dfrac{\tilde{P}_n}{1 - \tilde{P}_1 - ... - \tilde{P}_{n-1}}$, la proportion de la n^{e} sous-population tirée dans les sous-populations non encore tirées à l'étape n. Cette formule pour les définir n'est valable que sur $\{K \geq n\}$. Sur $\{K < n\}$ on pose $V_n = 1$.

Les $\tilde{P}_n$ se réécrivent

$$\tilde{P}_n = \bar{V}_1 ... \bar{V}_{n-1} V_n \tag{8}$$

où on introduit la notation $\bar{v} = 1 - v$, que nous utiliserons dans la suite.

Toute probabilité aléatoire s'écrit de manière unique

$$\begin{cases} P_1 = V_1 \\ P_2 = \bar{V}_1 V_2 \\ ... \\ P_n = \bar{V}_1 ... \bar{V}_{n-1} V_n \quad \forall n > 0. \end{cases} \tag{9}$$

pour $(V_n)_{n>0}$ variables aléatoires à valeurs dans [0,1] p.s., où $V_n = \dfrac{P_n}{1 - P_1 - ... - P_{n-1}}$ est une fonction déterministe de P.

On appelle *modèle à allocation résiduelle* (MAR, RAM ou *random allocation model* en anglais) tout processus P tel que les $V_1, ..., V_n, ...$ sont des variables aléatoires *indépendantes*. On l'appelle aussi *stick breaking model* : on casse une proportion V_1 d'un bâton de longueur 1, puis une proportion V_2, *indépendante de* V_1, du reste du bâton, etc. La suite des longueurs des morceaux découpés est la probabilité aléatoire associée. L'indépendance est ici la seule vraie contrainte : en effet, pour toute probabilité aléatoire P, il existe une suite V, pas nécessairement indépendante telle que P peut s'écrire de cette manière.

3.1 Théorème de Pitman - caractérisation des MAR IPBT

La propriété d'IPBT est, comme nous l'avons vu, peu contraignante, puisque la PBT d'une probabilité aléatoire quelconque est IPBT. Le fait d'être un MAR est également une hypothèse assez faible, puisque, bien qu'on force l'indépendance entre les différentes étapes, on n'impose aucune contrainte sur leurs lois. Cependant, combiner les deux propriétés s'avère extrêmement restrictif. En effet, la structure de MAR impose l'indépendance entre les V_n, et la propriété d'IPBT impose des relations entre les lois des V_n.

On peut montrer que l'indépendance des V_n implique que K sera p.s. constant. Cela justifie l'hypothèse de constance de K que nous faisons dans certains énoncés. Mais ce n'est pas la seule contrainte, ni la plus forte. En fait, elle est incluse dans le résultat prouvé par Pitman [10], qui est le suivant :

Théorème 1. *Les MAR IPBT sont les processus de Poisson-Dirichlet généralisés et quelques processus dégénérés. Plus précisément, si P est un MAR IPBT :*

(i) Si $P_n > 0$ ps pour tout n, il existe $0 \leq \alpha < 1$ et $\theta > -\alpha$ tels que V_n suit la loi Beta($1-\alpha$, $\theta + n\alpha$) pour tout n. Un tel processus est appelé processus GEM de paramètres (α, θ), et noté GEM(α, θ). Le processus composé de ses points triés par ordre décroissant est appelé processus de Poisson-Dirichlet généralisé de paramètres (α, θ), et noté PD(α, θ). Le processus ponctuel associé est simple.

Sinon, il existe un entier m tel que p.s. $\{n, P_n > 0\} = \{1, ..., m\}$ (autrement dit K est p.s. constant et fini), et on est dans un des cas suivants :

(ii) Il existe $\beta > 0$ tel que V_n suive la loi Beta($1+\beta, m\beta - n\beta$) pour $n \leq m$.

(iii) $V_n = 1/(m-n+1)$ p.s., soit $P_n = 1/m$ pour tout $n \leq m$.

(iv) $m = 2$, et la probabilité F sur $]0,1[$ définie par

$$F(\mathrm{d}v) = \bar{v}\mathbb{P}(V_1 \in \mathrm{d}v)/\mathbb{E}(\bar{V}_1) \tag{10}$$

est symétrique par rapport à $1/2$, où $\bar{v} = 1 - v$.

Réciproquement, tous les cas énumérés ici sont des MAR IPBT.

Toute probabilité aléatoire IPBT sur $\{1,2\}$ vérifie (iv), et réciproquement, si P satisfait (iv), c'est une probabilité aléatoire IPBT sur $\{1,2\}$.

Remarque 1. *Les processus de Poisson-Dirichlet ont d'abord été introduits par Kingman [8], dans le cas PD$(0,\theta)$. Ils ont une forme particulièrement simple, puisque la suite des $(V_n)_{n>0}$ est alors iid. Kingman a également démontré la représentation des processus de Poisson-Dirichlet à l'aide de processus ponctuels de Poisson, que nous verrons dans la section 4, pour ces paramètres. Ces résultats ont été ensuite généralisés, notamment par Pitman et Yor [11], d'où leur appellation de* processus de Poisson-Dirichlet généralisés, *ou parfois Pitman-Yor, pour désigner ces processus.*

3.2 Propriété caractéristique des MAR IPBT

Soit P une probabilité aléatoire IPBT, et Φ le processus ponctuel associé. On définit deux suites de processus ponctuel : $(\tilde{\Phi})_{n\geq 0}$, le nuage résiduel (ce qui reste après avoir enlevé les premiers tirages), et $(\tilde{\Phi}^r)_{n\geq 0}$, le nuage résiduel renormalisé (ce qui reste, mais renormalisé pour obtenir un processus de somme 1), de la manière suivante :

$$\begin{cases} \tilde{\Phi}_n = \Phi \setminus \{P_1, ..., P_n\} \\ \tilde{\Phi}^r_n = \dfrac{1}{1 - P_1 - ... - P_n}\tilde{\Phi}_n \end{cases} \tag{11}$$

en utilisant l'abus de notation consistant à identifier un processus ponctuel à ses atomes, et en notant

$$\lambda\left(\sum_{n>0} \delta_{P_n}\right) := \sum_{n>0} \delta_{\lambda P_n}. \tag{12}$$

où on pose, par convention, $\frac{1}{0}\Psi = 0$ pour tout Ψ.

Un petit calcul donne $1 - P_1 - ... - P_n = \bar{V}_1 ... \bar{V}_n$, avec $V_n = \frac{P_n}{1-P_1-...-P_{n-1}}$ comme défini précédemment. On en déduit la relation de récurrence :

$$\tilde{\Phi}_0 = \tilde{\Phi}_0^r = \Phi \tag{13}$$

et pour tout $n \geq 0$,

$$\begin{cases} \tilde{\Phi}_{n+1} & = \tilde{\Phi}_n \setminus \{P_{n+1}\} = \{P_{n+2}, P_{n+3}, ...\} \\ \tilde{\Phi}_{n+1}^r & = \frac{1}{\bar{V}_{n+1}} \left(\tilde{\Phi}_n^r \setminus \{V_{n+1}\}\right) = \frac{1}{\bar{V}_1 \bar{V}_2 ... \bar{V}_{n+1}} \tilde{\Phi}_{n+1} \\ & = \{V_{n+2}, \bar{V}_{n+2} V_{n+3}, \bar{V}_{n+2} \bar{V}_{n+3} V_{n+4} ...\}. \end{cases} \tag{14}$$

Remarquons que la somme de $\tilde{\Phi}_n^r$ est p.s. 0 ou 1 ; elle reste à 1 tant que $V_n < 1$, et est nulle après. Si la somme de $\tilde{\Phi}_n^r$ ou $\tilde{\Phi}_n$ est nulle, alors V_k et $\tilde{P}_k$ sont nuls pour tout $k > n$.

La propriété suivante est une reformulation de la propriété de MAR (la proportion de chaque tirage dans le nuage qui reste à cette étape est indépendante de ces proportions aux autres étapes) qui exploite l'IPBT pour assouplir les hypothèses. Elle fournit une interprétation simple des processus de Poisson-Dirichlet.

Proposition 6 (Propriété caractéristique des MAR IPBT)**.** *Soit P une probabilité aléatoire IPBT. Les propriétés suivantes sont équivalentes :*

(i) Pour tout n, V_1, ..., V_n et $\tilde{\Phi}_n^r$ sont tous indépendants.

(ii) Pour tout n, V_n et $\tilde{\Phi}_n^r$ sont indépendants.

(iii) P est un MAR IPBT.

Remarquons que cette propriété est la définition d'un MAR, mais avec des hypothèses affaiblies grâce à la propriété d'IPBT. En effet, l'IPBT assure l'unicité (en loi) de la numérotation de la suite : le processus ponctuel associé contient autant d'information en loi. Sans elle, il aurait donc fallu remplacer $\tilde{\Phi}_n^r$ par $(V_k)_{k>n}$ dans les points (i) et (ii). On retomberait alors sur la définition des MAR.

Cette propriété s'interprète de la manière suivante. Si on tire un élément d'un MAR IPBT de manière biaisée par la taille et qu'on renormalise les poids restants pour obtenir à nouveau une probabilité, alors cet élément et le nuage résiduel renormalisé sont indépendants. De plus, le nuage résiduel renormalisé, numéroté de manière IPBT, est toujours un MAR IPBT ! Cette propriété reste encore vraie en tirant un nombre fini arbitraire de poids de manière biaisée par la taille.

Démonstration. (ii) découle immédiatement de (i).

Réciproquement, pour (ii) implique (i), procédons par récurrence. Soit $m > 0$ donné, et supposons qu'on a montré (i) pour $n \leq m$; l'initialisation provient de (ii) pour $n = 1$. Par construction, la loi de V_{m+1} conditionnée à $(V_1, ..., V_m, \tilde{\Phi}_m^r)$ ne dépend que de $\tilde{\Phi}_m^r$, et par hypothèse de récurrence $\tilde{\Phi}_m^r$ est indépendant de $(V_1, ..., V_m)$, donc $(V_{m+1}, \tilde{\Phi}_{m+1}^r)$ est indépendant de $(V_1, ..., V_m)$. Par (ii), on a V_{m+1} est indépendant de $\tilde{\Phi}^r_{m+1}$. D'où (i) pour $n = m+1$.

(i) implique que les $(V_n)_{n>0}$ sont indépendants. Or

$$\begin{cases} P_1 = V_1 \\ P_2 = (1 - V_1)V_2 \\ ... \\ P_n = (1 - V_1)...(1 - V_{n-1})V_n \end{cases} \tag{15}$$

P est donc un MAR ; comme il est IPBT, on a (iii).

Réciproquement, si P est un MAR IPBT, alors la suite $(V_n)_{n>0}$ est indépendante. Par conséquent, pour n donné, en posant

$$Y^n_k = \frac{P_k}{(1 - V_1)...(1 - V_n)} = V_k \prod_{j=n+1}^{k-1} (1 - V_j), \tag{16}$$

$(Y^n_k)_{k>n}$ est indépendante de $(V_1, ..., V_n)$. On conclut donc que

$$\begin{aligned} \tilde{\Phi}^r_n &= \frac{1}{(1 - V_1)...(1 - V_n)} \tilde{\Phi}_n \\ &= \sum_{k>n} \delta_{\frac{P_n}{(1-V_1)...(1-V_n)}} \\ &= \sum_{k>n} \delta_{Y^n_k} \end{aligned} \tag{17}$$

est indépendant de $(V_1, ..., V_n)$. Les V_i étant indépendants entre eux, on conclut (i).

□

3.3 Cas des processus de Poisson-Dirichlet généralisés

Une formulation plus précise, mais spécifique aux processus de Poisson-Dirichlet, est la suivante :

Proposition 7. *Soit P une probabilité aléatoire IPBT. Les propriétés suivantes sont équivalentes :*

(i) Pour tout n, V_1, ..., V_n et $\tilde{\Phi}^r_n$ sont tous indépendants, et pour tout $1 \leq i \leq n$, V_i suit la loi Beta$(1 - \alpha, \theta + i\alpha)$, et $\tilde{\Phi}^r_n$ suit la loi PD$(\alpha, \theta + n\alpha)$.

(ii) Pour tout n, V_n et $\tilde{\Phi}^r_n$ sont indépendants, V_n suit la loi Beta$(1 - \alpha, \theta + n\alpha)$, et $\tilde{\Phi}^r_n$ suit la loi PD$(\alpha, \theta + n\alpha)$.

(iii) P suit la loi PD(α, θ).

Démonstration. Montrons $(i) \Rightarrow (iii)$. La propriété 6 donne que P est un MAR IPBT. Comme $K = \infty$ p.s., on est dans le cas (i) du Théorème 1. Les informations sur la loi de V_1 permettent d'identifier les paramètres. □

En réalité, si on se donne P un MAR IPBT, on voit qu'on peut identifier quel MAR IPBT il est à partir d'indicateurs beaucoup plus restreints : il suffit de disposer d'assez d'informations sur les V_i pour conclure grâce au Théorème 1. Par exemple, par le Théorème 1, on sait que $K = \mathrm{Card}\{n : P_n > 0\}$ est déterministe. S'il est fini, on est dans le cas (ii), (iii) ou (iv). Pour identifier lequel, il suffit si $K > 3$ de calculer par exemple deux moments de P_1. Si $K = 2$, on est automatiquement dans le cas (iv). Enfin, si K est infini, on est dans le cas (i), et il suffit de calculer par exemple deux moments non triviaux de P_1 pour identifier les paramètres du processus de Poisson-Dirichlet.

