

TER*- La composante géante d'un graphe aléatoire

Christoph Temmel †

Résumé

Dans ce TER je présente certains aspects du comportement de la plus grande composante connexe du graphe $G(1, p)$, spécialement la phase de transition autour du cas critique avec $p = \frac{1}{n}$. On va voir qu'il y a une grande probabilité pour que dans le cas sous-critique avec $p < \frac{1}{n}$, la taille de la plus grande composante soit d'ordre $O(\log n)$, dans le cas critique d'ordre $O(n^{2/3})$ et d'ordre $O(n)$ dans le cas sur-critique, où $p > \frac{1}{n}$.

*Travail d'étude de la recherche

†encadré par Prof. Pierre Mathieu, UdP

Table des matières

1	Introduction	3
2	Autour du graphe aléatoire	3
2.1	Le modèle du graphe aléatoire	3
2.2	Le processus d'exploration	4
2.3	Le couplage	9
3	La taille de la plus grande composante connexe	10
3.1	Commentaires	10
3.2	Majoration sous-critique	11
3.3	Majoration critique	12
3.4	Minoration critique	14
3.5	Minoration sur-critique	18
4	Notations et théorèmes de base	19
4.1	Énoncés de la probabilité	19
4.2	Énoncés de la distribution binomiale	21
4.3	Le dépassement d'un marche aléatoire	24

Table des figures

1	Le K_6 et 2 réalisations de $G(6, \frac{1}{6})$	4
2	Une réalisation de $G(9, \frac{1}{9})$	5

Liste des tableaux

1	L'algorithme et le processus d'exploration sur la réalisation de $G(9, \frac{1}{9})$ en détail	5
2	Relation entre Y et N en temps t et t-1	7

Remerciements

Je voudrais remercier ici mon professeur Pierre Mathieu pour son encadrement, son support et ses explications détaillées. Je tiens aussi à le remercier pour sa patience et pour m'avoir poussé à travailler soigneusement.

1 Introduction

Dans ce TER je détaille le calcul pour déterminer (avec une grande probabilité) la taille de la plus grande composante connexe du graphe aléatoire $G(n, p)$. La définition de ce graphe et un processus d'exploration, qui va aider à montrer les résultats souhaités, et son couplage avec une martingale seront présentés dans la première partie 2.

Dans la partie principale 3 on trouve le calcul détaillé pour les valeurs différentes de p . On va remarquer qu'il y a un changement drastique autour du cas critique $p = \frac{1}{n}$: au-dessous le graphe consiste en de nombreuses petites composantes (d'ordre $\log n$), dans le cas critique on va avoir des composantes d'une taille $n^{2/3}$ et au-dessus le graphe est constitué presque exclusivement d'une composante géante d'ordre n . Ce changement rapide peut être comparé avec la percolation, c'est-à-dire le comportement de l'eau autour de zéro degré et sa transformation de l'état liquide à l'état solide.

La discussion est effectuée principalement avec l'aide des processus stochastiques discrètes, spécialement du processus d'exploration et des martingales diverses. Le concept des démonstrations pour le cas critique (voir les sections 3.4 et 3.3) est pris d'un article de Asaf Nachmias et Yuval Peres [7], pendant que les deux autres cas (voir section 3.2 pour $p < \frac{1}{n}$ et section 3.5 pour $p > \frac{1}{n}$) suivent le livre de Alon, Spencer et Paul Erdős [1], chapitre 10.

La dernière partie 4 rassemble tous les énoncés de la probabilité et théorie des martingales qui sont utilisés (voir la section 4.1), pris de [4], certains résultats sur les grandes déviations d'une loi binomiale (de l'appendix de [1]) et la loi binomiale itérée (section 4.2) et la discussion en détail du dépassement d'une marche aléatoire (de [7]) dans la section 4.3.

2 Autour du graphe aléatoire

2.1 Le modèle du graphe aléatoire

Pour les définitions reliées à la théorie des graphes voir [2]. On considère le modèle d'un graphe aléatoire suivant, proposé pour la première fois pendant les années 1960 par Erdős [3] : pour un entier naturel n et une probabilité $0 < p < 1$ (pour $p = 1$ on aura le graphe complet K_n et pour $p = 0$ n points, donc on exclut ces 2 cas dégénérés) on définit le graphe aléatoire $G(n, p)$ en prenant le graphe complet de degré n et en supprimant l'une après l'autre indépendamment chaque arête avec la probabilité $1-p$ (ou : retenir indépendamment chaque arête

avec probabilité p). Pour un exemple voir la figure 1.

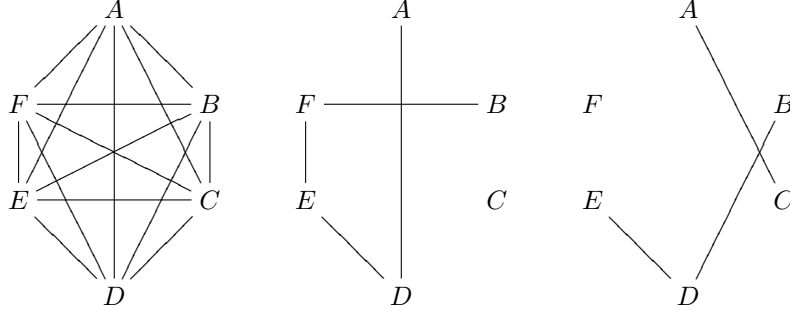


FIG. 1 – Le K_6 et 2 réalisations de $G(6, \frac{1}{6})$

On note par C_1, \dots, C_k les k composantes connexes de $G(n, p)$, ordonnées par taille décroissante (c'est-à-dire que C_1 est la plus grande composante) et on remarque ici que k est une variable aléatoire discrète à valeurs dans $\{0, \dots, n\}$. Suivant deux définitions prises de [1], chapitre 6 :

Définition 1 (Propriété des graphes) *On appelle une propriété des graphes P un sous-ensemble de l'ensemble de tous les graphes qui est stable sous isomorphisme.*

Définition 2 (Propriété monotone) *On appelle une propriété des graphes P monotone croissante (décroissante) si pour tout graphe G et tout graphe H obtenu à partir de G en ajoutant (supprimant) des arêtes on a*

$$G \in P \Rightarrow H \in P.$$

Une exemple pour une propriété des graphes monotone croissante est la taille de la plus grande composante connexe et une exemple pour une propriété des graphes monotone décroissante est le nombre des composantes connexes d'un graphe. Pour être complet il faut remarquer qu'il existe un deuxième modèle d'un graphe aléatoire, aussi propose par Erdős [3] : $G(n, m)$, qui est définit en prenant un graphe avec n sommets et exactement m arêtes. Mais pour n assez grand ce modèle se comporte presque comme $G(n, \frac{2m}{n(n+1)})$ pour les propriétés monotones, donc on préfère le premier modèle qui est plus facile à traiter.

