Divertissements probabilistes – Corrigé

0 – Exercice du TD 13 à chercher

Exercice 0. (**Sommes aléatoires**) Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles indépendantes de même loi, centrées, de variance σ^2. On pose $S_n = \sum_{m=1}^n X_m$. Soit $(N_k)_{k\geq 1}$ une suite de v.a. à valeurs dans $\mathbb{N}_*$, toutes indépendantes de la suite $(X_n)_{n\geq 1}$. On pose finalement

$$Z_k = \frac{1}{\sqrt{N_k}} S_{N_k}.$$

On suppose que $N_k \to \infty$ p.s. lorsque $k \to \infty$. Montrer que Z_k converge en loi vers une variable aléatoire que l'on déterminera.

Corrigé :

Posons $Y_n = S_n/\sqrt{n}$ et soit $\mathcal{N}$ une variable aléatoire normale centrée réduite. Nous allons montrer que Z_k converge en loi vers $\mathcal{N}$. Soit $F : \mathbb{R} \to \mathbb{R}_+$ une fonction continue bornée. Fixons $\epsilon > 0$. D'après le théorème central limite, Y_n converge en loi vers $\mathcal{N}$. Il existe donc N tel que pour $n \geq n_0$:

$$|\mathbb{E}[F(Y_n)] - \mathbb{E}[F(\mathcal{N})]| \leq \epsilon.$$

On écrit ensuite :

$$\begin{aligned} \mathbb{E}[F(Z_k)] &= \mathbb{E}\left[F(Y_{N_k})\right] = \sum_{j=1}^{\infty} \mathbb{E}\left[F(Y_j)\mathbb{1}_{\{N_k=j\}}\right] \\ &= \sum_{j=1}^{\infty} \mathbb{P}[N_k = j]\,\mathbb{E}\left[F(Y_j)\right] \qquad \text{par indépendance de } N_k \text{ et } Y_j \end{aligned}$$

Ainsi :

$$\begin{aligned} |\mathbb{E}[F(Z_k)] - \mathbb{E}[F(\mathcal{N})]| &\leq \sum_{j=1}^{\infty} \mathbb{P}[N_k = j]\left|\mathbb{E}\left[F(Y_j)\right] - \mathbb{E}[F(\mathcal{N})]\right| \\ &\leq 2\|F\|_\infty \mathbb{P}[N_k \leq n_0] + \sum_{j\geq n_0} \mathbb{P}[N_k = j]\left|\mathbb{E}\left[F(Y_j)\right] - \mathbb{E}[F(\mathcal{N})]\right| \\ &\leq 2\|F\|_\infty \mathbb{P}[N_k \leq n_0] + \sum_{j\geq n_0} \mathbb{P}[N_k = j]\epsilon \\ &\leq 2\|F\|_\infty \mathbb{P}[N_k \leq n_0] + \epsilon. \end{aligned}$$

Or N_k converge presque sûrement vers 0, et donc en probabilité (ou utiliser le théorème de convergence dominer pour le faire à la main). Le premier terme de la dernière somme converge donc vers 0. On en déduit que $|\mathbb{E}[F(Z_k)] - \mathbb{E}[F(\mathcal{N})]| \leq 2\epsilon$ pour k suffisamment grand, ce qui conclut. $\square$

1 – Processus de Galton-Watson

Soit $\mu = (\mu_i)_{i \geq 0}$ une mesure de probabilité sur $\{0, 1, 2, \dots\}$ telle que $\mu_0 + \mu_1 < 1$. Le processus de Galton-Watson $(Z_n)_{n \geq 0}$ est définie récursivement par $Z_0 = 1$ et, pour $n \geq 0$:

$$Z_{n+1} = \sum_{j=1}^{Z_n} X_j^{(n)},$$

où les variables aléatoires $(X_j^{(n)})_{j,n \geq 0}$ sont i.i.d. de loi μ. Ainsi, $(Z_n)_{n \geq 0}$ modélise l'évolution d'une population dont à chaque instant n les individus meurent en donnant naissance à des nombres d'enfants i.i.d de loi μ. On introduit la fonction génératrice ϕ associée à μ :

$$\phi(s) = \sum_{n \geq 0} \mu_n s^n, \qquad s \in [0, 1].$$

On note finalement $m = \phi'(1)$ la moyenne du nombre d'enfants. Le but est de démontrer le théorème suivant :

Théorème (Probabilité d'extinction).

