

Théorèmes limites pour des processus de branchement et de coalescence spatiaux

Amandine Véber

Thèse sous la direction d'Alison Etheridge et Jean-François Le Gall

Ce travail est composé de deux grandes parties : l'étude de processus de branchement en milieu aléatoire et celle de généalogies de populations structurées spatialement.

1 De l'art de piéger des souris

Imaginons une population de souris, se déplaçant dans un espace à leur disposition et se reproduisant et mourant au bout d'un certain temps. On place dans cet espace des pièges, dans lesquels les souris qui passent sont capturées avec une certaine probabilité et on considère qu'une souris attrapée disparaît de la population. Pour finir, supposons que les pièges sont disposés de manière assez homogène dans l'espace. La question à laquelle on s'intéresse est la suivante : avec quelle efficacité capture-t-on les souris ?¹

Plus formellement, considérons tout d'abord un nombre fini de particules (d'individus, de souris, ...) qui, indépendamment les unes des autres :

- se déplacent dans \mathbb{R}^d en suivant une trajectoire brownienne ;
- « branchent » à taux 1, i.e., meurent au bout d'un temps distribué selon une loi exponentielle de paramètre 1 et laissent à leur emplacement un nombre aléatoire K (pouvant être égal à 0) de descendants qui évoluent ensuite indépendamment suivant le même mécanisme.

La loi du nombre de descendants est la même pour toutes les particules et on supposera dans la suite que sa moyenne vaut 1 et sa variance, notée σ^2 , est finie. La figure 1 illustre le mécanisme d'évolution de la population.

On représente l'état à l'instant $t \geq 0$ de cette population par sa mesure empirique Z_t , somme des masses de Dirac en la position au temps t de

¹Malgré l'offense manifeste à l'intelligence des souris que cela constitue, on supposera que celles-ci ne voient pas les obstacles.

chacune des particules en vie à cet instant. Autrement dit, $Z_t(A)$ donne le nombre de particules présentes dans $A \subset \mathbb{R}^d$ à l'instant t . On obtient ainsi un processus markovien $\{Z_t, t \geq 0\}$ à valeurs dans l'espace des mesures ponctuelles sur \mathbb{R}^d , appelé *mouvement brownien branchant*.

Remarque 1. *Notons au passage que la taille totale de la population forme ici un processus de Galton-Watson en temps continu. Celui-ci étant critique (on a supposé que le nombre moyen d'enfants d'un individu valait 1), la population s'éteint donc en temps fini presque sûrement (voir par exemple [AN72]).*

Le *supermouvement brownien* standard est obtenu en faisant tendre le nombre n de particules initiales vers l'infini tout en divisant la masse totale de la population par n (pour garder une masse finie) et en renormalisant le temps et l'espace de sorte que chaque particule suive toujours la trajectoire d'un mouvement brownien, mais branche à taux n . Plus précisément, si $(\{Z_t^{(n)}, t \geq 0\})_{n \geq 1}$ est une suite de mouvements browniens branchants de même mécanisme de branchement et tels que $Z_0^{(n)}$ est constitué de n particules, on pose pour tout $t \geq 0$:

$$X_t^{(n)} := \frac{1}{n} \sum_{i \sim nt} \delta_{\xi_{nt}^{i,n}/\sqrt{n}}, \quad (1)$$

où $i \sim nt$ signifie que la particule de $Z^{(n)}$ indexée par i est en vie à l'instant nt et $\xi_{nt}^{i,n} \in \mathbb{R}^d$ est sa position au temps nt . Si à présent on fait tendre n vers l'infini et qu'on suppose que $X_0^{(n)}$ converge vers une mesure finie μ , alors un résultat classique assure que la suite de processus $(X^{(n)})_{n \geq 1}$ converge en loi vers un processus markovien X à valeurs dans les mesures finies sur \mathbb{R}^d et de valeur initiale μ , appelé *supermouvement brownien*. Ce dernier peut donc être vu comme une collection infinie de particules qui se déplacent suivant des mouvements browniens et branchent « à taux infini » en donnant naissance à un nombre de descendants d'espérance 1 et de variance finie.

Introduisons à présent les types de pièges auxquels on s'intéresse. Dans les deux problèmes décrits ci-dessous, les zones dangereuses seront placées autour de points aléatoires donnés par un processus homogène sur \mathbb{R}^d (voir la figure 1).