2.2 Le processus d'exploration

On regarde le graphe $G(n, p)$. Pour un sommet v on note $C(v)$ la composante connexe qui contient v . On commence un processus d'exploration qui a été développé dans les années 80 par Karp [5] et Martin-Löf [6]. Pendant l'exécution de ce processus les sommets du graphe sont partitionnés dans 3 ensembles selon leur état : $w \in \mathcal{A}_t$ si w actif en temps t , $w \in \mathcal{E}_t$ si w exploré en temps t et $w \in \mathcal{N}_t$ si w neutre en temps t . On fixe un ordre sur les sommets et on note par $\min(X)$ l'élément minimal pour l'ordre dans l'ensemble X . On note par w_t l'élément pivot en temps t .

Algorithme 3 (Algorithme d'exploration) Initialisation Au temps $t = 0$ on a $\mathcal{A}_0 = \{v\}$, $\mathcal{E}_0 = \emptyset$ et $\mathcal{N}_0 = \{v_1, \dots, v_n\} \setminus \{v\}$.

Boucle $\forall t \in \{1, \dots, n\}$:

1. si $\mathcal{A}_{t-1} \neq \emptyset$ alors $w_{t-1} = \min(\mathcal{A}_{t-1})$, sinon $w_{t-1} = \min(\mathcal{N}_{t-1})$;
2. $\mathcal{V}(w_{t-1}) = \{v \in \mathcal{N}_{t-1} \text{ tq } w_{t-1} \text{ relié à } v\}$;
3. $\mathcal{A}_t = (\mathcal{A}_{t-1} \cup \mathcal{V}(w_{t-1})) \setminus \{w_{t-1}\}$;
4. $\mathcal{N}_t = \mathcal{N}_{t-1} \setminus (\mathcal{V}(w_{t-1}) \cup \{w_{t-1}\})$;
5. $\mathcal{E}_t = \mathcal{E}_{t-1} \cup \{w_{t-1}\}$.

Invariant $\forall t \in \{0, \dots, n\}$ on a $\{v_1, \dots, v_n\} = \mathcal{E}_t \cup \mathcal{A}_t \cup \mathcal{N}_t$.

Un exemple d'algorithme dans le cas de la réalisation présentée dans la figure 2 est donné dans le tableau 1.

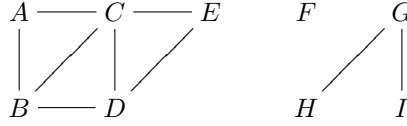


FIG. 2 – Une réalisation de $G(9, \frac{1}{9})$

t	Y_t	\mathcal{A}_t	\mathcal{E}_t	\mathcal{N}_t	w_t
0	1	$\{A\}$	\emptyset	$\{B, \dots, I\}$	A
1	2	$\{B, C\}$	$\{A\}$	$\{D, \dots, I\}$	B
2	1	$\{C, D\}$	$\{A, B\}$	$\{E, \dots, I\}$	C
3	2	$\{D, E\}$	$\{A, B, C\}$	$\{F, \dots, I\}$	D
4	1	$\{E\}$	$\{A, B, C, D\}$	$\{F, G, H, I\}$	E
5	0	\emptyset	$\{A, \dots, E\}$	$\{F, G, H, I\}$	F
6	0	\emptyset	$\{A, \dots, F\}$	$\{G, H, I\}$	G
7	2	$\{H, I\}$	$\{A, \dots, G\}$	$\{H, I\}$	H
8	1	$\{I\}$	$\{A, \dots, H\}$	$\{I\}$	I
9	0	\emptyset	$\{A, \dots, I\}$	\emptyset	-

TAB. 1 – L'algorithme et le processus d'exploration sur la réalisation de $G(9, \frac{1}{9})$ en détail

On note que cet algorithme est utilisable pour n'importe quel graphe, contrairement au processus suivant, qui est déjà lié au modèle particulier qui on utilise ici.

Définition 4 (Processus d'exploration) On définit maintenant $Y_t = |\mathcal{A}_t|$ (le processus d'exploration), $N_t = |\mathcal{N}_t \setminus \{w_t\}|$ et $\eta_t = |\mathcal{V}(w_t)|$ (défini en pas 2 d'algorithme 3).

Proposition 5 (Invariants) *Les invariants sont vrais pour tout $t \in \{1, \dots, n\}$:*

$$|\mathcal{E}_t| = t \quad (1a)$$

$$n = t + Y_t + N_t + \mathbf{1}_{\{Y_t=0\}} \quad (1b)$$

$$\eta_t \mathbf{1}_{\{N_{t-1}=k\}} \sim \text{Bin}(k, p) \mathbf{1}_{\{N_{t-1}=k\}} \quad (1c)$$

$$Y_t = Y_{t-1} + \eta_t - \mathbf{1}_{\{Y_{t-1} \neq 0\}} \quad (1d)$$

PREUVE. Pour montrer (1a), il suffit de voir que dans chaque itération d'algorithme on ajoute w_{t-1} à \mathcal{E}_t . L'invariant (1b) est bien compris en classifiant les sommets en temps t

$$\{v_1, \dots, v_n\} = \mathcal{E}_t \cup \mathcal{A}_t \cup \mathcal{N}_t$$

ce qui se traduit en cardinaux en comptant "tout = explorés + actifs + neutres + on commence à explorer une nouvelle composante" (voir pas 1 et 4 d'algorithme 3) :

$$n = t + Y_t + N_t + \mathbf{1}_{\{Y_t=0\}}.$$

La loi conditionnelle de η_t sachant N_{t-1} aussi ne pose aucun problème : le sommet pivot w_t a une probabilité p d'être relié indépendamment avec chacun des $N_{t-1} = k$ sommets neutres. Ainsi (1c). Finalement, la définition récurrente de Y_t est immédiate à voir si on pense à ajouter 1 au début de chaque nouvelle composante connexe (voir pas 3 d'algorithme 3). \square