Presque sûrement, il existe $n \geq 0$ tel que $Z_n = 0$ si et seulement si $m \leq 1$.

(0) Prouver que ϕ est strictement croissante, que ϕ' est strictement croissante et que $\phi(1) = 1$.

(i) Pour $s \in [0, 1]$, on note $\phi_n(s) = \mathbb{E}\left[s^{Z_n}\right]$. Montrer que $\phi_{n+1} = \phi_n \circ \phi$.

(ii) Soit $T = \min\{n \geq 0; Z_n = 0\}$. Montrer que $\mathbb{P}(T < \infty)$ est le plus petit point fixe de ϕ. Conclure.

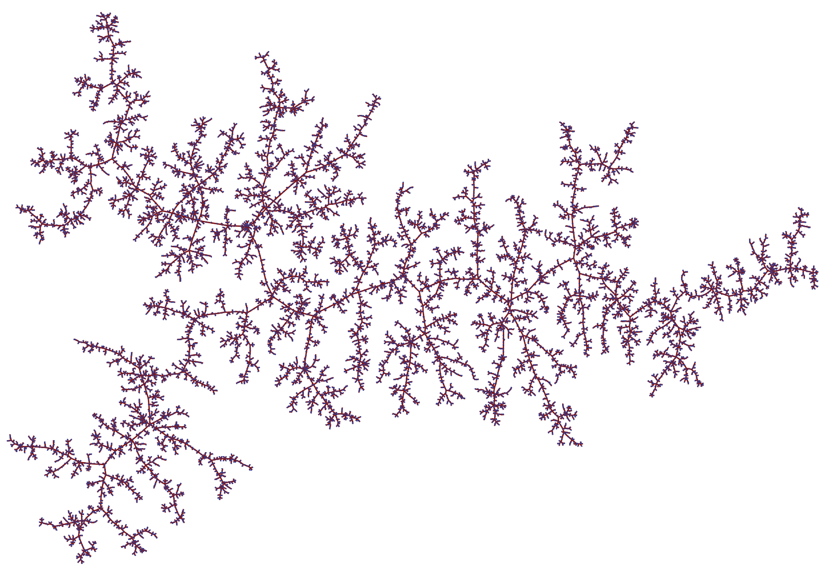


Figure 1 – Simulation d'un grand arbre généalogique du processus de Galton-Watson (les sommets correspondent aux individus, et deux individus sont reliés par une arête si l'un a donné naissance à l'autre).

Corrigé :

(0) Pas de difficulté en dérivant sous le signe somme.

(i) On démontre aisément par récurrence sur n que pour tout entier $n \geq 1$, Z_n et $(X_j^{(k)}; 0 \leq j, 0 \leq k \leq n)$ sont indépendants. Ainsi :

$$\begin{aligned}
\phi_{n+1}(s) &= \mathbb{E}\left[s^{\sum_{j=1}^{Z_n} X_j^{(n)}}\right] \\
&= \sum_{k=0}^{\infty} \mathbb{E}\left[s^{\sum_{j=1}^{k} X_j^{(n)}} \mathbb{1}_{\{Z_n = k\}}\right] \\
&= \sum_{k=0}^{\infty} \mathbb{P}[Z_n = k]\, \mathbb{E}\left[s^{\sum_{j=1}^{k} X_j^{(n)}}\right] \quad \text{car } Z_n \text{ et } (X_1^{(n)}, \ldots, X_k^{(n)}) \text{ sont indépendants} \\
&= \sum_{k=0}^{\infty} \mathbb{P}[Z_n = k]\, \mathbb{E}\left[s^{X_1^{(n)}}\right]^k \quad \text{car } X_1^{(n)}, \ldots, X_k^{(n)} \text{ sont iid} \\
&= \sum_{k=0}^{\infty} \mathbb{P}[Z_n = k]\, \phi(s)^k \\
&= \phi_n(\phi(s)).
\end{aligned}$$

D'où le résultat.