1.1 Pièges instantanés

On considère ici le cas où la probabilité d'être capturé en passant dans un piège vaut 1. Par ailleurs, on fait dépendre la distribution des pièges d'un

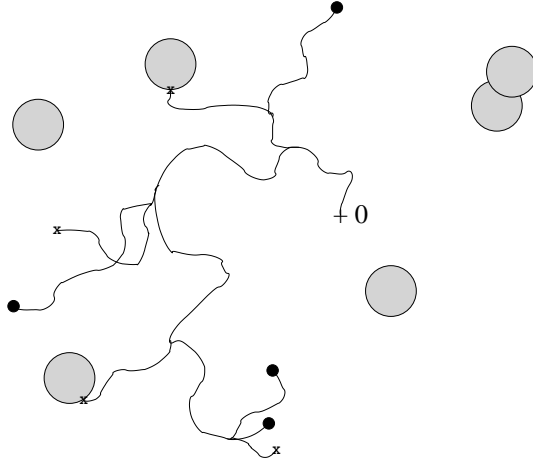


FIG. 1 – Les pièges sont représentés par les zones grisées. On s’intéresse à la famille issue d’une particule initialement en 0, qu’on laisse évoluer jusqu’à un temps $t > 0$. Les croix marquent l’emplacement où une particule disparaît de la population avant le temps t parce qu’elle rencontre un piège ou parce qu’elle branche en ne donnant naissance à aucun enfant, tandis que les points noirs représentent les particules en vie à l’instant t .

paramètre ε de telle sorte que lorsque ε tend vers 0, le rayon des pièges tend vers 0 mais leur densité tend vers l’infini.

Fixons donc $d \geq 2$ (la dimension 1 a peu d’intérêt ici) et une constante $c > 0$. Pour tout $\varepsilon \in (0, 1/2)$, définissons l’ensemble de pièges Γ_ε par :

$$\Gamma_\varepsilon := \bigcup_{x \in \mathcal{P}^\varepsilon} \bar{B}(x, \varepsilon),$$

où $\bar{B}(x, \varepsilon)$ est la boule fermée de centre x et de rayon ε , et \mathcal{P}^ε est une collection aléatoire de points de \mathbb{R}^d telle que le nombre moyen de points contenus dans un compact K est donné par $c \text{Vol}(K) \log(\varepsilon^{-1})$ si $d = 2$ et $c \text{Vol}(K) \varepsilon^{2-d}$ si $d \geq 3$. On notera \mathbf{P} la loi de $\{\Gamma_\varepsilon, \varepsilon \in (0, 1/2)\}$.

À Γ_ε fixé, définissons maintenant le processus $X^{(\Gamma_\varepsilon)}$ comme un supermouvement brownien de déplacement sous-jacent la loi d’un mouvement brownien tué instantanément dans Γ_ε . Pour finir, notons $k_2 = \pi$ et k_d la capacité newtonienne de la boule unité si $d \geq 3$, et appelons X^* le supermouvement brownien de déplacement sous-jacent la loi d’un mouvement brownien tué

à taux $k_d c$ (indépendant des obstacles). Le résultat principal du chapitre 2 de [V09] est un théorème de convergence qui établit que pour toute mesure finie μ :

1. Si $(\varepsilon_n)_{n \geq 1}$ décroît suffisamment rapidement vers 0, alors avec \mathbf{P} -probabilité 1 la loi de $X^{(\Gamma_{\varepsilon_n})}$ partant de μ converge lorsque $n \rightarrow \infty$ vers celle de X^* partant de μ (phénomène d'homogénéisation *quenched*, i.e., à réalisation des pièges fixée).
2. Sans condition de décroissance sur ε , la loi de $X^{(\Gamma_\varepsilon)}$ partant de μ converge en \mathbf{P} -probabilité lorsque $\varepsilon \rightarrow 0$ vers celle de X^* partant de μ .

Lorsque $\varepsilon \rightarrow 0$, les pièges deviennent donc denses dans l'espace et les particules du supermouvement brownien tuées dans Γ_ε finissent par en sentir l'effet de manière homogène : le taux de capture à la limite dépend de c mais plus de la disposition précise des pièges.