Conséquence de cette propriété, on peut donner une expression simple des mesures de Palm d'un processus de Poisson-Dirichlet généralisé Φ. On dispose également d'expressions explicites pour la mesure moyenne.

Corollaire 1 (Mesures moyennes et de Palm des processus de Poisson-Dirichlet généralisés)**.** *Soient $x_1, ..., x_n > 0$ de somme inférieure à 1. Les mesures de Palm associées à Φ un PD(α, θ) sont*

$$\begin{cases} \mathcal{P}^!_{x_1,...,x_n}(\Gamma) = P_{(\alpha,\theta+n\alpha)}((1 - x_1 - ... - x_n)\Phi \in \Gamma) \\ \mathcal{P}_{x_1,...,x_n}(\Gamma) = P_{(\alpha,\theta+n\alpha)}((1 - x_1 - ... - x_n)\Phi \cup \{x_1, ..., x_n\} \in \Gamma), \end{cases} \tag{18}$$

et sa mesure moyenne d'ordre n est

$$M^{(n)}(\mathrm{d}(p_1, ..., p_n)) = \frac{1}{p_1}\frac{1-p_1}{p_2}...\frac{1-p_1-...-p_{n-1}}{p_n}\mu_n\left(\mathrm{d}(p_1, ..., p_n)\right) \tag{19}$$

avec $\mu_n\left(\mathrm{d}(p_1, ..., p_n)\right)$ la loi du n-uplet $(P_1, ..., P_n)$, avec P une numérotation IPBT de Φ (on a son expression à l'aide du Théorème 1). [5]

Rappelons que la mesure de corrélation $M^{(n)}$ et les probabilités de Palm $\mathcal{P}_{x_1,...,x_n}$ et $\mathcal{P}^!_{x_1,...,x_n}$ (qu'on notera respectivement $\mathcal{P}_x$ et $\mathcal{P}^!_x$, avec $x = (x_1, ..., x_n)$) d'un processus ponctuel Φ sur E, avec $\mathbb{M}$ l'espace des mesures ponctuelles de Radon sur E (où Φ prend ses valeurs), sont définies comme vérifiant pour tout f mesurable :

$$\mathbb{E}\left(\sum_{i_1,...,i_n distincts} f(X_{i_1}, ..., X_{i_n}, \Phi)\right) \\ = \int_E M^{(n)}(\mathrm{d}x_1...\mathrm{d}x_n)\int_{\mathbb{M}} f(x_1, ..., x_n, \phi)\mathcal{P}_x(\mathrm{d}\phi) \tag{20}$$

$$\mathbb{E}\left(\sum_{i_1,...,i_n distincts} f\left(X_{i_1}, ..., X_{i_n}, \Phi \setminus \{X_{i_1}, ..., X_{i_n}\}\right)\right) \\ = \int_E M^{(n)}(\mathrm{d}x_1...\mathrm{d}x_n)\int_{\mathbb{M}} f(x_1, ..., x_n, \phi)\mathcal{P}^!_x(\mathrm{d}\phi) \tag{21}$$

5. Le lecteur intéressé pourra trouver une autre expression dans [12].

Remarquons aussi la relation entre ces deux mesures de Palm : M_n-p.p.,

$$\begin{cases}\mathcal{P}_{x_1,...,x_n}(\Phi \in \Gamma) = & \mathcal{P}^!_{x_1,...,x_n}(\Phi \cup \{x_1, ..., x_n\} \in \Gamma) \\ \mathcal{P}^!_{x_1,...,x_n}(\Phi \in \Gamma) = & \mathcal{P}_{x_1,...,x_n}(\Phi \setminus \{x_1, ..., x_n\} \in \Gamma).\end{cases} \tag{22}$$

Démonstration du corollaire. Si Φ est un processus de Poisson-Dirichlet généralisé, il est simple, et si P est une numérotation IPBT de ses points, p.s. tous les P_n sont strictement positifs. On note toujours $(P_{(n)})_{n>0}$ la suite des points de Φ numérotés dans l'ordre décroissant ; rappelons que P est une PBT de $(P_{(n)})_{n>0}$. Rappelons également (2) : P étant une PBT de $(P_{(n)})_{n>0}$, on a p.s. pour tout $i_1, ..., i_n > 0$ distincts deux à deux :

$$\begin{aligned}\mathbb{P}\left((P_1, ..., P_k) = (P_{(i_1)}, ..., P_{(i_n)})|P\right) =& P_{(i_1)}\frac{P_{(i_2)}}{1-P_{(i_1)}}...\frac{P_{(i_n)}}{1-P_{(i_n)}-...-P_{(i_{n-1})}} && (23)\\ =& T(P_{(i_1)}, ..., P_{(i_n)}), && (24)\end{aligned}$$

Donc

$$\begin{aligned}(*) :=& \mathbb{E}\left[\sum_{X_1,...,X_n\in\Phi} f\left(X_{i_1}, ..., X_{i_n}, \Phi \setminus \{X_{i_1}, ..., X_{i_n}\}\right)\right] && (25)\\ =& \mathbb{E}\left[\sum_{i_1,...,i_n>0 \text{ distincts}} f\left(P_{(i_1)}, ..., P_{(i_n)}, \Phi \setminus \{P_{(i_1)}, ..., P_{(i_n)}\}\right)\frac{T(P_{(i_1)}, ..., P_{(i_n)})}{T(P_{(i_1)}, ..., P_{(i_n)})}\right] && (26)\end{aligned}$$

En faisant intervenir une intégrale conditionnelle par rapport à Φ, comme les $(N_i)_{1\leq i\leq n}$ sont indépendants sachant Φ :

$$\begin{aligned}(*) =& \mathbb{E}\left[\frac{f\left(P_1, ..., P_n, \Phi \setminus \{P_1, ..., P_n\}\right)}{T(P_{(i_1)}, ..., P_{(i_n)})}\right] && (27)\\ =& \mathbb{E}\left[\frac{f\left(P_1, ..., P_n, (1-P_n-...-P_n)\tilde{\Phi}^r_n\right)}{T(P_{(i_1)}, ..., P_{(i_n)})}\right] && (28)\\ =& \mathbb{E}\left[\frac{g\left(P_1, ..., P_n\right)}{T(P_{(i_1)}, ..., P_{(i_n)})}\right] && (29)\end{aligned}$$

par la proposition 7, où

$$\begin{aligned}g\left(p_1, ..., p_n\right) =& \mathbb{E}\left[f\left(p_1, ..., p_n, (1-p_n-...-p_n)\tilde{\Phi}^r_n\right)\right] && (30)\\ =& \int_{\mathbb{M}} f(p_1, ..., p_n, (1-p_1-...-p_n)\phi)P_{(\alpha,\theta+n\alpha)}(\mathrm{d}\phi) && (31)\end{aligned}$$

avec $P_{(\alpha,\theta+n\alpha)}$ la loi du processus de Poisson-Dirichlet généralisé de paramètre (α, θ).

Donc, en inversant le passage de (25) à (29) :

$$(*) = \mathbb{E}\left[\sum_{X_1,...,X_n \in \Phi} g\left(X_1, ..., X_n\right)\right] \tag{32}$$
$$= \int_E M_n(\mathrm{d}x_1...\mathrm{d}x_n) g(x_1, ..., x_n) \tag{33}$$

On en déduit donc les formules suivantes pour $x_1, ..., x_n > 0$ de somme strictement inférieure à 1 :

$$\begin{cases} \mathcal{P}^!_{x_1,...,x_n}(\Gamma) = P_{(\alpha,\theta+n\alpha)}((1 - x_1 - ... - x_n)\Phi \in \Gamma) \\ \mathcal{P}_{x_1,...,x_n}(\Gamma) = P_{(\alpha,\theta+n\alpha)}((1 - x_1 - ... - x_n)\Phi \cup \{x_1, ..., x_n\} \in \Gamma). \end{cases} \tag{34}$$

L'expression de la mesure moyenne découle directement de l'équation (29), pour f qui ne dépend pas de Φ.

□

La propriété caractéristique est au cœur des deux sections suivantes. La section 5 sera consacrée à déterminer quels sont les processus ponctuels qui la vérifient. On exhibera dans la section 4 une famille de processus ponctuels construits pour que leurs probabilités de Palm aient une certaine structure, qui assure que les processus sont bien en réalité des processus de Poisson-Dirichlet généralisés. Cette structure est explicitée dans le Théorème 2 de la sous-section suivante.

3.4 Théorème Slivnyak-like

On dispose d'une réciproque à ce corollaire, qui permet d'affirmer que tout processus ponctuel dont la famille des probabilités de Palm possède une structure semblable à celle des processus de Poisson-Dirichlet est en fait un processus de Poisson-Dirichlet.

Théorème 2 (caractérisation des MAR IPBT par leurs probabilités de Palm)**.**

(i) Soit $(\Phi^n)_{n>0}$ une famille de processus ponctuels. Notons pour tout $k > 0$ et pour tout $n > 0$, $(\mathcal{P}^n_x)_{x \in E^k}$ et $\left(\mathcal{P}^{n,!}_x\right)_{x \in E^k}$ les familles des probabilités de Palm d'ordre k de Φ^n et P^n sa loi. Si on a pour tout $k > 0$ et pour tout $x = (x_1, ..., x_k)$ tel que $x_1, ..., x_k > 0$ de somme inférieure à 1 :

$$\begin{cases} \mathcal{P}^{n,!}_x(\Gamma) = P^{n+1}((1 - x_1 - ... - x_k)\Phi \in \Gamma) \\ \mathcal{P}^n_x(\Gamma) = P^{n+1}((1 - x_1 - ... - x_k)\Phi \cup \{x_1, ..., x_k\} \in \Gamma), \end{cases} \tag{35}$$

alors les Φ^n sont des MAR IPBT.

(ii) *Si de plus Φ^1 n'a pas un nombre p.s. fini de points, alors il existe $0 \leq \alpha < \theta$, $\theta > -\alpha$ tels que pour tout n, P^n soit la loi du processus ponctuel associé à un processus de Poisson-Dirichlet de paramètres $(\alpha, \theta + (n-1)\alpha)$.*

(iii) *Si tous les Φ^n ont même loi et que Φ^1 n'est pas p.s. nul, alors pour tout n, Φ^n est un processus de Poisson-Dirichlet de paramètres $(0, \theta)$. Autrement dit, soit Φ un processus ponctuel non p.s. nul. Notons pour tout $k > 0$, $(\mathcal{P}_x)_{x \in E^k}$ et $\left(\mathcal{P}_x^!\right)_{x \in E^k}$ les familles des probabilités de Palm d'ordre k de Φ et P sa loi. Si on a pour tout $k > 0$ et pour tout $x = (x_1, ..., x_k)$ tel que $x_1, ..., x_k > 0$ de somme inférieure à 1 :*

$$\begin{cases} \mathcal{P}_x^!(\Gamma) = P((1 - x_1 - ... - x_k)\Phi \in \Gamma) \\ \mathcal{P}_x(\Gamma) = P((1 - x_1 - ... - x_k)\Phi \cup \{x_1, ..., x_k\} \in \Gamma), \end{cases} \tag{36}$$

alors il existe $\theta > 0$ tel que Φ est un processus de Poisson-Dirichlet de paramètres $(0, \theta)$.

En fait, toute famille de lois $(P^n)_{n>0}$ telles que les processus ponctuels associés ont p.s. un nombre infini de points (ceci sert uniquement à éliminer les cas (ii), (iii) et (iv) du Théorème 1) et des probabilités de Palm possédant cette structure correspondent à une famille de loi de processus de Poisson-Dirichlet, pour un certain choix de paramètres.

Le point *(iii)* permet de caractériser les processus de Kingman. C'est l'énoncé le plus semblable au théorème de Slivnyak, puisque la famille de ses probabilités de Palm s'expriment en fonction de sa propre loi.