Ainsi on définit le temps d'arrêt $\tau = \min\{t \geq 1 : Y_t = 0\}$ et on observe qu'au temps τ l'ensemble des sommets déjà explorés est égal à $C(v)$ (dans le dernier pas on a marqué $w_{\tau-1}$ exploré et on n'a ajouté aucun sommet parce que $\eta_{\tau-1} = 0$), donc on a $|C(v)| = \tau$.

Proposition 6 (Lien entre Y_t et N_t) *Pour tout $t \in \{1, \dots, n\}$ on a :*

$$0 \leq N_t \leq n - t - 1 \quad (2a)$$

$$0 \leq Y_t \leq n - t \quad (2b)$$

$$\left. \begin{aligned} \mathbf{1}_{\{Y_t=k\}} &= \mathbf{1}_{\{N_t=n-t-k\}} & \forall k \in \{2, \dots, n-t\} \\ \mathbf{1}_{\{Y_t \in \{0,1\}\}} &= \mathbf{1}_{\{N_t=n-t-1\}} \end{aligned} \right\} \quad (2c)$$

PREUVE. Les assertions (2a) et (2b) sont une conséquence directe de (1b). De même façon (2c) - ici on voit que Y_t porte un peu plus d'information que N_t . Pour une visualisation de la relation entre N_t et Y_t voir le tableau 2. \square

C'est n'est pas possible de déterminer un forme simple et explicite en t pour la loi de Y_t , comme on voit dans la récurrence suivante :

$$\begin{aligned}
& N_t \mathbf{1}_{\{N_{t-1}=l\}} \\
& \sim [n - Y_t - \mathbf{1}_{\{Y_t=0\}} - t] \mathbf{1}_{\{N_{t-1}=l\}} \\
& \sim [n - [Y_{t-1} + \eta_t - \mathbf{1}_{\{Y_{t-1} \neq 0\}}] - \mathbf{1}_{\{Y_t=0\}} - t] \mathbf{1}_{\{N_{t-1}=l\}} \\
& \sim [n - Y_{t-1} - \eta_t + (1 - \mathbf{1}_{\{Y_{t-1}=0\}}) - \mathbf{1}_{\{Y_t=0\}} - (t-1) - 1] \mathbf{1}_{\{N_{t-1}=l\}} \\
& \sim [[n - Y_{t-1} - \mathbf{1}_{\{Y_{t-1}=0\}} - (t-1)] - \eta_t + 1 - \mathbf{1}_{\{Y_t=0\}} - 1] \mathbf{1}_{\{N_{t-1}=l\}} \\
& \sim [N_{t-1} - \eta_t - \mathbf{1}_{\{Y_t=0\}}] \mathbf{1}_{\{N_{t-1}=l\}} \\
& \sim [N_{t-1} - \text{Bin}(N_{t-1}, p) - \mathbf{1}_{\{Y_t=0\}}] \mathbf{1}_{\{N_{t-1}=l\}} \\
& \sim [\text{Bin}(N_{t-1}, 1-p) - \mathbf{1}_{\{Y_t=0\}}] \mathbf{1}_{\{N_{t-1}=l\}}
\end{aligned}$$

ce qui interdit l'application de la proposition d'une loi binomiale itérée 27. Donc on commence par définir une deuxième variante de processus d'exploration (voir [1]), qui echange une loi explicite contre un lien explicite avec l'algorithme 3 :

Définition 7 (Le processus d'exploration, 2^{ème} variante) *On définit un processus d'exploration par récurrence par $X_0 = 1$ et*

$$M_t = n - t - X_t$$

$$\kappa_t \mathbf{1}_{\{M_{t-1}=k\}} \sim \text{Bin}(M_{t-1}, p) \mathbf{1}_{\{M_{t-1}=k\}}$$

$$X_t = X_{t-1} + \kappa_t - 1.$$

Remarques :

1. On a encore un invariant :

$$n = X_t + M_t + t \tag{3}$$

2. Le lien avec le processus d'exploration défini en 4 est le suivant : Si on prend les temps d'arrêt pour $i \in \{1, \dots, n\}$

$$T_0 = 0 \text{ et } T_i = \min\{t > 0 : X_t = 1 - i \text{ la } 1^{\text{ère}} \text{ fois}\} \wedge (n + 1)$$

		n-t	n-t	n-t-1	n-t-2	...	1	0	N_{t-1}
		0	1	2	3	...	n-t	n-t+1	Y_{t-1}
n-t-1	0	x	x						
n-t-1	1	x	x	x					
n-t-2	2	x	x	x	x				
n-t-3	3	x	x	x	x				
⋮	⋮	⋮	⋮	⋮	⋮	⋮			
1	n-t-1	x	x	x	x	...	x		
0	n-t	x	x	x	x	...	x	x	
N_t	Y_t								

TAB. 2 – Relation entre Y et N en temps t et t-1

et

$$\tau_0 = 0 \text{ et } \tau_i = \min\{t > 0 : Y_t = 0 \text{ la } i^{\text{ème}} \text{ fois}\} \wedge (n + 1)$$

on a

$$X_{(t \vee (T_{i-1}+1)) \wedge T_i} = Y_{(t \vee (\tau_{i-1}+1)) \wedge \tau_i} - (i - 1) \quad \forall t \in \{0, \dots, n\}$$

où Y_t le processus d'exploration défini en 4. En effet, la différence entre les 2 processus est seulement l'ajout de 1 au début d'exploration d'une nouvelle composante, comme montre par :

$$X_t = Y_t - \sum_{s=0}^{t-1} \mathbf{1}_{\{Y_s=0\}} \quad \forall t \in \{0, \dots, n\} \quad (4)$$

3. 4 donne immédiatement $X_t \stackrel{\text{st}}{\leq} Y_t$.
4. La distribution de κ_t est encore décrite seulement par une distribution conditionnelle.