Remarque 2. *On peut faire varier la densité de pièges dans l'espace en remplaçant la constante c par une fonction $c : \mathbb{R}^d \rightarrow \mathbb{R}_+$ mesurable et bornée. Le nombre moyen de points de \mathcal{P}_ε contenu dans un compact K est alors donné par $\varepsilon^{2-d} \int_K c(x) dx$ (en dimension $d \geq 3$) et le processus limite est un supermouvement brownien dont les particules sont tuées à taux $k_d c(x)$.*

1.2 Pièges à retardement²

Supposons à présent que les captures ne soient pas instantanées, mais qu'elles se produisent après un certain temps passé dans les pièges. Prenons également des pièges d'une forme un peu plus générale, définis de la manière suivante : on fixe un ensemble compact $K \subset \mathbb{R}^d$ ($d \geq 1$ cette fois) et on pose

$$\Gamma := \bigcup_{x \in \mathcal{P}} (x + K),$$

où \mathcal{P} est une collection aléatoire de points de \mathbb{R}^d telle que le nombre moyen de points de \mathcal{P} tombant dans un compact C est égal au volume de C . On appelle κ la probabilité que $0 \in \Gamma$.

À Γ et $\varepsilon \geq 0$ fixés, définissons le processus $Z^{\Gamma, \varepsilon}$ comme un mouvement brownien branchant dont les particules se déplacent dans \mathbb{R}^d suivant la loi d'un mouvement brownien tué à taux ε dans Γ . Autrement dit, en laissant de côté le phénomène de branchement, la probabilité qu'un tel mouvement brownien survive jusqu'au temps t s'écrit

$$p(\Gamma) := \mathbf{E} \left[e^{-\varepsilon \int_0^t \mathbf{1}_\Gamma(\xi_s) ds} \right],$$

²Travail en collaboration avec J.-F. Le Gall (Univ. Paris-Sud 11).

où ξ est un mouvement brownien standard (et Γ est fixé dans l'espérance).

On obtient alors à nouveau une homogénéisation spatiale de l'effet des pièges : si on définit le processus $X^{\Gamma,\varepsilon}$ comme dans (1), en remplaçant $1/n$ par ε , on obtient un mouvement brownien branchant renormalisé dont les particules sont en outre tuées à taux 1 dans $\sqrt{\varepsilon} \Gamma$. On montre alors qu'à Γ fixé (pour presque tous les Γ), la loi de $X^{\Gamma,\varepsilon}$ partant de δ_0 converge lorsque $\varepsilon \rightarrow 0$ vers celle d'un supermouvement brownien X^∞ indépendant de Γ , de déplacement sous-jacent la loi d'un mouvement brownien tué partout à taux κ .

Ce résultat est l'un des ingrédients principaux du théorème central du chapitre 3 de [V09], qui donne le comportement asymptotique *quenched* de la probabilité que l'une des particules de $Z^{\Gamma,\varepsilon}$ (partant d'une seule particule à l'origine) réussisse à sortir de $\bar{B}(0,R)$ lorsque $R \rightarrow \infty$ et $\varepsilon \rightarrow 0$. Cette asymptotique s'exprime en fonction de la solution d'une équation aux dérivées partielles avec conditions aux bords infinies, liée à des probabilités d'atteinte pour X^∞ (ces liens sont détaillés par exemple dans [LG99]).

2 Arbre généalogique des souris piégées

Dans cette deuxième partie, on utilise les pièges pour échantillonner des souris à différents endroits de l'espace où vit la population. On cherche alors à décrire la forme que prend l'arbre généalogique de l'échantillon d'individus obtenu en fonction de l'emplacement des pièges et de la manière dont la population se renouvelle au fil du temps.

Pour ce faire, on considère plusieurs types de modèles de population structurée en espace : dans le paragraphe 2.1, on suppose que les souris sont réparties en des sous-populations discrètes reliées par des migrations d'individus d'une communauté à une autre ; au contraire, dans le paragraphe 2.2 l'espace dans lequel elles évoluent est continu (on choisira par exemple \mathbb{R}^d , voire \mathbb{R}^2 en pratique).

Remarque 3. *Bien qu'il n'en soit pratiquement pas fait mention dans la suite, la motivation principale de cette étude est de comprendre comment évolue la diversité génétique au sein de la population. On s'intéresse à un gène porté par les souris et qui peut prendre plusieurs formes (i.e., qui possède plusieurs allèles). Puisque les descendants héritent de l'allèle de leur parent (parfois avec des erreurs, des mutations), une bonne compréhension des relations généalogiques entre les individus fournit des indications sur les mécanismes d'évolution ayant conduit à la répartition allélique observée dans l'échantillon.*