(ii) La probabilité d'extinction $\mathbb{P}(T < \infty)$ est la réunion croissante des événements $\{Z_n = 0\}$. En remarquant que $\phi_n(0) = \mathbb{P}[Z_n = 0]$, il s'ensuit que q est la limite de $\phi^{(n)}(0)$ lorsque $n \to \infty$.

Lorsque $m \leq 1$, on introduit la fonction $h(s) = \phi(s) - s$ qui vérifie, pour $0 \leq s < 1$ $h'(s) = \phi'(s) - 1 < \phi'(1) - 1 \leq 0$. Ainsi h est strictement décroissante sur $[0,1]$ avec $h(1) = 0$. On en déduit que $\phi(t) > t$ pour tout $t \in [0,1)$. Lorsque $m > 1$, on démontre de manière similaire que que $\phi(s) = s$ admet une unique solution sur $[0,1)$. Il est alors classique de montrer que la suite $\phi^{(n)}(0)$ converge vers le plus petit point fixe de ϕ sur $[0,1]$ lorsque $n \to \infty$.

2 – Percolation

On considère un graphe $\mathcal{G}$ formé d'un ensemble (fini ou dénombrable) de sites $\mathcal{S}$ et d'un ensemble d'arêtes $\mathcal{A}$ (une famille de couples de sites).

On définit une famille de variables aléatoires $(\omega(a), a \in \mathcal{A})$ indépendantes [1], de même loi de Bernoulli de paramètre $p \in [0,1]$. En d'autres termes, pour chaque arête a, on tire indépendamment pile (càd 1) avec probabilité p ou face (càd 0) avec probabilité $1-p$. Lorsque $\omega(a) = 1$, on dit que l'arête a est ouverte, et sinon fermée. On note P_p la loi correspondante.

Une réalisation de ω définit donc un sous-graphe aléatoire de $\mathcal{G}$ formé des sites $\mathcal{S}$ et des arêtes ouvertes pour ω. Souvent, on identifie la réalisation $\omega = (\omega(a), a \in \mathcal{A})$ avec le graphe qu'elle définit.

On considère le cas du graphe $\mathcal{G} = \mathbb{Z}^d$.

(i) Lorsque $d = 1$, p.s. existe-t-il une composante connexe infinie ?

Notre but est de démontrer le résultat suivant :

Théorème (Transition de phase pour la percolation)**.**

Lorsque $d \geq 2$, il existe $p_c = p_c(d) \in (0,1)$ tel que :

- pour tout $p < p_c$, ω n'a presque sûrement pas de composante connexe infinie.
- pour tout $p > p_c$, ω a presque sûrement (au moins) une composante connexe infinie.

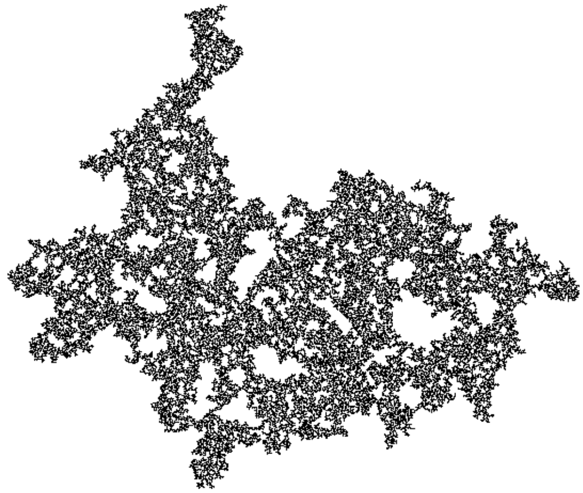


Figure 2 – Simulation de la composante connexe contenant l'origine dans $\mathbb{Z}^2$ (simulation de Daniel Tiggemann).