Il se trouve, qu'en raison de sa définition plus simple, on peut calculer la distribution de ce processus explicitement, comme le montre la proposition suivante :

Proposition 8 (Distribution du 2^{ème} processus d'exploration) *On a*
 $\forall t \in \{0, \dots, n\}$:

$$X_t \sim \text{Bin}(n - 1, 1 - (1 - p)^t) + 1 - t$$

$$M_t \sim \text{Bin}(n - 1, (1 - p)^t) \text{ et } \kappa_t \sim \text{Bin}(n - 1, p(1 - p)^t).$$

PREUVE. La façon la plus simple est de le montrer par récurrence sur M_t : pour $M_0 = n - 1 \sim \text{Bin}(n - 1, (1 - p)^0)$ et pour $t - 1 \rightarrow t$

$$\begin{aligned} & M_t \mathbf{1}_{\{M_{t-1}=k\}} \\ & \sim (n - X_t - t) \mathbf{1}_{\{M_{t-1}=k\}} \\ & \sim [n - (X_{t-1} + \kappa_t - 1) - t] \mathbf{1}_{\{M_{t-1}=k\}} \\ & \sim [n - X_{t-1} - \text{Bin}(M_{t-1}, p) + 1 - t] \mathbf{1}_{\{M_{t-1}=k\}} \\ & \sim [n - X_{t-1} - (t - 1) - \text{Bin}(M_{t-1}, p)] \mathbf{1}_{\{M_{t-1}=k\}} \\ & \sim [M_{t-1} - \text{Bin}(M_{t-1}, p)] \mathbf{1}_{\{M_{t-1}=k\}} \\ & \sim \text{Bin}(M_{t-1}, 1 - p) \mathbf{1}_{\{M_{t-1}=k\}} \end{aligned}$$

et on applique la proposition de la loi binomiale itérée 27 à ce résultat pour obtenir

$$M_t \sim \text{Bin}(n - 1, (1 - p)^t).$$

Finalement on applique ce résultat à l'invariant (3) pour arriver à

$$X_t \sim n - M_t - t \sim n - \text{Bin}(n - 1, (1 - p)^t) - t \sim \text{Bin}(n - 1, 1 - (1 - p)^t) + 1 - t$$

et la proposition de la loi binomiale itérée 27 à κ_t , or

$$\kappa_t \sim \text{Bin}(n - 1, p(1 - p)^t).$$

□

Proposition 9 (Nombre des composantes connexes de $G(n, p)$) Soit k le nombre des composantes connexes dans $G(n, p)$, alors

$$k \sim 1 + \text{Bin}(n-1, (1-p)^n)$$

est k d'ordre $O(n)$ pour $p = \frac{c}{n}$ avec $c \neq 0$ et n assez grand.

PREUVE. Comme dans le début de l'exploration de chaque nouvelle composante après la première X_t décroît par 1 comparé avec Y_t on a

$$\begin{aligned} k & \\ & \sim -(X_n - 1) \\ & \sim 1 - \text{Bin}(n-1, 1 - (1-p)^n) - 1 + n \\ & \sim n - \text{Bin}(n-1, 1 - (1-p)^n) \\ & \sim 1 + \text{Bin}(n-1, (1-p)^n) \end{aligned}$$

De plus on voit que pour n assez grand et $p = \frac{c}{n}$ on a $(1-p)^n \approx e^{-c}$ donc $\mathbf{E}k = 1 + e^{-c}n \in O(n)$. \square

2.3 Le couplage du processus d'exploration avec une marche aléatoire

Maintenant je détaille comment coupler une marche aléatoire

$$S_t = 1 + \sum_{i=1}^t (\xi_i - 1)$$

avec le processus d'exploration Y_t décrit dans section 2.2 de telle sorte qu'à chaque instant $t \leq \tau$ on a $Y_t \stackrel{\text{st}}{\leq} S_t$. On prend n^2 variables de loi Bernoulli(p), qu'on divise en 2 parties, \mathcal{I} et \mathcal{J} , de cardinaux $|\mathcal{I}| = \frac{n(n-1)}{2}$ et $|\mathcal{J}| = \frac{n(n+1)}{2}$ respectivement. La première partie va former le graphe $G(n, p)$, la deuxième partie va permettre la construction des ξ_t . On va noter $\bar{Y}_t = Y_t + \mathbf{1}_{\{Y_t=0\}}$, ce qui donne la relation $n = t + N_t + \bar{Y}_t$ (voir (1b)).

Pour chaque $t \in \{1, \dots, n-1\}$ on fait la chose suivante : on choisit les $n-1$ variables de \mathcal{I} qui déterminent les arêtes de w_{t-1} , dont on prend les N_{t-1} qui représentent les arêtes entre w_{t-1} et les sommets neutres pour η_t . On ajoute les \bar{Y}_{t-1} variables pour les arêtes entre w_{t-1} et les sommets actifs (de \mathcal{I}) et $n - N_{t-1} - \bar{Y}_{t-1} = t$ variables additionnelles (de \mathcal{J}), ce qui donne ξ_t . On a donc la loi conditionnelle de η_t (voir 1c)

$$\eta_t \mathbf{1}_{\{N_{t-1}=k\}} \sim \text{Bin}(k, p) \text{ et } \xi_t \sim \text{Bin}(N_{t-1} + \bar{Y}_{t-1} + t, p) \sim \text{Bin}(n, p)$$

est la loi de ξ_t . Maintenant on montre que $\eta_t \stackrel{\text{st}}{\leq} \xi_t$ pour tout t . On a pour $c \in \mathbb{R}$ quelconque

$$\begin{aligned} & \mathbf{P}(\eta_t > c | N_{t-1} = m_{t-1}, \dots, N_1 = m_1) \\ &= \frac{\mathbf{P}(\text{Bin}(m_{t-1}, p) \sim \eta_t > c, N_{t-1} = m_{t-1}, \dots, N_1 = m_1)}{\mathbf{P}(N_{t-1} = m_{t-1}, \dots, N_1 = m_1)} \\ &\leq \frac{\mathbf{P}(\text{Bin}(n - m_{t-1}, p) + \eta_t \sim \text{Bin}(n, p) \sim \xi_t > c, N_{t-1} = m_{t-1}, \dots, N_1 = m_1)}{\mathbf{P}(N_{t-1} = m_{t-1}, \dots, N_1 = m_1)} \end{aligned}$$

par la construction détaillée au début. $\text{Bin}(n - m_{t-1}, p)$ représente les $n - m_{t-1}$ variables qui sont ajoutées à η_t pour obtenir ξ_t (voir ci-dessus).