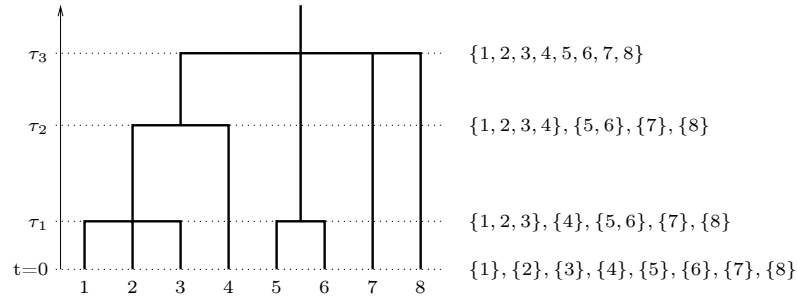


FIG. 2 – Réalisation possible d’un coalescent pour $n = 8$: on part de la partition triviale au temps 0, puis au temps τ_1 on observe une première transition au cours de laquelle $\{1\}, \{2\}$ et $\{3\}$ fusionnent en un seul bloc et $\{5\}, \{6\}$ en un autre. Les singletons restants sont inchangés. L’évolution continue jusqu’au temps τ_3 auquel le processus atteint la valeur finale $\{1, \dots, 8\}$.

Introduisons en préambule la manière dont nous décrirons ces relations ancestrales. Nous supposerons dans la suite que chaque individu n’a qu’un seul parent. Lorsque la population n’a pas de structure spatiale, l’objet couramment utilisé pour décrire la généalogie de n individus est le *coalescent*. Formellement, il s’agit d’un processus aléatoire à valeurs dans les partitions de $\{1, \dots, n\}$, dont la valeur initiale est la partition triviale en singletons et dont les blocs fusionnent à un certain taux jusqu’à ce que le processus atteigne la valeur absorbante $\{1, \dots, n\}$. Si l’état du coalescent à l’instant $t \geq 0$ est la partition $\{a_1, \dots, a_k\}$, alors on interprète chaque bloc a_i comme contenant les numéros de tous les individus de l’échantillon dont l’*ancêtre commun le plus récent* (le père pour des frères, le grand-père pour des cousins, etc.) vivait au plus t unités de temps dans le passé. Sur la figure 2 par exemple, le processus ancestral à tout temps $t \in [\tau_1, \tau_2)$ est égal à $\{\{1, 2, 3\}, \{4\}, \{5, 6\}, \{7\}, \{8\}\}$ puisque les individus de l’échantillon qu’on avait appelés 1, 2 et 3 ont leur premier ancêtre commun τ_1 unités de temps dans le passé, les lignées ancestrales de 5 et 6 fusionnent aussi à cet instant, mais les autres individus n’ont pas encore d’ancêtre en commun si on remonte t unités de temps dans le passé. La figure 2 montre également qu’on peut voir la réalisation d’un coalescent comme un arbre.

Lorsque la population a une structure spatiale, celle-ci influence la manière dont ses membres se reproduisent (en fonction du mécanisme de reproduction qu’on a spécifié). Par exemple, on peut imposer que les descendants

d'un individu naissent toujours dans la même communauté que leur parent, auquel cas deux lignées ancestrales ne pourront fusionner que si elles appartiennent à la même sous-population. Par conséquent, il faut également garder en mémoire la position géographique des ancêtres de l'échantillon. On utilise donc des partitions marquées de $\{1, \dots, n\}$: au temps t , chaque bloc de la partition ancestrale contient à nouveau le numéro des individus ayant un ancêtre commun vivant t unités de temps dans le passé et la marque d'un bloc donne la position spatiale de l'ancêtre correspondant à cet instant. C'est à ces coalescents spatiaux que nous nous intéressons dans ce qui suit.

2.1 Population structurée en communautés discrètes³

Dans le chapitre 4 de [V09], on se place dans le cadre du *modèle d'îles de Wright* : la population totale est répartie sur D îles (ou communautés) connectées entre elles par des migrations d'individus et des événements d'*extinction massive* au cours desquels une fraction des îles voit sa population disparaître et être remplacée par les descendants d'un certain nombre d'individus choisis dans la population totale. Un mécanisme de reproduction intra-île assure en outre une évolution locale de chaque communauté. On étudie alors le processus retraçant la généalogie de n individus lorsque le nombre D d'îles tend vers l'infini.