(ii) On note $A = \{\omega$ a au moins une composante connexe infinie$\}$. Montrer que $P_p(A) = 0$ ou 1.

(iii) Montrer que $p \mapsto P_p(A)$ est croissante.

Indication. On pourra définir une famille $(X(a), a \in \mathcal{A})$ de variables aléatoires indépendantes uniformes sur $[0,1]$ et considérer $\omega_p(a) = \mathbb{1}_{X(a) \leq p}$.

(iv) Montrer que pour p suffisamment petit, presque sûrement il n'existe pas de chemin ouvert de longueur infinie issue de l'origine.

(v) En déduire que $p_c > 0$.

(vi) Montrer que si p est suffisamment proche de 1 alors presque sûrement il existe un chemin ouvert de longueur infinie issue de l'origine. En déduire que $p_c < 1$.

Indication. Pour $d = 2$, on pourra remarquer que s'il n'existe pas de chemin ouvert de longueur infinie issue de l'origine, alors il existe un cycle "dual" fermé entourant l'origine.

(vii) Quelle est la probabilité critique p_c pour la percolation (par arêtes) dans un arbre binaire (chaque site, sauf la racine, est relié à trois voisins, et le graphe ne comporte pas de cycle) ? Pour $p = p_c$, existe-t-il une composante connexe infinie ?

Petits compléments (hors TD) :

- Montrer le théorème analogue dans le cas de la percolation par sites (et non plus sur arêtes) sur $\mathbb{Z}^d$.
- (★★) Montrer que sur $\mathbb{Z}^d$, ou bien il n'y a p.s. pas de composante connexe infinie, ou bien il y a p.s. une unique composante connexe infinie.

On peut démontrer que $p_c = 1/2$ pour la percolation par arêtes pour $\mathbb{Z}^2$ et que sous $P_{1/2}$ il n'y a p.s. pas de composante connexe infinie. On ne sait pas ce qu'il en est en dimension 3.

Pour davantage de détails, consulter le livre *Percolation et modèle d'Ising* de Wendelin Werner.

Ébauche de corrigé :

(i) Oui, uniquement lorsque $p = 1$. Si $p < 1$, le lemme de Borel-Cantelli implique que presque sûrement, $\omega(i) = 0$ pour une infinité de $i \geq 0$ et $i \leq 0$.

1. l'espace d'états est donc $\{0,1\}^{\mathcal{A}}$ muni de la tribu produit.

(ii) Vérifions tout d'abord que A eſt bien mesurable pour la tribu produit. À cet effet, notons $A_x = \{\omega$ a au moins une composante connexe infinie contenant $x\}$ pour $x \in \mathbb{Z}^d$, de sorte que

$$A = \bigcup_{x\in\mathbb{Z}^d} A_x = \bigcup_{x\in\mathbb{Z}_d} \bigcap_{n\geq 1} \{X \text{ eſt relié à la boule de centre } x \text{ et de rayon } n \text{ par un chemin ouvert}\}.$$

Or toute partie qui ne dépend que de l'état d'un nombre fini d'arêtes eſt mesurable pour la tribu produit. Formellement, cela veut dire que si on note $\pi_a : \{0,1\}^{\mathcal{A}} \to \{0,1\}$ la projeċtion définie par $\pi_a(\omega) = \omega(a)$, alors toute partie de la forme

$$\pi_{a_1}^{-1}(C_1) \cap \cdots \cap \pi_{a_j}^{-1}(C_j) \qquad (j \geq 1, \qquad a_1, \ldots, a_j \in \mathcal{A}, \qquad C_1, \ldots, C_j \subset \{0,1\})$$

eſt mesurable pour la tribu produit. Ainsi A eſt bien mesurable.