Démonstration. Démontrons d'abord le point *(i)*. Notons $\tilde{Q}^n$ une numérotation invariante par permutation biaisée par la taille de Φ_n ; Q^n sa numérotation dans l'ordre décroissant ; et $\tilde{\Phi}_k^{n,r} = \frac{1}{1 - \tilde{Q}_1^n - ... - \tilde{Q}_k^n}\left(\Phi \setminus \{\tilde{Q}_1^n, ..., \tilde{Q}_k^n\}\right)$.

$$\begin{aligned}
&\mathbb{E}\left[f(\tilde{Q}_1^n, ..., \tilde{Q}_k^n) g(\tilde{\Phi}_k^{n,r})\right] && (37)\\
=&\mathbb{E}\left[\sum_{i_1, ..., i_k \text{ distincts}} T(Q_{i_1}^n, ..., Q_{i_k}^n) f(Q_{i_1}^n, ..., Q_{i_k}^n) g\left(\frac{\Upsilon \setminus \{Q_{i_1}^n, ..., Q_{i_k}^n\}}{1 - Q_{i_1}^n - ... - Q_{i_k}^n}\right)\right] && (38)\\
=&\int M_{n,k}(\mathrm{d}(q_1, ..., q_k)) \int \mathcal{P}_{(q_1, ..., q_k)}^{n,!}(\mathrm{d}\mu) T(q_1, ..., q_k) f(q_1, ..., q_k) g\left(\frac{\mu}{1 - q_1 - ... - q_k}\right) && (39)\\
=&\int M_{n,k}(\mathrm{d}(q_1, ..., q_k)) T(q_1, ..., q_k) f(q_1, ..., q_k) \mathbb{E}[g(\Phi^{n+k})] && (40)\\
=&\mathbb{E}\left[\sum_{i_1, ..., i_k \text{ distincts}} T(Q_{i_1}^n, ..., Q_{i_k}^n) f(Q_{i_1}^n, ..., Q_{i_k}^n)\right] \mathbb{E}[g(\Phi^{n+k})] && (41)\\
=&\mathbb{E}\left[f(\tilde{Q}_1^n, ..., \tilde{Q}_k^n)\right] \mathbb{E}[g(\Phi^{n+k})] && (42)
\end{aligned}$$

Le passage de la première à la seconde ligne s'effectue en prenant une permutation biaisée par la taille de Q^n. $T(Q_{i_1}^n, ..., Q_{i_k}^n)$ est la probabilité sachant Q^n que les k premiers éléments de la PBT $\tilde{Q}^n$ de Q^n soient $(Q_{i_1}^n, ..., Q_{i_k}^n)$.

On en déduit l'indépendance de $(\tilde{Q}_1^n, ..., \tilde{Q}_k^n)$ et $\tilde{\Phi}_k^{n,r}$ pour tout $k > 0$. Par la propriété caractéristique (proposition 6), comme $\tilde{Q}^n$ est IPBT, il s'agit d'un MAR IPBT.

Pour montrer le point *(ii)*, il suffit de constater que les seuls MAR IPBT qui n'ont pas un nombre p.s. fini de points sont les processus de Poisson-Dirichlet.

Pour montrer le point *(iii)*, il suffit de remarquer qu'il s'agit de la seule possibilité parmi les MAR IPBT.

□

Nous verrons dans la partie suivante une démonstration de la change of measure formula de Pitman, démontrée à l'aide de ce théorème.

4 Description des processus de Poisson-Dirichlet à l'aide de processus ponctuels de Poisson

Pitman, Perman et Yor [9] ont relié les processus de Poisson-Dirichlet généralisés à un processus ponctuel de Poisson. Le lien est particulièrement simple pour certains choix de paramètre ($\alpha = 0$ ou $\theta = 0$) ; la loi d'un processus de Poisson-Dirichlet est alors celle d'un processus ponctuel de Poisson d'une certaine intensité dont on a renormalisé tous les points par la somme de ses points pour obtenir une probabilité aléatoire. Dans les cas intermédiaires ($\alpha \neq 0$ et $\theta \neq 0$), il faut ajouter une densité.

Soient $0 < \alpha < 1$, $\theta > -\alpha$ quelconques. Soit Ψ un processus ponctuel de Poisson sur $\mathbb{R}_+$ d'intensité $y^{-1-\alpha}\mathrm{d}y$. Il est simple, sa somme est p.s. finie, et on peut numéroter ses points par ordre décroissant. Notons $(Y_{(n)})_{n>0}$ cette numérotation. On définit la probabilité aléatoire $(P_{(n)})_{n>0}$ par $P_{(n)} = \frac{Y_{(n)}}{\int y\Psi(\mathrm{d}y)}$, et on note Φ le processus ponctuel associé (dont elle est la numérotation dans l'ordre décroissant). Alors

$$L(\Phi) := \lim_{n\to\infty} n(P_{(n)})^\alpha \tag{43}$$

existe et est finie et non nulle p.s..

Soit $\theta' > 0$. Soit Ψ' un processus ponctuel de Poisson sur $\mathbb{R}_+$ d'intensité $\frac{\theta}{y}e^{-y}\mathrm{d}y$. Il est simple, sa somme est p.s. finie, et on peut numéroter ses points par ordre décroissant. Notons $(Y'_{(n)})_{n>0}$ cette numérotation. On définit la probabilité aléatoire $(P'_{(n)})_{n>0}$ par $P'_{(n)} = \frac{Y'_{(n)}}{\int y\Psi'(\mathrm{d}y)}$, et on note Φ' le processus ponctuel associé (dont elle est la numérotation dans l'ordre décroissant).

Théorème 3 (Lien entre PD et PPP)**.** *Soit $\hat{\Phi}$ un processus de Poisson-Dirichlet de paramètres (α, θ). On a*

$$\mathbb{E}\left[f(\hat{\Phi})\right] = \mathbb{E}\left[L(\Phi)^{\theta/\alpha} f(\Phi)\right]. \tag{44}$$

Soit $\hat{\Phi}'$ un processus de Poisson-Dirichlet de paramètres $(0, \theta')$. On a

$$\mathbb{E}\left[f(\hat{\Phi}')\right] = \mathbb{E}\left[f(\Phi')\right]. \tag{45}$$

On montrera d'abord que la construction dérivée des processus de Poisson donne des MAR IPBT, avant d'identifier les paramètres.

Ce résultat provient de la remarquable similitude entre les processus ponctuels de Poisson et les processus de Poisson-Dirichlet. En effet, le Théorème de Slivnyak-Mecke assure que, conditionnellement à ce que des points donnés appartiennent au nuage, le reste du nuage est de même loi que le nuage initial. Le corollaire 1, lui, assure que, conditionnellement à ce que des points donnés (de somme inférieure à 1 strictement) appartiennent au nuage, le reste du nuage, *renormalisé pour être de somme 1*, a la loi d'un processus de Poisson-Dirichlet avec des paramètres différents, indépendamment des points choisis. C'est d'ailleurs le Théorème de Slivnyak-Mecke qui permet de montrer que les probabilités de Palm ont la même structure dans les deux constructions (à partir de processus ponctuels de Poisson, et par la construction par MAR).

4.1 Notations

On note $\Delta_k = \{(y_1, ..., y_k) \in [0,1]^k, y_1 + ... + y_k \leq 1\}$ le simplexe en dimension k.

Si Ξ est un processus ponctuel ayant ses points dans un espace polonais E, $\mathbb{M}$ l'ensemble des mesures ponctuelles de Radon sur E, $(X_k)_{k>0}$ une numérotation mesurable des points de Ξ, $k \in \mathbb{N}^*$, et $(x_1, ..., x_k) \in E^k$. On définit $M^{(k)}$ sa mesure moyenne d'ordre k et $\left(\mathcal{P}^!_{(x_1,...,x_k)}\right)_{x_1,...,x_k \in E}$ sa famille de probabilités de Palm d'ordre k comme les uniques mesures vérifiant pour tout f mesurable positive,

$$\mathbb{E}\left[\sum_{i_1,...,i_k \text{ distincts}} f(X_{i_1}, ..., X_{i_k}, \Xi \setminus \{X_{i_1}, ..., X_{i_k}\})\right]$$
$$= \int_E M^{(k)}(\mathrm{d}(x_1, ..., x_k)) \int_{\mathbb{M}} \mathcal{P}^!_{(x_1,...,x_k)}(\mathrm{d}\xi) f(x_1, ..., x_k, \xi). \quad (46)$$

Quelques fonctionnelles de processus ponctuels Pour tout processus ponctuel $\Xi = (X_n)_{n>0}$ ayant ses points dans $\mathbb{R}_+^*$, on définit sa somme

$$\Sigma(\Xi) = \int_{\mathbb{R}_+^*} x\Xi(\mathrm{d}x) \quad (47)$$

0 est p.s. le seul point d'accumulation de $(X_n)_{n>0}$ dans $\mathbb{R}_+$, car Ξ est p.s. de Radon sur $(0, \infty)$. Si $+\infty$ n'est pas un point d'accumulation, on peut numéroter ses points par ordre décroissant. On note $(X_{(n)})_{n>0}$ une telle numérotation. On définit $L(\Xi)$ comme

$$L(\Xi) = \lim_{n\to\infty} n(X_{(n)})^{\alpha}. \quad (48)$$

Si Ξ est de somme finie ($\Sigma(\Xi) < \infty$), on note $\Phi(\Xi)$ l'ensemble de ses points renormalisés par la somme $\Sigma(\Xi)$. Dans ce cas, $\Phi(\Xi)$ est p.s. de Radon sur $(0, \infty)$, donc c'est bien un processus ponctuel. De plus, $+\infty$ ne peut pas être un point d'accumulation de la suite des atomes de Ξ, donc ne peut pas non plus être un point d'accumulation de la suite des atomes de $\Phi(\Xi)$. On note $(P(\Xi)_{(n)})_{n>0}$ ses points numérotés par ordre décroissant ; on a $P(\Xi)_{(n)} = X_{(n)}/\Sigma(\Xi)$.

Un processus ponctuel de Poisson On se donne une famille $(\mathbb{P}_\alpha)_{0<\alpha<1}$ de probabilités, $(\mathbb{E}_\alpha)_{0<\alpha<1}$ les espérances associées, et un processus ponctuel Ψ tel que pour tout $0 < \alpha < 1$, sous $\mathbb{P}_\alpha$, Ψ est un processus ponctuel de Poisson (PPP) sur $\mathbb{R}_+^*$ d'intensité $\Gamma_\alpha(\mathrm{d}t) = t^{-1-\alpha}\mathrm{d}t$.

Soit $\alpha \in]0,1[$. Sous $\mathbb{P}_\alpha$, 0 est le seul point d'accumulation de Ψ, car pour tout $\epsilon > 0$, par la formule de Palm,

$$\mathbb{E}_\alpha\left[\int_\epsilon^\infty \Psi(\mathrm{d}y)\right] = \int_\epsilon^\infty y^{-1-\alpha}\mathrm{d}y < \infty. \tag{49}$$

donc p.s. il n'y a qu'un nombre fini de points supérieur à ϵ. Remarquons que, son intensité n'étant pas intégrable, il y a p.s. une infinité de points sur $\mathbb{R}_+^*$. Par conséquent, on peut se donner une numérotation $(Y_{(n)})_{n>0}$ dans l'ordre décroissant. Il est par ailleurs de somme p.s. finie (mais pas intégrable), car par la formule de Palm

$$\mathbb{E}_\alpha\left[\sum_{n>0} Y_{(n)}\mathbb{1}_{Y_n<1}\right] = \int_0^1 yy^{-1-\alpha}\mathrm{d}y < \infty \tag{50}$$

et il n'y a p.s. qu'un nombre fini de points plus grands que 1.

De même, on se donne $(\mathbb{P}^\theta)_{\theta>0}$ (et $(\mathbb{E}^\theta)_{\theta>0}$ les espérances associées) telle que sous $\mathbb{P}^\theta$, Ψ est un PPP d'intensité $\Gamma^\theta(\mathrm{d}t) = \frac{\theta}{t}e^{-t}\mathrm{d}t$. Sous $\mathbb{P}^\theta$, 0 est le seul point d'accumulation de Ψ, Ψ a un nombre infini de points, et il est de somme finie. On a en outre la propriété que $\Sigma(\Psi)$ et $\Phi(\Psi)$ sont indépendants sous $\mathbb{P}^\theta$ (voir le théorème 2.2 de [4]).

Dans les sections suivantes, lorsqu'il n'y a pas ambiguïté, Φ désigne $\Phi(\Psi)$ et Σ désigne $\Sigma(\Psi)$.

Soit $\alpha \in]0,1[$. Sous $\mathbb{P}_\alpha$, le processus $(Y_{(n)}^{-\alpha})_{k>0}$, image du processus ponctuel de Poisson $(Y_{(n)})_{n>0}$ d'intensité Γ_α par $y \mapsto y^{-\alpha}$, est un processus ponctuel de Poisson d'intensité Lebesgue sur $]0,\infty[$, numéroté par ordre croissant, donc les accroissements sont des variables aléatoires iid exponentielles de paramètre 1. Par la loi des grands nombres, $(Y_{(n)}^{-\alpha})/n \overset{n\to\infty}{\longrightarrow} 1$. $L(\Psi)$ et $L(\Phi)$ sont donc bien définie, et on a les relations suivantes :

$$\begin{cases} L(\Psi) = \lim_{n\to\infty} n(Y_{(n)})^\alpha = 1 \\ L(\Phi) = \lim_{n\to\infty} n(P(\Phi)_{(n)})^\alpha = 1/\Sigma(\Psi)^\alpha \\ \Sigma(\Psi) = L(\Phi)^{-1/\alpha}. \end{cases} \tag{51}$$

Soit Υ un processus ponctuel, $(P_n)_{n>0}$ une numérotation mesurable donnée des points de Υ (par exemple en ordre décroissant, mais pas forcément. Elle ne sera pas forcément IBPT), et $(\mathbb{P}_{\alpha,\theta})_{0\le\alpha<1,\theta>-\alpha}$ une famille de probabilités (on note toujours $(\mathbb{E}_{\alpha,\theta})_{0\le\alpha<1,\theta>-\alpha}$ les espérances associées) tels que pour $\theta > 0$,

$$\mathbb{E}_{0,\theta}\left[f(\Upsilon)\right] = \mathbb{E}^\theta\left[f(\Phi(\Psi))\right] \tag{52}$$

et pour $0 < \alpha < 1$, $\theta > -\alpha$,

$$\begin{cases} \mathbb{E}_{\alpha,0}\left[f(\Upsilon)\right] = \mathbb{E}_{\alpha}\left[f(\Phi(\Psi))\right] \\ \mathbb{E}_{\alpha,\theta}\left[f(\Upsilon)\right] = c_{\alpha,\theta}\mathbb{E}_{\alpha,0}\left[L(\Upsilon)^{\theta/\alpha}f(\Upsilon)\right] \end{cases} \tag{53}$$

avec $c_{\alpha,\theta} = 1/\mathbb{E}_{\alpha,0}\left[L(\Phi)^{\theta/\alpha}\right] = 1/\mathbb{E}_{\alpha}\left[\Sigma(\Psi)^{-\theta}\right]$ une constante de renormalisation. Notre objectif est de montrer que ces probabilités coïncident avec celles des processus de Poisson-Dirichlet de mêmes paramètres.