$$= \mathbf{P}(\xi_t > c | N_{t-1} = m_{t-1}, \dots, N_1 = m_1)$$

ce qui donne

$$\begin{aligned} \mathbf{P}(\eta_t > c) &= \mathbf{E}[\mathbf{P}(\eta_t > c | \sigma(N_{t-1}, \dots, N_1))] \\ &\leq \mathbf{E}[\mathbf{P}(\xi_t > c | \sigma(N_{t-1}, \dots, N_1))] = \mathbf{P}(\xi_t > c) \end{aligned}$$

Il reste à montrer que les ξ_t produits par cette démarche sont indépendants. Ici les $I_i \in \mathcal{I}$, et les $J_j \in \mathcal{J}$. On a $\forall t \in \{1, \dots, n\}, \forall (c_1, \dots, c_t) \in \mathbb{N}^t$:

$$\begin{aligned} &\mathbf{P}(\xi_t = c_t, \xi_{t-1} = c_{t-1}, \dots, \xi_1 = c_1) \\ &= \mathbf{P}\left(\sum_{i_t=1}^t I_{i_t} + \sum_{j_t=1}^{n-t} J_{j_t} = c_t, \dots, \sum_{i_1=1}^t I_{i_1} + \sum_{j_1=1}^{n-t} J_{j_1} = c_1\right) \end{aligned}$$

Ici nous observons que

1. Les I_i sont indépendants des J_j par définition ;
2. Pour $s < t$ les J_{j_t} sont indépendants des J_{j_s} parce qu'on peut fixer un ordre dans \mathcal{J} pour le choix des J_j ;
3. Pour $s < t$ les I_{i_t} sont indépendants des I_{i_s} parce que si on connaît la valeur d'un I_{i_t} à temps s , ça implique que c'était une arête d'un sommet exploré avant temps s , qui est en contradiction avec le choix des $N_{t-1} + \bar{Y}_{t-1}$ éléments de \mathcal{I} dans l'algorithme.

$$\begin{aligned} &= \prod_{s=1}^t \mathbf{P}\left(\sum_{i_s=1}^s I_{i_s} + \sum_{j_s=1}^{n-s} J_{j_s} = c_s\right) \\ &= \prod_{s=1}^t \mathbf{P}(\xi_s = c_s) \end{aligned}$$

ce qui donne l'indépendance souhaitée. De plus, $\{S_t\}_{t=1}^n$ est une martingale pour $p = \frac{1}{n}$:

$$\mathbf{E}[S_t | \mathcal{F}_{t-1}] = \mathbf{E}[S_{t-1} + (\xi_t - 1) | \mathcal{F}_{t-1}] = S_{t-1} + \mathbf{E}[\xi_t | \mathcal{F}_{t-1}] - 1 = S_{t-1},$$

sinon il est une sous- respectivement une surmartingale.

3 La taille de la plus grande composante connexe

3.1 Commentaires

Les sous-sections suivants contiennent les théorèmes qui constituent le coeur de ce travail. Je voudrais mentionner ici tous les choses qui ne sont pas importantes pour les preuves elles-mêmes, mais néanmoins intéressantes dans le

contexte general.

Comme n est fini, tous les temps d'arrêts sont finis. Il en est de même pour toutes les martingales, qui sont basées sur des lois finies. Alors, on peut appliquer le théorème 25 (théorème d'arrêt) sans problème dans tous les preuves.

On note la symétrie étonnante entre les preuves de la majoration sous-critique et de la minoration sur-critique : ils utilisent les deux la côté exponentielle de la proposition de grandes déviations 26 car les lois concernées sont exactement dans les senses opposées et il y a la même expression $1 - x - \log x$ dans un point crucial dans les deux preuves, chaque fois avec une signe différenté.

Le couplage de la section 2.3 est utilisé pour les deux majorations et la 2^{ème} variante de processus d'exploration 7 seulement dans la minoration sur-critique. Il est aussi possible de la utiliser dans la majoration sous-critique, mais le resultat reste exactement de la même ordre (si on sait seulement $c < 1$), comme un calcul vite le montre. Donc, en utilisant le lien entre les deux processus d'exploration détaillé dans (4) dans section 2.2 et le couplage dans la section 2.3 on a $\forall t \in \{0, \dots, n\}$:

$$X_t \stackrel{\text{st}}{\leq} Y_t \stackrel{\text{st}}{\leq} S_t.$$

La minoration sur-critique est normalement effectuée en passant avec $n \rightarrow \infty$, où X_t devient un processus de branchement (avec les $\kappa_t \sim \text{Poi}(c)$), voir [1] chapitre 10). Dans le processus de branchement il existe une dualité qui on peut utiliser pour déterminer que l'ordre de tous les autres composants petits qui restent est $O(\log n)$. Ca ne sera pas possible avec la preuve présentée ici, car la dualité n'a pas un équivalent discret. D'autre côté la preuve présentée ici est plus courte et directe.

3.2 Majoration sous-critique

Théorème 10 (Majoration sous-critique) *Dans le graphe $G(n, \frac{c}{n})$, où $c < 1$, on a pour chaque $A > 0$*

$$\mathbf{P}(|C_1| \geq A \log n) \leq n^{1-c''A} \tag{5}$$

où c'' est une constante dépendant de c .