Généralisant des résultats de Wakeley (voir en particulier [W99]), on obtient une séparation d'échelles de temps entre l'évolution des processus généalogiques intra-îles très rapides et ceux retraçant la généalogie d'individus échantillonnés dans des communautés différentes, bien plus lents. On établit également des conditions sur les mécanismes de reproduction et de migration pour que le processus ancestral sur l'échelle de temps la plus lente converge en loi vers un processus aléatoire sans structure spatiale appelé *coalescent à collisions multiples et simultanées* (on renvoie à [S00] pour une définition précise de cet objet). Sous ces conditions, lorsque la population est disséminée sur un très grand nombre d'îles l'arbre généalogique d'un nombre fini d'individus échantillonnés dans des communautés distinctes ne dépend donc plus des positions spatiales précises de leurs ancêtres.

2.2 Population répartie dans un espace continu⁴

³Travail en collaboration avec J.E. Taylor (Arizona State Univ.)

⁴Travail en collaboration avec N.H. Barton (biologiste à l'IST Austria) et A.M. Etheridge (Univ. of Oxford).

L'objectif de cette dernière partie est d'étudier un nouveau modèle de population vivant dans un espace continu, introduit par N. Barton et A. Etheridge en 2008. Dans ce modèle, les individus portent un allèle, transmis de parent à descendants, et sont répartis sur \mathbb{R}^2 (ou plus généralement sur \mathbb{R}^d). La population à l'instant $t \geq 0$ est représentée par une fonction Φ_t qui associe à chaque site de \mathbb{R}^2 une mesure de probabilité sur l'espace des allèles possibles. Informellement, $\Phi_t(x)$ donne la distribution de l'allèle d'un individu qu'on échantillonnerait au site x à l'instant t . Un processus ponctuel de Poisson d'événements en dicte alors l'évolution : si (t, x, r, u) est un point du processus ponctuel, l'événement se produit au temps t et dans la boule fermée centrée en x et de rayon r . On échantillonne un allèle k suivant la distribution moyenne des allèles au temps $t-$ dans $\bar{B}(x, r)$ (i.e., on choisit uniformément un parent dans la zone affectée) et à chaque site de cette boule une fraction u de la population est remplacée par les descendants de ce parent, tous d'allèle k . Autrement dit, pour tout $y \in \bar{B}(x, r)$ on pose

$$\Phi_t(y) = (1 - u)\Phi_{t-}(y) + u\delta_k.$$

Les autres sites de \mathbb{R}^2 ne sont pas affectés par l'événement.

Dans un premier temps, on montre l'existence et l'unicité en loi du *processus Λ -Fleming-Viot spatial* $\{(\Phi_t(x), x \in \mathbb{R}^2)\}_{t \geq 0}$. On étudie ensuite les généalogies associées à ce modèle, en supposant que la population évolue sur le tore $\mathbb{T}(L) \subset \mathbb{R}^2$ de côté L . Des travaux de Cox & Griffeath et Zähle, Cox & Durrett sur des marches aléatoires coalescentes suggèrent que lorsque le rayon des événements est borné, le processus généalogique de n individus échantillonnés uniformément sur $\mathbb{T}(L)$ converge en loi lorsque $L \rightarrow \infty$ vers un *coalescent de Kingman* non-spatial (cf. [ZCD05]). Afin d'obtenir un comportement limite plus riche, on ajoute à ces *petits* événements des événements de taille $\mathcal{O}(L^\alpha)$ ($\alpha \in (0, 1]$) beaucoup plus rares, modélisant une extinction partielle de la population d'une région étendue (à la suite d'un feu de forêt ou d'une glaciation, par exemple) suivie d'une recolonisation de cette zone par les descendants de quelques individus « chanceux ». On en tire alors des processus limites variés, mais qui presque tous manifestent également une perte de la structure spatiale à la limite.

Références

[AN72] K.B. Athreya et P.E. Ney. *Branching processes*. Springer, 1972.

- [LG99] J.-F. Le Gall. *Spatial branching processes, random snakes and partial differential equations*. Lectures in Mathematics ETH Zürich. Birkhäuser, 1999.
- [S00] J. Schweinsberg. Coalescents with simultaneous multiple collisions. *Electron. J. Probab.*, 5 :1–50, 2000.
- [V09] A. Véber. *Théorèmes limites pour des processus de branchement et de coalescence spatiaux*. Thèse de doctorat, Univ. Paris-Sud 11, 2009.
- [W99] J. Wakeley. Nonequilibrium migration in human history. *Genetics*, 153 :1863–1871, 1999.
- [ZCD05] I. Zähle, J.T. Cox et R. Durrett. The stepping stone model II : genealogies and the infinite sites model. *Ann. Appl. Probab.*, 15 :671–699, 2005.