$c_{\alpha,\theta}$ est fini pour tout $0 < \alpha < 1$, $\theta > -\alpha$. En effet, pour tout $\epsilon > 0$, si $\theta \neq 0$:

$$\mathbb{P}_{\alpha}\left(\Sigma < \epsilon\right) \leq \mathbb{P}_{\alpha}\left(\Psi([\epsilon, \infty[) = 0\right) \tag{54}$$

$$\leq e^{\int_{\epsilon}^{\infty} y^{-1-\alpha}\mathrm{d}y} \int_{\epsilon}^{\infty} y^{-1-\alpha}\mathrm{d}y \tag{55}$$

$$= e^{-\epsilon^{-\alpha}}\epsilon^{-\alpha} \tag{56}$$

donc par la layer-cake formula,

$$\mathbb{E}_{\alpha}\left[\Sigma^{-\theta}\right] = \int_0^{\infty} \mathbb{P}_{\alpha}\left(\Sigma^{-\theta} > z\right)\mathrm{d}z \tag{57}$$

$$= \int_0^{\infty} \mathbb{P}_{\alpha}\left(\Sigma < z^{-1/\theta}\right)\mathrm{d}z \tag{58}$$

$$\leq \int_0^{\infty} e^{-z^{\alpha/\theta}} z^{\alpha/\theta}\mathrm{d}z \tag{59}$$

$$= \frac{|\theta|}{\alpha}\int_0^{\infty} e^{-u}u^{\theta/\alpha}\mathrm{d}u < \infty \tag{60}$$

et si $\theta = 0$, $c_{\alpha,\theta} = 1$.

Remarquons que si $f(\xi) = \mathbb{1}_{\xi(\mathbb{R})=\infty}$, alors sous $\mathbb{P}_{\alpha}$, $f(\Phi(\Psi)) = 1$ p.s., donc sous $\mathbb{P}_{\alpha,\theta}$, $L(\Upsilon)^{\theta/\alpha}f(\Upsilon) = L(\Upsilon)^{\theta/\alpha}$ p.s., et donc $\mathbb{P}_{\alpha,\theta}(\Upsilon \text{ infini}) = 1$: p.s. sous $\mathbb{P}_{\alpha,\theta}$, Υ est infini.

4.2 Probabilité de Palm

lemme 1 (Probabilités de Palm du processus sous $\mathbb{P}_{\alpha,\theta}$)**.** *Pour tout* $0 \leq \alpha < 1$, $\theta > -\alpha$, $k \in \mathbb{N}^*$ *et* $(p_1, ..., p_k) \in \Delta_k$, *la mesure de Palm d'ordre* k *de* Υ *sous* $\mathbb{P}_{\alpha,\theta}$ *est*

$$\mathcal{P}^!_{(\alpha,\theta)_{(p_1,...,p_k)}}(\Upsilon \in A) = \mathbb{P}_{\alpha,\theta+k\alpha}\left((1 - p_1 - ... - p_k)\Upsilon \in A\right). \tag{61}$$

Remarquons la similitude avec les probabilités de Palm des processus de Poisson-Dirichlet, présentées dans le corollaire 1. Cette similitude n'est pas anodine. C'est en effet cette forme particulière qui permettra de prouver grâce au théorème 2 que les processus ainsi construits sont des processus de Poisson-Dirichlet.

4.2.1 Ordre 1

Prouvons d'abord ce Lemme pour $k = 1$. Les probabilités de Palm d'ordre supérieur en découlent, voir le paragraphe suivant. Soit $0 < \alpha < 1$, $\theta > -\alpha$; calculons la probabilité de Palm. Pour tout f mesurable positive,

$$\mathbb{E}_{\alpha,\theta}\left[\sum_{n>0} f(P_n, \Upsilon\setminus\{P_n\})\right] \tag{62}$$

$$=C_{\alpha,\theta}\mathbb{E}_{\alpha,0}\left[L(\Upsilon)^{\theta/\alpha}\sum_{n>0} f\left(P_n, \Upsilon\setminus\{P_n\}\right)\right] \tag{63}$$

$$=C_{\alpha,\theta}\mathbb{E}_{\alpha}\left[L(\Phi)^{\theta/\alpha}\sum_{n>0} f\left(\frac{Y_n}{\Sigma}, \frac{1}{\Sigma}\left(\Psi\setminus\{Y_n\}\right)\right)\right] \tag{64}$$

$$=C_{\alpha,\theta}\mathbb{E}_{\alpha}\left[\sum_{n>0} \Sigma(\Psi)^{-\theta} f\left(\frac{Y_n}{\Sigma}, \frac{1}{\Sigma}\left(\Psi\setminus\{Y_n\}\right)\right)\right] \tag{65}$$

$$=C_{\alpha,\theta}\int M_\alpha^{(1)}(\mathrm{d}x)\int \mathcal{P}^{\Psi,!}_{\alpha,x}(\mathrm{d}\mu)\left(\frac{1}{x+\Sigma(\mu)}\right)^\theta f\left(\frac{x}{x+\Sigma(\mu)}, \frac{1}{x+\Sigma(\mu)}\mu\right) \tag{66}$$

où $M_\alpha^{(1)}$ et $\mathcal{P}^{\Psi,!}_{\alpha,x}$ sont respectivement la mesure moyenne et la probabilité de Palm associée à Ψ sous $\mathbb{P}_\alpha$. Or par le Théorème de Slivnyak-Mecke, $\mathcal{P}^{\Psi,!}_{\alpha,x} = \mathbb{P}_\alpha$. Par ailleurs, $M_\alpha^{(1)}(\mathrm{d}x) = x^{-1-\alpha}\mathrm{d}x$. Comme f est mesurable positive, on peut appliquer le Théorème de Fubini :

$$= C_{\alpha,\theta}\mathbb{E}_\alpha\left[\int x^{-1-\alpha}\mathrm{d}x\left(\frac{1}{x+\Sigma}\right)^\theta f\left(\frac{x}{x+\Sigma}, \frac{1}{x+\Sigma}\Psi\right)\right] \tag{67}$$

Faisons le changement de variable suivant :

$$y = \frac{x}{x+\Sigma} \quad\Rightarrow\quad x = \frac{y\Sigma}{1-y} \quad , \quad \mathrm{d}x = \frac{\Sigma}{(1-y)^2}\mathrm{d}y \quad , \quad \frac{1}{x+\Sigma} = \frac{1-y}{\Sigma} \tag{68}$$

Une miraculeuse séparation des variables nous simplifie alors énormément les calculs :

$$=C_{\alpha,\theta}\mathbb{E}_\alpha\left[\int \mathrm{d}y\frac{(1-y)^{1+\alpha}}{y^{1+\alpha}\Sigma^{1+\alpha}}\frac{\Sigma}{(1-y)^2}\left(\frac{1-y}{\Sigma}\right)^\theta f\left(y, \frac{1-y}{\Sigma}\Psi\right)\right] \tag{69}$$

$$=C_{\alpha,\theta}\int (1-y)^{\alpha+\theta-1}y^{-1-\alpha}\mathrm{d}y\,\mathbb{E}_\alpha\left[\Sigma^{-\theta-\alpha} f\left(y, \frac{1-y}{\Sigma}\Psi\right)\right] \tag{70}$$

$$=\int M_{\alpha,\theta}^{(1)}(\mathrm{d}y)C_{\alpha,\theta+\alpha}\mathbb{E}_\alpha\left[L(\Phi)^{(\theta+\alpha)/\alpha} f\left(y, (1-y)\Phi\right)\right] \tag{71}$$

$$=\int M_{\alpha,\theta}^{(1)}(\mathrm{d}y)C_{\alpha,\theta+\alpha}\mathbb{E}_{\alpha,0}\left[L(\Upsilon)^{(\theta+\alpha)/\alpha} f\left(y, (1-y)\Upsilon\right)\right] \tag{72}$$

$$=\int M_{\alpha,\theta}^{(1)}(\mathrm{d}y)\mathbb{E}_{\alpha,\theta+\alpha}\left[f\left(y, (1-y)\Upsilon\right)\right] \tag{73}$$

$$=\int M_{\alpha,\theta}^{(1)}(\mathrm{d}y)\int \mathcal{P}^!_{(\alpha,\theta)_y}(\mathrm{d}\mu) f(y,\mu) \tag{74}$$

avec $M_{\alpha,\theta}^{(1)}(\mathrm{d}y) = \frac{C_{\alpha,\theta}}{C_{\alpha,\theta+\alpha}}(1-y)^{\alpha+\theta-1}y^{-1-\alpha}\mathrm{d}y$, et en prenant $f(y,\mu) = g(y)\mathbb{1}_A(\mu)$ pour A un mesurable quelconque et g mesurable positive :

$$\mathcal{P}^!_{(\alpha,\theta)_y}(\Upsilon\in A) = \mathbb{P}_{\alpha,\theta+\alpha}((1-y)\Upsilon\in A) \tag{75}$$

ce qui est l'expression cherchée pour les probabilités de Palm d'ordre 1 pour $0 < \alpha < 1$.

Pour $\alpha = 0$,

$$\mathbb{E}_{0,\theta}\left[\sum_{n>0} f(P_n, \Upsilon \setminus \{P_n\})\right] \tag{76}$$

$$=\mathbb{E}^{\theta}\left[\sum_{n>0} f\left(\frac{Y_n}{\Sigma}, \frac{1}{\Sigma}(\Psi \setminus \{Y_n\})\right)\right] \tag{77}$$

$$=\int M^{\theta,(1)}(\mathrm{d}x)\int \mathcal{P}_x^{\Psi,\theta,!}(\mathrm{d}\mu) f\left(\frac{x}{x+\Sigma(\mu)}, \frac{1}{x+\Sigma(\mu)}\mu\right) \tag{78}$$

où $M^{\theta,(1)}$ et $\mathcal{P}_x^{\Psi,\theta,!}$ sont respectivement la mesure moyenne et la probabilité de Palm associée à Ψ sous $\mathbb{P}^\theta$. Or par le Théorème de Slivnyak-Mecke, $\mathcal{P}_x^{\Psi,\theta,!} = \mathbb{P}^\theta$. Par ailleurs, $M^{\theta,(1)}(\mathrm{d}x) = \frac{\theta}{x}e^{-x}\mathrm{d}x$. Comme f est mesurable positive, on peut appliquer le Théorème de Fubini :

$$= \mathbb{E}^{\theta}\left[\int \frac{\theta}{x}e^{-x}\mathrm{d}x f\left(\frac{x}{x+\Sigma}, \frac{1}{x+\Sigma}\Psi\right)\right] \tag{79}$$

Le même changement de variable donne, en rappelant que Σ et Φ sont indépendants sous $\mathbb{P}^\theta$:

$$=\mathbb{E}^{\theta}\left[\int \frac{\Sigma}{(1-y)^2}\mathrm{d}y\frac{\theta(1-y)}{y\Sigma}e^{-\frac{y\Sigma}{1-y}} f\left(y, \frac{1-y}{\Sigma}\Psi\right)\right] \tag{80}$$

$$=\int \frac{\theta}{y(1-y)}\mathbb{E}^{\theta}\left[e^{-\frac{y\Sigma}{1-y}}\right]\mathrm{d}y\mathbb{E}^{\theta}\left[f\left(y, \frac{1-y}{\Sigma}\Psi\right)\right] \tag{81}$$

$$=\int \frac{\theta}{y(1-y)}\mathbb{E}^{\theta}\left[e^{-\frac{y\Sigma}{1-y}}\right]\mathrm{d}y\mathbb{E}_{0,\theta}\left[f\left(y, \frac{1-y}{\Sigma}\Upsilon\right)\right] \tag{82}$$

$$=\int M_{0,\theta}^{(1)}(\mathrm{d}y)\int \mathcal{P}^!_{(0,\theta)_y}(\mathrm{d}\mu) f(y,\mu) \tag{83}$$

avec $M_{0,\theta}^{(1)}(\mathrm{d}y) = \frac{\theta}{y(1-y)}\mathbb{E}^\theta\left[e^{-\frac{y\Sigma}{1-y}}\right]\mathrm{d}y$, et en prenant $f(y,\mu) = g(y)\mathbb{1}_A(\mu)$ pour A un mesurable quelconque et g mesurable positive :

$$\mathcal{P}^!_{(0,\theta)_y}(\Upsilon \in A) = \mathbb{P}_{0,\theta}((1-y)\Upsilon \in A). \tag{84}$$