PREUVE. On commence par remarquer que pour $0 < h < n$ on a :

$$S_h = 1 + \sum_{t=1}^h (\xi_t - 1) = \sum_{t=1}^h \xi_t - (h - 1) \sim \text{Bin}(hn, \frac{c}{n}) - (h - 1)$$

on a pour chaque sommet v et $0 < h < n$:

$$\begin{aligned}
& \mathbf{P}(|C(v)| > h) \\
&= \mathbf{P}(\tau > h) \\
&= \mathbf{P}(\min\{Y_1, \dots, Y_h\} > 0) \\
&\leq \mathbf{P}(\min\{S_1, \dots, S_h\} > 0) && \text{utilisant le couplage de 2.3} \\
&\leq \mathbf{P}(S_h > 0) \\
&= \mathbf{P}\left(\text{Bin}\left(hn, \frac{c}{n}\right) > h - 1\right) \\
&= \mathbf{P}\left(\text{Bin}\left(hn, \frac{c}{n}\right) > \frac{1}{c'}hc\right) && \text{où } c' \in]0, c[\text{ pour } h \text{ assez grand}
\end{aligned}$$

Remarque : $\frac{hc}{h-1} < 1$ pour $\frac{1}{1-c} < h$

$$\begin{aligned}
&\leq [e^{1/c' - 1 - (1/c') \log(1/c')}]^{hc} && \text{par (18b)} \\
&= e^{-f(c')hc/c'}
\end{aligned}$$

en voyant que $\forall x \in]0, 1[$ on a $f(x) = 1 - x + \log x < 0$

$$\leq e^{-c''h} \quad \text{où } c'' = -\frac{f(c')c}{c'} > 0.$$

Maintenant on prend $h = A \log n$ avec $A > 0$, ce qui donne

$$\mathbf{P}(|C(v)| > h) \leq n^{-c''A}$$

or

$$\mathbf{P}(|C_1| > h) \leq \sum_j \mathbf{P}(|C_j| > h) \leq \sum_v \mathbf{P}(|C(v)| > h) \leq \sum_v n^{-c''A} = n^{1-c''A}$$

ce qui tend vers 0 quand $A \rightarrow \infty$ ou $n \rightarrow \infty$ si $A > \frac{1}{c''}$. \square

3.3 Majoration critique

Fixons un sommet v de $G(n, \frac{1}{n})$. On va analyser la composante de v dans $G(n, \frac{1}{n})$ en utilisant le processus d'exploration de 2.2 et la notion établie dans cette section. On définit S_t et γ comme dans le théorème 30 en utilisant le couplage détaillé en 2.3 et on prend un $H > 0$ (dont la valeur sera fixée plus tard).

La preuve est divisée en deux : dans le lemme 11 on majore γ et la probabilité que S_t monte au-dessus de H . Ce resultat est utilisé dans le théorème 12 pour montrer une majoration simple de la taille de la plus grande composante connexe.

Lemme 11 *On a*

$$\mathbf{P}(S_\gamma \geq H) \leq \frac{1}{H} \tag{6}$$

et

$$\mathbf{E}\gamma \leq H + 3 \tag{7}$$

PREUVE. Comme S_t est un martingale, on applique la théorème 25 à S_t , ce qui donne $1 = S_0 = \mathbf{E}[S_\gamma] \geq H\mathbf{P}(S_\gamma \geq H)$ et on obtient (6). On écrit $S_\gamma^2 - H^2 = 2H(S_\gamma - H) + (S_\gamma - H)^2$ et on applique le corollaire 31 avec $f(x) = 2Hx + x^2$ et $X \sim \text{Bin}(n, \frac{1}{n})$:

$$\mathbf{E}[S_\gamma^2 - H^2 | S_\gamma \geq H] = \mathbf{E}[f(S_\gamma - H) | S_\gamma \geq H] \leq \mathbf{E}[2HX + X^2] = 2H + (1 - \frac{1}{n})$$

ce qui donne $\forall H \geq 3$

$$\mathbf{E}[S_\gamma^2 | S_\gamma \geq H] \leq H^2 + 2H + 2 \leq H^2 + 3H \quad (8)$$

On constate que $B_t = S_t^2 - (1 - \frac{1}{n})t$ est aussi une martingale :

$$\begin{aligned} \mathbf{E}[B_t | \mathcal{F}_{t-1}] &= \mathbf{E}[S_t^2 - (1 - \frac{1}{n})t | \mathcal{F}_{t-1}] \\ &= \mathbf{E}[S_{t-1}^2 + 2S_{t-1}(\xi_t - 1) + (\xi_t - 1)^2 | \mathcal{F}_{t-1}] - (1 - \frac{1}{n})t \\ &= S_{t-1}^2 + 2S_{t-1}\mathbf{E}[(\xi_t - 1)] + \mathbf{E}[(\xi_t - 1)^2] - (1 - \frac{1}{n})t \\ &= S_{t-1}^2 + 2S_{t-1}0 + (1 - \frac{1}{n}) - (1 - \frac{1}{n})t \\ &= S_{t-1}^2 - (1 - \frac{1}{n})(t - 1) \\ &= B_{t-1} \end{aligned}$$

et en appliquant le théorème 25 et les equations (6) et (8) on obtient

$$\begin{aligned} 1 &= \mathbf{E}[B_0] = \mathbf{E}[B_\gamma] = \mathbf{E}[S_\gamma^2] - (1 - \frac{1}{n})\mathbf{E}\gamma \\ &= \mathbf{E}[S_\gamma^2 | S_\gamma \geq H]\mathbf{P}(S_\gamma \geq H) - (1 - \frac{1}{n})\mathbf{E}\gamma \leq \frac{H^2 + 3H}{H} - (1 - \frac{1}{n})\mathbf{E}\gamma \end{aligned}$$

donc pour $H \leq n - 3$

$$\mathbf{E}\gamma \leq \frac{n}{n-1}(H+2) \leq H+3.$$

□

Théorème 12 (Borne simple) *Pour chaque $A > 0$ on a*

$$\mathbf{P}(|C_1| \geq An^{2/3}) \leq \frac{3}{A^{3/2}}. \quad (9)$$

PREUVE. On a pour $3 \leq H \leq n - 3$ et en utilisant (7)

$$\mathbf{P}(\gamma \geq H^2) \leq \frac{H+3}{H^2} \leq \frac{H}{2}.$$

Maintenant on définit $\gamma^* = \gamma \wedge H^2$, ce qui donne pour $H < \sqrt{n}$

$$\begin{aligned} &\mathbf{P}(S_{\gamma^*} > 0) \\ &= \mathbf{P}(S_{\gamma^*} > 0, \gamma < H^2) + \mathbf{P}(S_{\gamma^*} > 0, \gamma < H^2) \\ &\leq \mathbf{P}(S_\gamma \geq H) + \mathbf{P}(\gamma \geq H) \\ &\leq \frac{1}{H} + \frac{2}{H} = \frac{3}{H} \end{aligned} \quad \text{par (6).}$$