Le fait que $\mathbb{P}_p(A) = 0$ ou 1 eſt une conséquence de la loi du 0-1 de Kolmogorov. Si on veut rédiger cela en détail, on peut procéder comme suit. Soit $(a_i)_{i\geq 1}$ une énumération des arêtes de $\mathcal{A}$ et notons $X_i = \omega(a_i)$ pour $i \geq 1$. Ainsi, sous $\mathbb{P}_p$, $(X_i)_{i\geq 1}$ eſt une suite de variables aléatoires i.i.d. Or A ne dépend pas de l'état d'un nombre fini d'arêtes, de sorte que pour tout $n \geq 1$, $A \in \sigma(X_n, X_{n+1}, \ldots)$. Donc A eſt dans la tribu asymptotique de $(X_i)_{i\geq 1}$, qui eſt triviale d'après la loi du 0-1 de Kolmogorov pour une suite de variables indépendantes.

(iii) La loi de ω_p eſt $\mathbb{P}_p$. On en déduit que pour $p \leq p'$:

$$\mathbb{P}_p(A) = \mathbb{P}(\omega_p \in A) \leq \mathbb{P}(\omega_{p'} \in A) = P_{p'}(A).$$

Pour l'inégalité, on a utilisé l'inclusion $\{\omega_p \in A\} \subset \{\omega_{p'} \in A\}$, vraie presque sûrement.

(iv) On rappelle la notation $A_x = \{\omega$ a au moins une composante connexe infinie contenant $x\}$ pour $x \in \mathbb{Z}^d$. Mais

$$A_0 \subset \{ \text{ il exiſte un chemin injeċtif ouvert infini issu de } 0\}.$$

Mais le nombre de chemins injeċtifs ouverts de longueur n issus de 0 eſt au plus $2d(2d-1)^{n-1}$. Donc, pour tout $n \geq 1$,

$$\mathbb{P}[A_0] \leq \mathbb{P}(\text{ il exiſte un chemin injeċtif ouvert de longueur } n \text{ issu de } 0) \leq p^n 2d(2d-1)^{n-1},$$

qui tend vers 0 lorsque $n \to \infty$ pour $p < 1/(2d)$. Ainsi $\mathbb{P}[A_0] = 0$ lorsque $p < 1/(2d)$.

(v) On a

$$\mathbb{P}[A] \leq \sum_{x\in\mathbb{Z}^d} \mathbb{P}[A_x] \leq \sum_{x\in\mathbb{Z}^d} \mathbb{P}[A_0]$$

par invariance par translation. On a vu que $\mathbb{P}[A_0] = 0$ pour $p < 1/(2d)$, donc $p_c \geq 1/(2d)$.

(vi) Comme $\mathbb{Z}^2 \subset \mathbb{Z}^d$, s'il exiſte p.s. une composante connexe infinie dans $\mathbb{Z}^2$, alors il en exiſte une p.s. dans $\mathbb{Z}^d$, ce qui implique $p_c(\mathbb{Z}^d) \leq p_c(\mathbb{Z}^2)$. Il suffit donc de prouver le résultat pour $d = 2$. Par dualité, s'il n'exiſte pas de chemin ouvert de longueur infinie issue de l'origine, alors il exiſte un cycle "dual" fermé entourant l'origine. Mais ce cycle dual interseċte nécessairement la demi-droite positive. Or si un tel cycle passe par $[x, x+1]$, alors il eſt de longueur au moins $2x+1$ et contient donc un chemin injeċtif de longueur $2x$ passant par $[x, x+1]$. Comme il y a u plus 4^{2x} chemins injeċtifs de longueur $2x$ passant par $[x, x+1]$, on a :

$$\mathbb{P}_p(0 \text{ n'eſt pas dans une composante connexe infinie}) \leq \sum_{x\geq 1} (4(1-p))^{2x}.$$