4.2.2 Ordre supérieur

Procédons par récurrence. Soit $k > 0$ un entier, et supposons que pour tout $(x_1, ..., x_k) \in \Delta_k$, pour tout $0 \leq \alpha < 1$, $\theta > -\alpha$,

$$\mathcal{P}^!_{(\alpha,\theta)_{(x_1,\dots,x_k)}}(\Upsilon \in A) = \mathbb{P}_{\alpha,\theta+k\alpha}((1-x_1-\dots-x_k)\Upsilon \in A). \tag{85}$$

Alors pour tout $0 \leq \alpha < 1$, $\theta > -\alpha$, pour tout f mesurable positive,

$$\mathbb{E}_{\alpha,\theta}\left[\sum_{i_1,...,i_{k+1}\text{ distincts}} f(P_{i_1},...,P_{i_{k+1}},\Upsilon\setminus\{P_{i_1},...,P_{i_{k+1}}\})\right] \tag{86}$$

$$=\mathbb{E}_{\alpha,\theta}\left[\sum_i\left(\sum_{i_2,...,i_{k+1}\neq i\text{ distincts}} f(P_i,P_{i_2},...,P_{i_{k+1}},\Upsilon\setminus\{P_i,P_{i_2},...,P_{i_{k+1}}\}\right)\right] \tag{87}$$

$$=\int M^{(1)}_{\alpha,\theta}(\mathrm{d}x_1)\mathbb{E}_{\alpha,\theta+\alpha}\left[\sum_{i_2,...,i_{k+1}\text{ distincts}} f(x_1,\bar{x}_1P_{i_2},...,\bar{x}_1P_{i_{k+1}},\bar{x}_1\Upsilon\setminus\{P_{i_2},...,P_{i_{k+1}}\})\right] \tag{88}$$

Or par hypothèse de récurrence,

$$\mathbb{E}_{\alpha,\theta+\alpha}\left[\sum_{i_2,...,i_{k+1}\text{ distincts}} f(x_1,\bar{x}_1P_{i_2},...,\bar{x}_1P_{i_{k+1}},\bar{x}_1\Upsilon\setminus\{P_{i_2},...,P_{i_{k+1}}\})\right]$$
$$=\int M^{(k)}_{\alpha,\theta+\alpha}(\mathrm{d}(y_2,...,y_{k+1}))\mathbb{E}_{\alpha,\theta+(k+1)\alpha}\left[f(x_1,\bar{x}_1y_2,...,\bar{x}_1y_{k+1},\bar{x}_1(1-y_2-...-y_{k+1})\Upsilon)\right] \tag{89}$$

Avec le changement de variable $x_j=\bar{x}_1y_j$ pour $2\leq j\leq k+1$, on obtient

$$\mathbb{E}\left[\sum_{i_1,...,i_{k+1}\text{ distincts}} f(P_{i_1},...,P_{i_{k+1}},\Upsilon\setminus\{P_{i_1},...,P_{i_{k+1}}\})\right]$$
$$=\int M^{(k+1)}_{\alpha,\theta}(\mathrm{d}(x_1,...,x_{k+1}))\mathbb{E}_{\alpha,\theta+(k+1)\alpha}\left[f(x_1,...,x_{k+1},(1-x_1-...-x_{k+1})\Upsilon)\right] \tag{90}$$

On en déduit l'hérédité, et donc le Lemme.

4.3 Identification avec un processus de Poisson-Dirichlet

Relions à présent cette construction avec les processus GEM.

Théorème 4. *Pour tout $0\leq\alpha>1$, pour tout $\theta>-\alpha$, il existe $0\leq\alpha'>1$, $\theta'>-\alpha'$ tel que sous $P_{\alpha,\theta}$, $(\tilde{P}_n)_{n>0}$ est un processus GEM de paramètres (α',θ').*

Grâce au lemme 1 donnant les probabilités de Palm du processus renormalisé, on peut appliquer le théorème 2 pour conclure directement que c'est un processus GEM, de paramètres à déterminer.

On peut aussi donner une preuve directe du résultat.

Démonstration.

$$\mathbb{E}_{\alpha,\theta}\left[f(\tilde{P}_1,...,\tilde{P}_k)g(\tilde{\Upsilon}_k^r)\right] \tag{91}$$

$$=\mathbb{E}_{\alpha,\theta}\left[\sum_{i_1,...,i_k \text{ distincts}} T(P_{i_1},...,P_{i_k})f(P_{i_1},...,P_{i_k})g\left(\frac{\Upsilon\setminus\{P_{i_1},...,P_{i_k}\}}{1-P_{i_1}-...-P_{i_k}}\right)\right] \tag{92}$$

$$=\int M_{(\alpha,\theta),k}(\mathrm{d}(p_1,...,p_k))\int \mathcal{P}^!_{(\alpha,\theta),(p_1,...,p_k)}(\mathrm{d}\mu)T(p_1,...,p_k)f(p_1,...,p_k)g\left(\frac{\mu}{1-p_1-...-p_k}\right) \tag{93}$$

$$=\int M_{(\alpha,\theta),k}(\mathrm{d}(p_1,...,p_k))T(p_1,...,p_k)f(p_1,...,p_k)\mathbb{E}_{\alpha,\theta+k\alpha}[g(\Upsilon)] \tag{94}$$

$$=\mathbb{E}_{\alpha,\theta}\left[\sum_{i_1,...,i_k \text{ distincts}} T(P_{i_1},...,P_{i_k})f(P_{i_1},...,P_{i_k})\right]\mathbb{E}_{\alpha,\theta+k\alpha}[g(\Upsilon)] \tag{95}$$

$$=\mathbb{E}_{\alpha,\theta}\left[f(\tilde{P}_1,...,\tilde{P}_k)\right]\mathbb{E}_{\alpha,\theta+k\alpha}[g(\Upsilon)] \tag{96}$$

Le passage de la première à la seconde ligne s'effectue en prenant une permutation biaisée par la taille de P. $T(P_{i_1},...,P_{i_k})$ est la probabilité sachant P que les k premiers éléments de la PBT $\tilde{P}$ de P soient $(P_{i_1},...,P_{i_k})$. Le passage de la troisième à la quatrième utilise l'expression de la probabilité de Palm démontrée précédemment.

On en déduit l'indépendance de $(\tilde{P}_1,...,\tilde{P}_k)$ et $\tilde{\Upsilon}_k^r$ pour tout $k>0$. Par la propriété caractéristique (proposition 6), comme $\tilde{P}$ est IPBT, il s'agit d'un MAR IPBT. Son nombre de points est p.s. infini, il s'agit donc d'un processus GEM, de paramètres (α',θ') à déterminer. □

4.4 Identification des paramètres

On a prouvé que la construction donnait un processus GEM, mais on ignore ses paramètres (α',θ'). Le théorème suivant comble ce manque, et conclut la preuve de la change of measure formula.

Théorème 5. *Pour tout* $0\leq\alpha<1$, $\theta>-\alpha$,

$$\alpha'=\alpha \quad , \quad \theta'=\theta. \tag{97}$$

Démonstration. Considérons la mesure moyenne. Sous $\mathbb{P}_{\alpha,\theta}$, la mesure moyenne de Υ est, pour une certaine constante D :

$$M^{(1)}_{\alpha,\theta}(\mathrm{d}y)=D\mathbb{1}_{]0,1[}(y)(1-y)^{\alpha+\theta-1}y^{-1-\alpha}\mathrm{d}y. \tag{98}$$

Or le corollaire 1 donne une formule explicite pour la mesure moyenne d'un processus de Poisson-Dirichlet de paramètre (α',θ') :

$$M^{(1)}_{PD(\alpha',\theta')}(\mathrm{d}y)=c_{1,\alpha',\theta'}\mathbb{1}_{]0,1[}(y)(1-y)^{\alpha'+\theta'-1}y^{-1-\alpha'}\mathrm{d}y \tag{99}$$

pour un certain $c_{1,\alpha',\theta'}$. Ceci permet de conclure que sous $\mathbb{P}_{\alpha,\theta}$, Υ est un processus de Poison-Dirichlet de paramètre (α,θ). □

démonstration de (99). La définition de l'échantillon biaisé par la taille et la propriété d'IPBT permettent d'exprimer facilement la mesure moyenne. Soit $(P_n)_{n>0}$ un $\mathrm{PD}(\alpha, \theta)$.

$$\mathbb{E}\left[\sum_{n>0} f(P_n)\right] = \mathbb{E}\left[\sum_{n>0} P_n \frac{f(P_n)}{P_n}\right] \tag{100}$$

$$= \mathbb{E}\left[\frac{f(P_1)}{P_1}\right] \tag{101}$$

$$= \int \frac{x^{-\alpha}(1-x)^{\alpha+\theta-1}}{\mathrm{Beta}(1-\alpha, \theta+\alpha)} \mathrm{d}x \frac{f(x)}{x} \tag{102}$$

$$= \int \frac{x^{-1-\alpha}(1-x)^{\alpha+\theta-1}}{\mathrm{Beta}(1-\alpha, \theta+\alpha)} \mathrm{d}x f(x) \tag{103}$$

$$= \int M(\mathrm{d}x) f(x) \tag{104}$$

où on passe de la première à la seconde ligne en utilisant que P_1 est un échantillon biaisé par la taille de P, via la même méthode que dans la démonstration du corollaire 1. $\square$

5 Marquages de processus de Poisson-Dirichlet

On a vu avec la propriété 6 une caractérisation des processus IPBT qui sont des MAR. On s'intéresse naturellement à savoir quels processus ponctuels plus généraux, plus forcément à valeurs dans]0,1], satisfont cette propriété caractéristique : les n premiers tirages biaisés par la taille et le nuage résiduel, convenablement renormalisé, doivent être indépendants. On ne se restreint plus à des processus ponctuels à atomes dans $\mathbb{R}_+^*$: ils prennent leurs atomes dans un polonais plus général. Il faut naturellement disposer d'une notion de "tirage biaisé par la taille" et d'une fonction de renormalisation.

La manière la plus naturelle d'y parvenir est de se donner un processus ponctuel Ψ sur un espace produit $\mathbb{R}_+^* \times E$, où E est un espace polonais quelconque, qu'on munit de sa tribu borélienne [6]. On peut se donner une numérotation mesurable $\left((\hat{P}_n, \hat{X}_n)\right)_{n>0}$. La première composante est un réel entre 0 et 1, qui représente une probabilité. Ce sera la seule qu'on renormalisera. La seconde est une marque $\hat{X}_n$ dans l'espace polonais E muni de sa tribu borélienne. On impose

$$\sum_{n>0} \hat{P}_n = 1. \tag{105}$$

Le processus $\hat{P}$ est une probabilité aléatoire. On définit la permutation biaisée par la taille du processus marqué Ψ comme la permutation de $(\hat{P}_n, \hat{X}_n)_{n>0}$ par une permutation biaisée par la taille adaptée à $\hat{P}$. Autrement dit, on détermine d'abord une permutation à l'aide de la composante des probabilités $\hat{P}$, puis on permute les deux composantes en

6. On aura besoin de cette hypothèse dans la preuve de la propriété caractéristique généralisée, et pour disposer d'une numérotation mesurable.

même temps suivant cette permutation. Cette notion de PBT étend naturellement celle définie précédemment, et vérifie notamment toutes les propriétés de la section 1.5.

Il est facile de se convaincre que si les $(X_n)_{n>0}$ sont iid et indépendants de P et que P est un MAR IPBT, le processus marqué satisfera la propriété caractéristique, convenablement généralisée aux processus ponctuels marqués. Le Théorème 6 affirme qu'en réalité, on ne peut observer que ce cas.

Soit Ψ un processus ponctuel sur $\mathbb{R}_+^* \times E$, avec E polonais muni de sa tribu borélienne, et $\int_{\mathbb{R}_+^* \times E} p\Psi(\mathrm{d}(p,x)) = 1$ p.s.. On se donne $((P_n, X_n))_{n>0}$ une numérotation mesurable et IPBT de Ψ. On définit sur $\{K \geq n\}$:

$$\begin{cases} V_n = \dfrac{P_n}{1 - P_1 - \ldots - P_{n-1}} \\ \tilde{\Psi}_n^r = \{(\bar{V}_{n+1}\ldots\bar{V}_{k-1}V_k, X_k), k > n\}. \end{cases} \tag{106}$$

et $V_n = 1$ et $\tilde{\Psi}_n^r = 0$ la mesure nulle sur $\{K < n\}$.

Théorème 6 (Marquages satisfaisant la propriété caractéristique)**.** *Les propriétés suivantes sont équivalentes :*

(i) Pour tout n, (V_1, X_1), ..., (V_n, X_n) et $\tilde{\Psi}_n^r$ sont tous indépendants.

(ii) Pour tout n, (V_n, X_n) et $\tilde{\Psi}_n^r$ sont indépendants.

(iii) P est un MAR IPBT, et il existe Q une probabilité sur $(E, \mathcal{E})$ telle que les $(X_n)_{n>0}$ soient iid indépendants de P et de loi Q.