Prend $9 \leq T = H^2 < n$, ce qui donne $\{|C(v)| \geq H^2\} \Rightarrow \{S_{\gamma^*} > 0\}$ et

$$\mathbf{P}(|C(v)| \geq T) \leq \mathbf{P}(S_{\gamma^*} > 0) \leq \frac{3}{\sqrt{T}}.$$

En définissant $G_T = \{v \in V : |C(v)| \geq T\}$ et utilisant Markov 15 on arrive à

$$\begin{aligned} \mathbf{P}(|C_1| \geq T) &\leq \mathbf{P}(|G_T| \geq T) \leq \frac{\mathbf{E}|G_T|}{T} \\ &= \frac{\mathbf{E} \sum_{i=1}^n \mathbf{1}_{\{v_i \in G_T\}}}{T} = \frac{\sum_{i=1}^n \mathbf{E}[v_i \in G_T]}{T} \\ &= \frac{n\mathbf{P}(|C(v)| \geq T)}{T} \leq \frac{3n}{T^{3/2}}, \end{aligned}$$

dont on deduit le resultat en prenant $T = An^{2/3}$ pour $A > 0$. \square

3.4 Minoration critique

Théorème 13 (Minoration critique) *Pour tout $0 < \theta < \frac{1}{3}$ et $n > 8\theta^{-3/5}$, le graphe aléatoire $G(n, \frac{1}{n})$ satisfait*

$$\mathbf{P}(|C_1| < \lceil \theta n^{2/3} \rceil) \leq 60\theta^{3/5}. \quad (10)$$

PREUVE. Soient h, T_1 et T_2 des entiers positifs, à préciser plus tard. La preuve est divisée en trois parties : dans la première partie on vérifie qu'avec une probabilité assez grande Y_t monte jusqu'à h avant le temps T_1 , et dans la deuxième partie on s'assure que Y_t , après avoir dépassé h , reste positif pendant un temps T_2 avec une assez grande probabilité. Finalement, dans la troisième partie tous est mis ensemble pour minorer la taille de la plus grande composante connexe..

Partie 1 : C'est partie peut être aussi effectué avec X_t à la place de Y_t avec un calcul un peu simplifié. On définit

$$\tau_h = \min\{1 \leq t \leq T_1 : Y_t \geq h\} \wedge T_1 \quad (11)$$

On a pour $2 \leq h \leq \frac{\sqrt{n}}{4}$, $T_1 = \lceil \frac{n}{8h} \rceil$ et $0 < k < h$:

$$\begin{aligned} &\mathbf{E}[Y_t^2 - Y_{t-1}^2 | Y_{t-1} = k, t-1 < \tau_h] \\ &= \mathbf{E}[(\eta_t - 1)^2 + 2Y_{t-1}(\eta_t - 1) | Y_{t-1} = k, t-1 < \tau_h] \\ &= \mathbf{E}[\eta_t^2 | Y_{t-1} = k, t-1 < \tau_h] + 2k\mathbf{E}[\eta_t - 1 | Y_{t-1} = k, t-1 < \tau_h] \end{aligned}$$

avec $\{\tau_h > t - 1\} \Rightarrow \{k < h\}$ et $\eta_t \sim \text{Bin}(n - (t - 1) - k, \frac{1}{n})$

$$\begin{aligned}
&= \frac{(n - (t - 1) - k)(n - 1)}{n^2} + \frac{(n - (t - 1) - k)^2}{n^2} + 2k \left[n - (t - 1) - k \frac{1}{n} - 1 \right] \\
&= \frac{(n - t + 1 - k)(2n - t - k)}{n^2} - \frac{2k(t - 1 + k)}{n} \\
&\geq \frac{(n - t - k)(2n - t - k)}{n^2} - \frac{2k(t + k)}{n} \\
&= \frac{2n^2 - 3nt - 3nk + (t + k)^2 - 2ntk - 2nk^2}{n^2} \\
&\geq 2 - \frac{3}{8} - \frac{3}{4\sqrt{n}} + \frac{4}{n^2} - \frac{1}{4} - \frac{1}{8} \\
&\geq \frac{5}{4} - \frac{3}{4} \\
&\geq \frac{1}{2}
\end{aligned}$$

et pour $k = 0$ on a $\eta_t \sim \text{Bin}(n - (t - 1) - 1, \frac{1}{n}) \sim \text{Bin}(n - t, \frac{1}{n})$

$$\begin{aligned}
&\mathbf{E}[Y_t^2 - Y_{t-1}^2 | Y_{t-1} = 0, t - 1 < \tau_h] = \mathbf{E}[\eta_t^2 | Y_{t-1} = 0, t - 1 < \tau_h] \\
&= \frac{n - t}{n} \left(1 - \frac{1}{n}\right) + \left(\frac{n - t}{n}\right)^2 = \frac{(n - t)(n - 1 + n - t)}{n^2} \\
&= \frac{(n - t)(2n - t + 1)}{n^2} \geq \frac{2n^2 - 3nt + t^2 + n - t}{n^2} \\
&\geq \frac{2n^2 - 2n\lceil \frac{n}{16} \rceil - \lceil \frac{n}{16} \rceil}{n^2} \geq 1
\end{aligned}$$

ce qui donne

$$\mathbf{E}[Y_t^2 - Y_{t-1}^2 | Y_{t-1}, t - 1 < \tau_h] \geq \frac{1}{2} \quad (12)$$