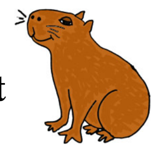
Lorsque p eſt suffisamment proche de 1, cette somme eſt ſtrcitement inférieure à 1, ce qui implique $\mathbb{P}_p(A_0) > 0$. D'où $\mathbb{P}_p(A) > 0$, et comme $\mathbb{P}_p(A) = 0$ ou 1, il s'ensuit que $\mathbb{P}_p(A) = 1$ pour p suffisamment proche de1.

(vii) La composante connexe de la racine est un arbre de Galton-Watson de loi de reproduction μ donnée par :

$$\mu_0 = (1-p)^2, \qquad \mu_1 = 2p(1-p), \qquad \mu_2 = p^2, \qquad \mu_i = 0 \quad (i \geq 3).$$

D'après l'exercice sur les processus de Galton-Watson, la composante connexe de la racine est presque sûrement finie si et seulement si la moyenne de μ est inférieure ou égale à 1. On en déduit immédiatement que $p_c = 1/2$ et que pour $p = 1/2$ il n'y a presque sûrement pas de composante connexe infinie.

3 – Le capybara savant



Un capybara tape un texte aléatoirement sur son capyclavier contenant les 42 lettres de l'alphabet capybarien (en choisissant chaque nouvelle lettre uniformément au hasard, les lettres étant tapées aux instants $1, 2, 3, \ldots$). On note X_n la lettre tapée à l'instant n. Pour un mot[2] $\mathbb{M} = m_1 m_2 \cdots m_k$, on note $T_{\mathbb{M}}$ le premier instant où le capybara a tapé exactement le mot $\mathbb{M}$, c'est-à-dire :

$$T_{\mathbb{M}} = \inf\{n \geq 0; \quad (X_{n-k+1}, X_{n-k}, \ldots, X_n) = (m_1, m_2, \ldots, m_k)\}.$$

On suppose que l'alphabet romain est inclus dans l'alphabet capybarien. Quelle est la probabilité de lire ABRACADABRA avant ABRACADABRI ? Qui est le plus grand entre $\mathbb{E}[T_{ABRACADABRA}]$ et $\mathbb{E}[T_{ABRACADABRI}]$?

Élements de réponse :

En s'arrêtant aux instants successifs où on lit ABRACADABR, on voit que la probabilité de lire ABRACADABRA avant ABRACADABRI vaut $1/2$.

Donnons une explication intuitive du fait que $\mathbb{E}[T_{ABRACADABRA}] > \mathbb{E}[T_{ABRACADABRI}]$:
- lorsqu'on lit ABRACADABR et que la lettre d'après n'est pas A, tout se passe comme si on recommençait de zéro si on veut lire ABRACADABRA.
- lorsqu'on lit ABRACADABR et que la lettre d'après n'est pas I, tout ne se pas **pas** comme si on recommençait de zéro si on veut lire ABRACADABRI. En effet, si la lettre d'après est A, cela nous favorise car on pourrait lire ABRACADABRA**BRACADABRA** ou ABRACADABRA**CADABRA** (tout se passe comme si on recommençait en sachant qu'on a déjà un A, ce qui nous favorise).

Il est possible calculer explicitement $\mathbb{E}[T_{\mathbb{M}}]$ pour tout mot $\mathbb{M}$ (cette quantité fait intervenir le nombre de préfixes de $\mathbb{M}$ qui sont également suffixes). Par exemple, on a

$$\mathbb{E}[T_{ABRACADABRA}] = 42^{11} + 42^4 + 42^1, \qquad \mathbb{E}[T_{ABRACADABRI}] = 42^{11}.$$

Une démonstration simple de ceci sera donnée en TD au second semestre en utilisant le théorème d'arrêt pour les martingales.

2. Un mot est toujours supposé de longueur finie.

Fin