Remarque 2.

- *Dans le cas où on n'écrit pas les marques, on retrouve la définition de la section 2.*
- *Lorsqu'on représente une probabilité aléatoire comme un processus ponctuel, on a vu qu'on excluait le point 0. C'est la même raison qui préside au choix de l'espace $]0,1] \times E$.*
- *Un marquage de processus ponctuel $U = (P_n, X_n)$ est dit iid lorsque les $(X_n)_{n>0}$ sont iid conditionnellement à $(P_n)_{n>0}$. Un marquage de MAR IPBT qui vérifie la proposition est, lui, un marquage iid dont la loi conditionnelle de X par rapport à P ne dépend pas de P. C'est donc bien plus fort : les $(X_n)_{n>0}$ sont iid* et indépendants de $(P_n)_{n>0}$.

En conséquence de (iii), tout processus ponctuel marqué qui vérifie la propriété caractéristique aura même loi qu'un MAR IPBT réel de même loi marqué de manière iid selon la probabilité Q.

Par analogie avec le cas standard, on appellera (V_n, X_n) le n^{e} tirage, et $\tilde{\Psi}_n^r$ le nuage résiduel renormalisé. Pour tout $n > 0$, on note $\mathbb{V}_n$ la loi de V_n.

On constate encore une fois le caractère contraignant de la propriété d'IPBT. Celui-ci réside principalement sur les contraintes qu'il impose sur les lois des P_n et des V_n, un changement de loi à un indice se répercutant sur toutes les autres après PBT. Celles-ci sont rendues visibles par le Lemme 4. Elles s'appliquent sur la loi des tirages, et non sur le nuage : l'IPBT seule n'impose rien du tout sur la loi du processus ponctuel, puisque n'importe quelle numérotation mesurable de ce processus, après PBT, devient IPBT. On ne peut ainsi l'exploiter qu'en présence de contraintes supplémentaires sur les tirages, par exemple la propriété d'indépendance des tirages successifs énoncée dans les propriétés caractéristiques.

Remarquons qu'il existe des marquages non dénués d'intérêt qui ne vérifient pas la propriété caractéristique. On peut par exemple considérer un marquage IPBT. Rien n'impose alors que le marquage soit iid : le marquage $X_n = P_n$ est IPBT sans être iid. Il s'agit ici de caractériser ceux qui vérifient la propriété caractéristique, puisqu'elle permet d'expliciter très facilement les probabilités de Palm.

Démonstration. (i) $\Rightarrow$ (ii) est immédiat.

(ii) $\Rightarrow$ (i) se prouve par récurrence, de la même manière que dans la preuve de la propriété 6. Le point délicat consiste à remarquer qu'on peut oublier la numérotation sans perdre d'information. Pour cela, il suffit de constater que par construction, la loi de $((\frac{P_k}{1-P_1-...-P_n}, X_k))_{k>n}$ sachant Ψ et $P_1, ..., P_n, X_1, ..., X_n$ ne dépend que de $\tilde{\Psi}_n^r$, et est celle d'une permutation aléatoire de $\tilde{\Psi}_n^r$.

(iii) $\Rightarrow$ (i) : Si P est un MAR IPBT, les $(V_n)_{0<n\leq m}$ sont indépendantes, où $m = K$ est le nombre (constant p.s. par le Théorème 1, éventuellement infini) de P_k non nuls du MAR IPBT. Les marques $(X_n)_{0<n\leq m}$ sont iid de loi Q et indépendantes de P. Ceci implique (i).

(i) $\Rightarrow$ (iii) est le Vpoint délicat. Pour le montrer, on constate d'abord que (i) implique que la suite $((V_n, X_n))_{0<n\leq m}$ est indépendante. En particulier, $(V_n)_{n>0}$ est indépendante, donc P est un MAR IPBT. On définit ensuite Q_n une version régulière de la loi conditionnelle de X_n conditionnellement à V_n de la manière suivante :

$$\mathbb{E}\left[f(V_n, X_n)\right] = \int \mathbb{V}_n(\mathrm{d}v) \int Q_n(v, \mathrm{d}x) f(v, x). \tag{107}$$

Un tel objet existe toujours, car E est polonais muni de sa tribu borélienne. Enfin, le Lemme 4, dont l'énoncé est reporté en annexe, exploite la propriété d'IPBT du processus. Couplé avec (i), il permettra de relier les lois conditionnelles de X_n et X_{n+1}, et de montrer que la loi conditionnelle Q_n de X_n conditionnellement à V_n ne dépend pas de V_n ni de n, donc X_n est indépendant de V_n et de même loi pour tout n. Comme on a l'indépendance entre les couples (V_n, X_n), on en déduit l'indépendance de V et X, et que les X_n sont iid. C'est cette dernière étape que nous détaillons ici.

Étape 1 : une équation reliant Q_n et Q_{n+1}. Pour tout $n > 0$, f mesurable positive, et $\mathcal{X} \in \mathcal{E}$, appliquons la deuxième partie du Lemme 4 (voir annexe) à $g(p_1, ..., p_n) = f(\frac{p_{n-1}}{1-p_1-...-p_{n-2}}, \frac{p_n}{1-p_1-...-p_{n-1}})$:

$$\mathbb{E}\left[T(\sigma) f(V_n, V_{n+1}) \mathbb{1}_{\mathcal{X}}(X_n)\right] = \mathbb{E}\left[T(\sigma) f(\bar{V}_n V_{n+1}, \frac{V_n}{1 - \bar{V}_n V_{n+1}}) \mathbb{1}_{\mathcal{X}}(X_{n+1})\right]. \tag{108}$$

(i) assure l'indépendance de (V_n, X_n) et (V_{n+1}, X_{n+1}). L'égalité (108) devient :

$$\int \mathbb{V}_n(\mathrm{d}v_n)\mathbb{V}_{n+1}(\mathrm{d}v_{n+1})T(v_n, v_{n+1})f(v_n, v_{n+1})Q_n(v_n, \mathcal{X})$$
$$= \int \mathbb{V}_n(\mathrm{d}v_n)\mathbb{V}_{n+1}(\mathrm{d}v_{n+1})T(v_n, v_{n+1})f(\bar{v}_n v_{n+1}, \frac{v_n}{1-\bar{v}_n v_{n+1}})Q_{n+1}(v_{n+1}, \mathcal{X}) \quad (109)$$

avec

$$T(v_n, v_{n+1}) = \int \mathbb{V}_1(\mathrm{d}v_1)...\mathbb{V}_{n-1}(\mathrm{d}v_{n-1})T(\sigma). \quad (110)$$

En particulier, pour tout g mesurable positive, si $\mathcal{X} = E$:

$$\int \mathbb{V}_n(\mathrm{d}v_n)\mathbb{V}_{n+1}(\mathrm{d}v_{n+1})T(v_n, v_{n+1})g(v_n, v_{n+1})$$
$$= \int \mathbb{V}_n(\mathrm{d}v_n)\mathbb{V}_{n+1}(\mathrm{d}v_{n+1})T(v_n, v_{n+1})g(\bar{v}_n v_{n+1}, \frac{v_n}{1-\bar{v}_n v_{n+1}}) \quad (111)$$

La dépendance particulière de (V_n, X_n) permet d'appliquer (111) à $g(v, w) = f(v, w)Q_n(v, \mathcal{X})$:

$$\int \mathbb{V}_n(\mathrm{d}v_n)\mathbb{V}_{n+1}(\mathrm{d}v_{n+1})T(v_n, v_{n+1})f(v_n, v_{n+1})Q_n(v_n, \mathcal{X})$$
$$= \int \mathbb{V}_n(\mathrm{d}v_n)\mathbb{V}_{n+1}(\mathrm{d}v_{n+1})T(v_n, v_{n+1})f(\bar{v}_n v_{n+1}, \frac{v_n}{1-\bar{v}_n v_{n+1}})Q_n(\bar{v}_n v_{n+1}, \mathcal{X}) \quad (112)$$

Si on prend $f : (x, y) \mapsto h(\bar{x}y, \frac{x}{1-\bar{x}y})/T(x, y)$, où h est une application mesurable positive quelconque (c'est bien défini car $T > 0$ p.s.) , on a

$$T(v_n, v_{n+1})f(\bar{v}_n v_{n+1}, \frac{v_n}{1-\bar{v}_n v_{n+1}}) = h(v_n, v_{n+1}) \quad (113)$$

d'où, en combinant les équations (109) et (112) :

$$\int \mathbb{V}_n \otimes \mathbb{V}_{n+1}(\mathrm{d}(v_n, v_{n+1}))h(v_n, v_{n+1})Q_{n+1}(v_{n+1}, \mathcal{X})$$
$$= \int \mathbb{V}_n \otimes \mathbb{V}_{n+1}(\mathrm{d}(v_n, v_{n+1}))h(v_n, v_{n+1})Q_n(\bar{v}_n v_{n+1}, \mathcal{X}) \quad (114)$$

donc pour tout $\mathcal{X} \in \mathcal{E}$, pour $\mathbb{V}_n \otimes \mathbb{V}_{n+1}$-presque tout (v_n, v_{n+1}),

$$Q_n(\bar{v}_n v_{n+1}, \mathcal{X}) = Q_{n+1}(v_{n+1}, \mathcal{X}). \quad (115)$$

Étape 2 : Constance de Q_n. Il s'agit à présent d'exploiter cette égalité essentielle, qui contient toute l'information dont on a besoin. On se donne une suite $(\mathcal{X}_k)_{k>0}$ dans $\mathcal{E}$ stable par intersections finies qui engendre la tribu $\mathcal{E}$. L'existence découle du caractère polonais de E.

Rappelons que K est p.s. constant. On suppose $K > 1$, le cas $K = 1$ étant immédiat. Soit $0 < n < K$. Il existe $\mathcal{V}$ tel que $\mathbb{V}_n \otimes \mathbb{V}_{n+1}(\mathcal{V}) = 1$ et tel que (115) soit vraie pour tout $(v_n, v_{n+1}) \in \mathcal{V}$, pour tout $\mathcal{X}_k$ (donc pour tout $\mathcal{X} \in \mathcal{E}$ par le Lemme de classes monotones).

P est un MAR IPBT. Le Théorème 1 assure donc que la mesure $\mathbb{V}_m$ ne peut prendre que deux formes pour $m \geq 2$: un Dirac, ou une mesure absolument continue par rapport à Lebesgue et telle que Lebesgue sur [0,1] est absolument continue par rapport à $\mathbb{V}_m$ (en fait c'est une loi Beta, mais ces deux propriétés sont suffisantes). Traitons ces deux cas séparément.

Cas 1 : $\mathbb{V}_{n+1} = \delta_w$. Cela implique que pour tout $k > 0$, p.s.

$$Q_n(\bar{V}_n V_{n+1}, \mathcal{X}_k) = Q_{n+1}(w, \mathcal{X}_k) =: Q(\mathcal{X}_k). \tag{116}$$

Reprenons alors l'équation (112) appliquée à $f : (v, x) \mapsto h(v)$, pour h mesurable positive quelconque et $\mathcal{X} = \mathcal{X}_k$, k quelconque.

$$\int \mathbb{V}_n(\mathrm{d}v_n) T(v_n, w) h(v_n) Q_n(v_n, \mathcal{X}_k) \tag{117}$$

$$= \int \mathbb{V}_n(\mathrm{d}v_n) \mathbb{V}_{n+1}(\mathrm{d}v_{n+1}) T(v_n, v_{n+1}) h(v_n) Q_n(v_n, \mathcal{X}_k) \tag{118}$$

$$= \int \mathbb{V}_n(\mathrm{d}v_n) \mathbb{V}_{n+1}(\mathrm{d}v_{n+1}) T(v_n, v_{n+1}) h(\bar{v}_n v_{n+1}) Q_n(\bar{v}_n v_{n+1}, \mathcal{X}_k) \tag{119}$$

$$= \int \mathbb{V}_n(\mathrm{d}v_n) \mathbb{V}_{n+1}(\mathrm{d}v_{n+1}) T(v_n, v_{n+1}) h(\bar{v}_n v_{n+1}) Q(\mathcal{X}_k) \tag{120}$$

$$= \int \mathbb{V}_n(\mathrm{d}v_n) \mathbb{V}_{n+1}(\mathrm{d}v_{n+1}) T(v_n, v_{n+1}) h(v_n) Q(\mathcal{X}_k) \tag{121}$$

$$= \int \mathbb{V}_n(\mathrm{d}v_n) T(v_n, w) h(v_n) Q(\mathcal{X}_k) \tag{122}$$

Le passage de la deuxième à la troisième ligne et de la quatrième à la cinquième provient de (112), celui de la troisième à la quatrième de (116). Comme $T > 0$ $\mathbb{V}_n \otimes \mathbb{V}_{n+1}$-p.s., on conclut que pour $\mathbb{V}_n$-presque tout v, $Q_n(v, \mathcal{X}_k)$ est constante p.s. égale à $Q(\mathcal{X}_k)$, et donc que pour $\mathbb{V}_n$-presque tout v, $Q_n(v, \cdot) = Q(\cdot)$. Naturellement, c'est également vrai pour Q_{n+1}, puisque V_{n+1} est p.s. constant. Cela implique que X_n et X_{n+1} sont indépendants de V_n et V_{n+1} respectivement.

Cas 2 : $\mathbb{V}_{n+1} \ll \mathcal{L}$ et $\mathcal{L} \ll \mathbb{V}_{n+1}$. On prouve le même résultat de constance de Q_n et Q_{n+1} :

$\mathbb{V}_n \otimes \mathbb{V}_{n+1}(\mathcal{V}) = 1 = \mathcal{L}(\mathcal{V})$, où $\mathcal{L}$ est la mesure de Lebesgue restreinte à [0,1]. Il existe donc $\mathcal{W} \subset]0, 1]$ de mesure de Lebesgue 1 tel que pour tout $w \in \mathcal{W}$, il existe $\mathcal{V}_w \subset]0, 1]$ de mesure de Lebesgue 1 tel que pour tout $v \in \mathcal{V}_w$,

$$Q_n(\bar{v}w, \cdot) = Q_{n+1}(w, \cdot), \tag{123}$$

c'est-à-dire il existe $\mathcal{N}_w \subset]0, w]$ Lebesgue-négligeable tel que pour tout $x \in]0, w] \setminus \mathcal{N}_w$,

$$Q_n(x, \cdot) = Q_{n+1}(w, \cdot). \tag{124}$$

Soit $w_0 > 0 \in \mathcal{W}$. Pour tout $w \in \mathcal{W}$, pour tout $x \in [0, \min(w, w_0)] \setminus (\mathcal{N}_w \cup \mathcal{N}_{w_0})$,

$$Q(\cdot) := Q_{n+1}(w_0, \cdot) = Q_n(x, \cdot) = Q_{n+1}(w, \cdot). \tag{125}$$

Cet ensemble est non vide car de mesure de Lebesgue strictement positive. On a donc la constance p.s. de Q_{n+1} par rapport à $\mathcal{L}$, donc $\mathbb{V}_{n+1}$ par absolue continuité. On peut ensuite trouver une suite $(w_n)_{n>0}$ croissant vers 1 dans $\mathcal{W}$; en notant $\mathcal{N} = \bigcup_{k>0} \mathcal{N}_{w_k}$ (toujours négligeable), on a pour tout $x \in]0, 1[\setminus\mathcal{N}$,

$$Q(\cdot) = Q_n(x, \cdot). \tag{126}$$

D'où la constance p.s. de Q_n par rapport à $\mathcal{L}$, donc $\mathbb{V}_n$.

Conclusion. Les (V_n, X_n) sont indépendants par (i). On a montré que pour tout $0 < n \leq m$, X_n est indépendant de V_n de loi Q, puisque la loi conditionnelle Q_n de X_n par rapport à V_n ne dépendait pas de V_n. Ceci montre que les $(X_n)_{0<n\leq m}$ sont iid de loi Q et indépendant de P, d'où la conclusion.

□

6 Annexes

6.1 Dynamiques qui laissent un processus IPBT invariant en loi

6.1.1 Tirage de k tirages BT

Soit P une probabilité aléatoire, et $k \in \mathbb{N}^*$. Soient $(Q_1, ..., Q_k)$ les k premiers tirages d'une PBT de P, définis en loi par la définition 3. On pose pour $j > k$:

$$Q_j = (j-k)^{\mathrm{e}} \text{ élément de } (P_n)_{n>0} \setminus \{Q_1, ..., Q_k\}. \tag{127}$$

On note $D^{(k)}(P)$ la loi de $Q = (Q_j)_{j>0}$.

lemme 2 (Invariance d'une probabilité aléatoire IPBT sous $D^{(k)}$)**.** *Si P une probabilité aléatoire IPBT, alors Q est une probabilité aléatoire IPBT, de même loi que P : $D^{(k)}(P)$ est égale à la loi de P.*

Démonstration. Q est obtenue en ajoutant k tirages au début de la suite de tirages iid dans la construction par tirage avec remise (voir section 2), et en conditionnant à ce que ces k tirages soient distincts (sur l'évènement $K \geq k$, ceci ne change pas la loi ; sur son complémentaire, le reste de la suite est nulle, et Q est bien une PBT.). □

Une démonstration intuitive repose sur la construction par tirage avec remise. On peut assez facilement visualiser cette opération avec l'autre construction. Reprenons donc l'image du mélange d'un jeu de carte biaisé présentée dans la première section. On représente la probabilité aléatoire P par un jeu de carte infini où la carte numéro 1 est en dessous du paquet, et où la n^e carte porte le poids P_n. On a la possibilité de tirer une carte dans le tas avec une probabilité proportionnelle à son poids ; si tous les poids sont nuls, on tire n'importe quelle carte (par exemple la première). En effectuant cette opération une infinité de fois, on construit un nouveau jeu de carte dont les poids, dans leur ordre d'apparition, forment la PBT de P. L'opération présentée dans ce Lemme consiste à s'arrêter après k tirages dans la construction de la PBT, et à ajouter le tas de destination, de k cartes, en-dessous du tas source.

On dispose d'un autre résultat sur ces dynamiques. Pour k fixé, en répétant cette opération un nombre infini de fois à une probabilité P non nécessairement IPBT, on convergera en loi vers la PBT de P.

lemme 3 (Convergence du mélange par $D^{(k)}$ vers la PBT)**.** *Pour tout $k > 0$, $(D^{(k)})^n(P)$ converge en loi quand $n \to \infty$ vers $D^{(\infty)}(P) = PBT(P)$.*

Démonstration. Nous ne le prouvons ici que pour $k = 1$, et dans le cas simple pour alléger les calculs. Le cas non simple se traite exactement de la même manière, avec des A_j en plus dans les calculs. Notons P^n la probabilité aléatoire obtenue en appliquant n fois l'opération $D^{(k)}$ à P. Soient $j > 0$ et $n_1, ..., n_j$ distincts deux à deux ; sur $\{P_{n_1} > 0, ..., P_{n_j} > 0\}$:

$$\begin{aligned}&\mathbb{P}\left((P^n_1, ..., P^n_j) = (P_{n_1}, ..., P_{n_j})|P\right)\\ &\qquad = \sum_{n=u_1>...>u_j>0} P_{n_j}(P_{n_1} + ... + P_{n_{j-1}})^{u_{k-1}-u_k-1}...P_{n_2}P_{n_1}^{n-u_1-1}P_{n_1} \end{aligned} \tag{128}$$

En effet, cet évènement est réalisé si, pour tout $1 \leq j \leq k$, on tire au moins une fois P_{n_j}, et on ne tire plus que $P_{n_1}, ..., P_{n_{j-1}}$ après l'avoir tiré pour la dernière fois, et au moins une fois chacun d'entre eux. L'expression sommée ci-dessus correspond à une disjonction de cas en fonction des u_j, le rang du dernier tirage de P_{n_j}. Faisons le changement de variable $v_j = n - u_j$. Sur $\{P_{n_1} > 0, ..., P_{n_j} > 0\}$:

$$\mathbb{P}\left((P^n_1, ..., P^n_j) = (P_{n_1}, ..., P_{n_j})|P\right) \tag{129}$$

$$= \sum_{0=v_1<...<v_j<n} P_{n_j}(P_{n_1} + ... + P_{n_{j-1}})^{v_k-v_{k-1}-1}...P_{n_2}P_{n_1}^{v_1-1}P_{n_1} \tag{130}$$

$$\underset{n\to\infty}{\longrightarrow} \sum_{0=v_1<...<v_j} P_{n_j}(P_{n_1} + ... + P_{n_{j-1}})^{v_k-v_{k-1}-1}...P_{n_1}P_{n_1}^{v_1-1} \tag{131}$$

$$= P_{n_j}...P_{n_1} \sum_{0=v_1<...<v_j} (P_{n_1} + ... + P_{n_{j-1}})^{v_k-v_{k-1}-1}...P_{n_1}^{v_1-1} \tag{132}$$

$$= P_{n_j}...P_{n_1}\frac{1}{1 - P_{n_1} - ... - P_{n_{j-1}}}...\frac{1}{1 - P_{n_1}} \tag{133}$$

$$= \mathbb{P}\left((P^\infty_1, ..., P^\infty_j) = (P_{n_1}, ..., P_{n_j})|P\right) \tag{134}$$

où P^∞ est une PBT de P.

P^n est une probabilité aléatoire. Or $[0,1]^{\mathbb{N}^*}$ est compact, donc la suite $(P_n)_{n>0}$ est tendue. Elle converge en loi vers Q, qui n'est pas forcément une probabilité aléatoire. Or pour tout f mesurable bornée, pour tout j,

$$\begin{aligned}
&\mathbb{E}\left[f(P_1^n,...,P_j^n)\right] \\
&\qquad = \sum_{n_1,...,n_j \text{ distincts}} \mathbb{E}\left[f(P_{n_1},...,P_{n_j})\mathbb{P}\left((P_1^n,...,P_j^n)=(P_{n_1},...,P_{n_j})|P\right)\right]
\end{aligned} \tag{135}$$

Comme

$$\begin{aligned}
&f(P_{n_1},...,P_{n_j})\mathbb{P}\left((P_1^n,...,P_j^n)=(P_{n_1},...,P_{n_j})|P\right) \\
&\qquad = f(P_{n_1},...,P_{n_j})\mathbb{1}_{P_{n_1}>0,...,P_{n_j}>0}\mathbb{P}\left((P_1^n,...,P_j^n)=(P_{n_1},...,P_{n_j})|P\right) \\
&+\sum_{u=1}^{j-1} f(P_{n_1},...,P_{n_u},0,...,0)\mathbb{1}_{P_{n_1}>0,...,P_{n_u}>0,P_{n_{u+1}}=0}\mathbb{P}\left((P_1^n,...,P_u^n)=(P_{n_1},...,P_{n_u})|P\right)
\end{aligned} \tag{136}$$

converge p.s. vers $f(P_{n_1},...,P_{n_j})\mathbb{P}\left((P_1^\infty,...,P_j^\infty)=(P_{n_1},...,P_{n_j})|P\right)$ et est dominé par $f(P_{n_1},...,P_{n_j})$ intégrable, par le Théorème de convergence dominée,

$$\mathbb{E}\left[f(P_1^n,...,P_j^n)\right] \underset{n\to\infty}{\longrightarrow} \mathbb{E}\left[f(P_1^\infty,...,P_j^\infty)\right]. \tag{137}$$

D'où on déduit par unicité la convergence en loi vers $Q \overset{(d)}{=} P^\infty$ (qui est bien une probabilité aléatoire) la permutation biaisée par la taille de P. □

6.1.2 Permutation aléatoire finie des composantes

Nous présentons et démontrons dans cette section un Lemme qui permet de relier les lois des différents poids d'une probabilité aléatoire IPBT. Il est essentiel pour obtenir la "rigidité" des MAR IPBT, moyennant des hypothèses structurelles supplémentaires.

lemme 4 (Invariance sous des permutations finie aléatoires)**.** *Soit $U=(U_n)_{n>0}$, avec $U_n=(P_n,X_n)$ un processus IPBT tel que le K défini dans la proposition 5 est constant p.s.. Soit $2\le k\le K$ un entier, Σ un sous-groupe de Σ_k l'ensemble des permutations de $\{1,...,k\}$, et U' un processus tel que, sachant U, $U'_n=U_n$ pour tout $n>k$, et pour tout $\sigma\in\Sigma_k$:*

$$(U'_1,...,U'_k)=(U_{\sigma(1)},...U_{\sigma(k)}) \text{ avec probabilité } T(\sigma) \text{ sachant } P \tag{138}$$

où

$$\begin{cases}
g(\sigma)=A_{\sigma(1)}P_{\sigma(1)}\dfrac{A_{\sigma(2)}P_{\sigma(2)}}{1-P_{\sigma(1)}}\dfrac{A_{\sigma(3)}P_{\sigma(3)}}{1-P_{\sigma(1)}-P_{\sigma(2)}}\cdots\dfrac{A_{\sigma(k)}P_{\sigma(k)}}{1-P_{\sigma(1)}-...-P_{\sigma(k-1)}} \\
T(\sigma)=\dfrac{\mathbb{1}_{\sigma\in\Sigma}g(\sigma)}{\sum_{\nu\in\Sigma}g(\nu)}.
\end{cases} \tag{139}$$

avec A_j comme dans la définition 3. Alors U' est IPBT, de même loi que U.

On en déduit, pour toute f mesurable positive, si $\sigma = (k \;\; k-1)$ est une transposition, et $\Sigma = \{\sigma, id\}$:

$$\mathbb{E}\left[f(U_1, ..., U_{k-2}, U_{k-1}, U_k)T(\sigma)\right] = \mathbb{E}\left[f(U_1, ..., U_{k-2}, U_k, U_{k-1})T(\sigma)\mathbb{1}_{P_1+...+P_{k-1}<1}\right] \quad (140)$$

où $T(\sigma)$ prend dans ce cas particulier la forme

$$T(\sigma) = \frac{\frac{A_{k-1}}{1 - P_1 - ... - P_{k-2} - P_k}}{\frac{A_{k-1}}{1 - P_1 - ... - P_{k-2} - P_k} + \frac{A_k}{1 - P_1 - ... - P_{k-2} - P_{k-1}}}. \quad (141)$$

Ce Lemme traduit le fait qu'un processus IPBT est invariant sous l'opération consistant à lui appliquer une permutation finie, choisie avec une certaine probabilité dans un groupe de permutation donné. L'hypothèse que l'ensemble dans lequel on prend la permutation est un groupe interviendra dans la démonstration, mais ce n'est pas simplement une hypothèse pour simplifier les calculs : elle est essentielle. En effet, si on pouvait prendre un singleton distinct de $\{id\}$ pour l'ensemble Σ dans lequel on prend les permutations, en prenant les transpositions, on aurait l'invariance par transposition, et donc tous les U_n auraient la même loi. Ce n'est bien sûr pas le cas. Par ailleurs, remarquons que la probabilité de choisir une permutation donnée n'est pas quelconque : à renormalisation près, elle est égale à la probabilité d'observer cette permutation des poids dans la PBT de U.

On peut se passer de l'hypothèse portant sur la constance de K en conditionnant par rapport à K. Un processus IPBT conditionné à K reste IPBT, on peut alors appliquer le Lemme.

Démonstration du Lemme. Dans cette démonstration, on ne se restreindra pas au cas simple.

Soit $k \geq 2$, Σ un sous-groupe de Σ_k, $\hat{U} = (U_{(n)})_{n>0}$ la probabilité aléatoire obtenue en triant les composantes de U dans l'ordre décroissant. On suppose pour simplifier que U est à valeur dans $\mathbb{R}$, c'est-à-dire qu'il n'y a que la composante correspondant à la probabilité aléatoire. Soit σ une permutation de $\mathbb{N}^*$. Sur l'évènement $\{\hat{U}_{\sigma(1)} > 0, ...\hat{U}_{\sigma(k)} > 0\}$:

$$\begin{aligned}
&\mathbb{P}\left((U'_1, ..., U'_k) = (\hat{U}_{\sigma(1)}, ..., \hat{U}_{\sigma(k)})|\hat{U}\right) \\
&\quad = \sum_{\nu\in\Sigma} \mathbb{P}\left((U'_1, ..., U'_k) = (U_{\nu(1)}, ..., U_{\nu(k)}), (U_1, ..., U_k) = (\hat{U}_{\sigma\nu^{-1}(1)}, ..., \hat{U}_{\sigma\nu^{-1}(k)})|\hat{U}\right) \quad (142)
\end{aligned}$$

Conditionnons par rapport à U. Comme $\hat{U}$ est une fonction déterministe de U :

$$\begin{aligned}
&\mathbb{P}\left((U'_1, ..., U'_k) = (U_{\nu(1)}, ..., U_{\nu(k)})|\hat{U}, U\right) \\
&\qquad = \mathbb{P}\left((U'_1, ..., U'_k) = (U_{\nu(1)}, ..., U_{\nu(k)})|U\right) = T(\sigma) \quad (143)
\end{aligned}$$

Donc

$$\mathbb{P}\left((U'_1,...,U'_k)=(\hat{U}_{\sigma(1)},...,\hat{U}_{\sigma(k)})|\hat{U}\right) \tag{144}$$

$$=\sum_{\nu\in\Sigma}\mathbb{E}\left[\frac{g(\nu)}{\sum_{\mu\in\Sigma}g(\mu)}\mathbb{1}_{(U_1,...,U_k)=(\hat{U}_{\sigma\nu^{-1}(1)},...,\hat{U}_{\sigma\nu^{-1}(k)})}|\hat{U}\right] \tag{145}$$

$$=\sum_{\nu\in\Sigma}\frac{\hat{g}(\sigma)}{\sum_{\mu\in\Sigma}\hat{g}(\sigma\nu^{-1}\mu)}\mathbb{P}\left((U_1,...,U_k)=(\hat{U}_{\sigma\nu^{-1}(1)},...,\hat{U}_{\sigma\nu^{-1}(k)})|\hat{U}\right) \tag{146}$$

$$=\hat{g}(\sigma)\sum_{\nu\in\Sigma}\frac{\hat{g}(\sigma\nu^{-1})}{\sum_{\mu\in\Sigma}\hat{g}(\sigma\nu^{-1}\mu)} \tag{147}$$

où $\hat{g}$ est obtenu en remplaçant U par $\hat{U}$ dans sa définition. On peut expliciter les probabilités ainsi car $k\leq K$ (donc pour tout $\nu\in\Sigma$, $U_{\sigma\nu^{-1}(1)}+...+U_{\sigma\nu^{-1}(k-1)}<1$) et U est IPBT, donc est égal en loi à une PBT de $\hat{U}$ (ce fait provient du choix judicieux de $\hat{U}$, voir section 1). Comme Σ est un groupe, la somme ci-dessus vaut 1, donc

$$\mathbb{P}\left((U'_1,...,U'_k)=(\hat{U}_{\sigma(1)},...,\hat{U}_{\sigma(k)})|\hat{U}\right)=\hat{g}(\sigma) \tag{148}$$

ce qui prouve par l'équation (3) que U' est une PBT de $\hat{U}$. Comme U est IPBT, U' est IPBT et de même loi.

Par l'égalité des lois, pour $\Sigma=\{\sigma,\mathrm{id}\}$ avec $\sigma=(k\ \ k-1)$:

$$\begin{aligned}\mathbb{E}&\left[f(U'_1,...,U'_{k-1},U'_k)(\mathbb{1}_{(U'_{k-1},U'_k)=(U_k,U_{k-1})}+\mathbb{1}_{(U'_{k-1},U'_k)=(U_{k-1},U_k)})\right]\\&=\mathbb{E}\left[f(U_1,...,U_{k-1},U_k)(\mathbb{1}_{(U'_{k-1},U'_k)=(U_k,U_{k-1})}+\mathbb{1}_{(U'_{k-1},U'_k)=(U_{k-1},U_k)})\right]\end{aligned} \tag{149}$$

donc

$$\mathbb{E}\left[f(U_1,...,U_k,U_{k-1})\mathbb{1}_{(U'_{k-1},U'_k)=(U_k,U_{k-1})}\right]=\mathbb{E}\left[f(U_1,...,U_{k-1},U_k)\mathbb{1}_{(U'_{k-1},U'_k)=(U_k,U_{k-1})}\right] \tag{150}$$

Or

$$\mathbb{E}\left[\mathbb{1}_{(U'_{k-1},U'_k)=(U_k,U_{k-1})}|U\right]=T(\sigma) \tag{151}$$

d'où l'égalité du Lemme. L'expression de $T(\sigma)$ découle de sa définition. □

6.2 Conséquences

Proposition 8 (Supports des poids d'une probabilité aléatoire IBPT)**.** *Si P est une probabilité aléatoire IPBT tel que K est constant p.s., alors pour tout $n,m\leq K$,*

$$Supp(P_n)=Supp(P_m), \tag{152}$$

et pour tout $n>K$, $Supp(P_n)=\{0\}$.

Corollaire 2 (Supports des poids d'une probabilité aléatoire IPBT)**.** *Si P est une probabilité aléatoire IPBT quelconque, alors*

$$\mathit{Supp}(P_1)\cap]0,1[=\mathit{Supp}(P_2)\cap]0,1[. \tag{153}$$

et pour tout $n \geq m$,

$$\mathit{Supp}(P_n)\cap]0,1] \subset \mathit{Supp}(P_m)\cap]0,1]. \tag{154}$$

Démonstration de la propriété. Par le Lemme 4, pour tout $0 < n < K$, pour toute f mesurable positive, si $\sigma = (n \;\; n+1)$ est une transposition, et $\Sigma = \{\sigma, \mathrm{id}\}$:

$$\mathbb{E}\left[f(U_n, U_{n+1})T(\sigma)\right] = \mathbb{E}\left[f(U_{n+1}, U_n)T(\sigma)\right]. \tag{155}$$

Par la proposition 5, p.s. $U_{n+1} > 0$ et $U_n > 0$, donc $0 < T(\sigma) < 1$.

On applique l'équation (155) à $f(x,y) = \mathbb{1}_A(x)$ pour A borélien quelconque de [0,1] :

$$\mathbb{E}\left[\mathbb{1}_{U_n \in A}T(\sigma)\right] = \mathbb{E}\left[\mathbb{1}_{U_{n+1} \in A}T(\sigma)\right]. \tag{156}$$

Soit $x \in \mathrm{Supp}(U_{n+1})$. Pour tout $\epsilon > 0$, prenons $A = I_\epsilon =]x-\epsilon, x+\epsilon[\cap[0,1]$. On a alors $\mathbb{P}(U_{n+1} \in I_\epsilon) = \mathbb{E}\left[\mathbb{1}_{U_{n+1} \in I_\epsilon}\right] > 0$. On en déduit que $\mathbb{E}\left[\mathbb{1}_{U_{n+1} \in I_\epsilon}T(\sigma)\right] > 0$, puis que $\mathbb{E}\left[\mathbb{1}_{U_n \in I_\epsilon}T(\sigma)\right] > 0$, donc que $\mathbb{P}(U_n \in I_\epsilon) > 0$. Ceci étant vrai pour tout ϵ, on a bien

$$\mathrm{Supp}(U_{n+1}) \subset \mathrm{Supp}(U_n). \tag{157}$$

La même méthode en sens inverse donne l'inclusion réciproque dans le cas où P est une probabilité aléatoire sur $\mathbb{N}^*$ entier, et on a donc égalité des supports à tout rang. □

Démonstration du corollaire. Soit k tel que $\mathbb{P}(K = k) > 0$. Conditionnellement à $K = k$, P est toujours une probabilité aléatoire IPBT, avec K constant p.s. égal à k. On peut appliquer la proposition 8 : si on note $\mathrm{Supp}_k(P_n)$ le support de P_n conditionnellement à $K = k$, alors pour tout $0 < m \leq n \leq k$,

$$\mathrm{Supp}_k(P_n)\cap]0,1] = \mathrm{Supp}_k(P_m)\cap]0,1], \tag{158}$$

et pour $n > k$,

$$\emptyset = \mathrm{Supp}_k(P_n)\cap]0,1] \subset \mathrm{Supp}_k(P_m)\cap]0,1]. \tag{159}$$

Or $\mathrm{Supp}(P_n) = \overline{\bigcup_{k>0} \mathrm{Supp}_k(P_n)}$. On en déduit la seconde partie du corollaire. Pour la première, observons que $\mathrm{Supp}_1(P_1) = \{1\}$, donc pour tout $k > 0$,

$$\mathrm{Supp}_k(P_1)\cap]0,1[= \mathrm{Supp}_k(P_2)\cap]0,1[\tag{160}$$

d'où la première partie du corollaire. □

Proposition 9 (Cas où le premier poids est constant). *Soit P un processus IPBT. Si P_1 est constant p.s., alors il existe $m \in \mathbb{N}^*$ tel que $P_k = 1/m$ pour $1 \leq k \leq m$, et $P_k = 0$ pour $k > m$.*

Démonstration. Posons p la valeur de P_1. Le Lemme 2 assure que pour tout n, $P_n \in \{0, p\}$. On en déduit que

$$1 = \sum_{i>0} P_i = p\, |\{i : P_i \neq 0\}| \tag{161}$$

donc $p \neq 0$, et p.s. Card $\{i : P_i \neq 0\} = 1/p =: m \in \mathbb{N}^*$. Par la proposition 5, les m premiers termes sont non nuls, et les suivants le sont. □

Remerciements

Je remercie Bartłomiej Błaszczyszyn pour ses conseils et sa relecture attentive de ce rapport.

Références

[1] Richard Arratia, Andrew D Barbour, and Simon Tavaré. *Logarithmic combinatorial structures : a probabilistic approach.* European Mathematical Society, 2003.

[2] Wray Buntine and Marcus Hutter. A bayesian view of the poisson-dirichlet process. *arXiv :1007.0296v2*, 2012.

[3] W.J. Ewens. Population genetics theory - the past and the future. In Sabin Lessard, editor, *Mathematical and Statistical Developments of Evolutionary Theory*, volume 299 of *NATO ASI Series*, pages 177–227. Springer Netherlands, 1990.

[4] Shui Feng. *The Poisson-Dirichlet Distribution and Related Topics - Models and Asymptotic Behaviors.* Springer Science & Business Media, 2010.

[5] Thomas S Ferguson. A bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2) :209–330, 1973.

[6] Kenji Handa. The two-parameter poisson-dirichlet point process. *Bernoulli*, 15(4) :1082–1116, 2009.

[7] Jennie C. Hansen. Order statistics for decomposable combinatorial structures. *Random Structures and Algorithms*, 5(4) :517–533, 1994.

[8] John FC Kingman. Random discrete distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–22, 1975.

[9] Mihael Perman, Jim Pitman, and Marc Yor. Size-biased sampling of poisson point processes and excursions. *Probability Theory and Related Fields*, 92(1) :21–39, 1992.

[10] Jim Pitman. Random discrete distributions invariant under size-biased permutation. *Advances in Applied Probability*, 28 :525–539, 1996.

[11] Jim Pitman and Marc Yor. The two-parameter poisson-dirichlet distribution derived from a stable subordinator. *The Annals of Probability*, 25(2) :855–900, 1997.

[12] GA Watterson. The stationary distribution of the infinitely-many neutral alleles diffusion model. *Journal of Applied Probability*, pages 639–651, 1976.