qui fait de $Y_{t \wedge \tau_h}^2 - \frac{t \wedge \tau_h}{2}$ une sous-martingale. Maintenant on calcule que pour $0 < l \leq T_1, 0 < r < h$ et $c \in \mathbb{R}$ on a

$$\begin{aligned}
&\mathbf{P}(Y_{\tau_h} - h > c | \tau_h = l, Y_{l-1} = h - r, Y_l \geq h) \\
&= \mathbf{P}(Y_l - h > c | Y_{l-1} = h - r, Y_l \geq h, \eta_l \geq r) \\
&= \mathbf{P}(Y_{l-1} + \eta_l - 1 - h > c | Y_{l-1} = h - r, \eta_l \geq r) \\
&= \mathbf{P}(h - r + \eta_l - 1 - h > c | Y_{l-1} = h - r, \eta_l \geq r) \\
&= \mathbf{P}(\eta_l - r > c + 1 | N_{l-1} = n - (l - 1) - (h - r), \eta_l \geq r)
\end{aligned}$$

ici $\eta_l \sim \text{Bin}(n - (l - 1) - (h - r), \frac{1}{n})$

$$\begin{aligned}
&\leq \mathbf{P}\left(\text{Bin}\left(n - (l - 1) - (h - r), \frac{1}{n}\right) > c + 1\right) \quad \text{par proposition 28} \\
&\leq \mathbf{P}\left(\text{Bin}\left(n, \frac{1}{n}\right) > c + 1\right) \\
&\leq \mathbf{P}\left(\text{Bin}\left(n, \frac{1}{n}\right) > c\right)
\end{aligned}$$

et pour le cas $r = h$

$$\begin{aligned}
& \mathbf{P}(Y_{\tau_h} - h > c | \tau_h = l, Y_{l-1} = 0, Y_l \geq h) \\
&= \mathbf{P}(Y_l - h > c | Y_{l-1} = 0, Y_l \geq h, \eta_l \geq h) \\
&= \mathbf{P}(Y_{l-1} + \eta_l - h > c | Y_{l-1} = 0, \eta_l \geq h) \\
&= \mathbf{P}(h - h + \eta_l - h > c | Y_{l-1} = 0, \eta_l \geq h) \\
&= \mathbf{P}(\eta_l - h > c | N_{l-1} = n - (l - 1), \eta_l \geq h)
\end{aligned}$$

ici $\eta_l \sim \text{Bin}(n - (l - 1), \frac{1}{n})$

$$\begin{aligned}
&\leq \mathbf{P}(\text{Bin}(n - (l - 1), \frac{1}{n}) > c) \quad \text{par proposition 28} \\
&\leq \mathbf{P}(\text{Bin}(n, \frac{1}{n}) > c)
\end{aligned}$$

On applique les lemmata 29 et 18 pour obtenir pour $h \geq 3$

$$\begin{aligned}
& \mathbf{E}[Y_{\tau_h}^2 | Y_{\tau_h} \geq h] \\
&= \mathbf{E}[h^2 + 2h(Y_{\tau_h} - h) + (Y_{\tau_h} - h)^2 | Y_{\tau_h} \geq h] \\
&\leq h^2 + 2h\mathbf{E}[X] + \mathbf{E}[X^2] \\
&= h^2 + 2h + (\frac{n-1}{n} + 1) \\
&\leq 2h^2
\end{aligned}$$

Maintenant on applique le théorème d'arrêt à la sous-martingale $Y_{t \wedge \tau_h}^2 - \frac{t \wedge \tau_h}{2}$

$$0 \leq 1 = \mathbf{E}[Y_0^2] \leq \mathbf{E}[Y_{\tau_h}^2 - \frac{\tau_h}{2}] \Rightarrow \mathbf{E}[\frac{\tau_h}{2}] \leq \mathbf{E}[Y_{\tau_h}^2]$$

or

$$\begin{aligned}
& \frac{T_1}{2} \mathbf{P}(\tau_h = T_1) \\
&\leq \frac{1}{2} \mathbf{E}[\tau_h] \leq \mathbf{E}[Y_{\tau_h}^2] \\
&\leq \mathbf{E}[Y_{\tau_h}^2 | Y_{\tau_h} \geq h] \mathbf{P}(Y_{\tau_h} \geq h) + \mathbf{E}[Y_{\tau_h}^2 | Y_{\tau_h} < h] \mathbf{P}(Y_{\tau_h} < h) \\
&\leq 2h^2 + h^2 \\
&= 3h^2
\end{aligned}$$

et finalement

$$\mathbf{P}(\tau_h = T_1) \leq \frac{6h^2}{T_1} \leq \frac{48h^3}{n}. \quad (13)$$

Partie 2 : On définit

$$\tau_0 = \min\{t \leq T_2 : Y_{\tau_h+t} = 0\} \wedge T_2 \quad (14)$$

et un nouveau processus

$$B_t = h - \min\{h, Y_{\tau_h+t}\} = \min\{0, h - Y_{\tau_h+t}\}. \quad (15)$$

On a pour $0 < k < h, 2 < h < \frac{\sqrt{n}}{4}, 0 \leq l \leq T_2 \leq \frac{n}{8h} \leq \frac{n}{8}$ et $1 < t \leq \tau_0 \leq \frac{n}{8h} \leq \frac{n}{8}$

$$\begin{aligned} & \mathbf{E}[B_t^2 - B_{t-1}^2 | Y_{m+t-1} = k, t-1 < \tau_0, \tau_h = l] \\ &= \mathbf{E}[(1 - \eta_{\tau_h+t})^2 + 2(1 - \eta_{\tau_h+t})B_{t-1} | Y_{m+t-1} = k, t-1 < \tau_0, \tau_h = l] \end{aligned}$$

ici on voit que $\eta_{\tau_h+t} \sim \text{Bin}(n - (l + t - 1) - k, \frac{1}{n})$

$$\begin{aligned} &= \frac{(n - l - t + 1 - k)(n - 1 + n - l - t + 1 - k)}{n^2} \\ &+ 2 \frac{n - n + l + t - 1 + k}{n} (h - k) \\ &= \frac{(n - l - t - k + 1)(2n - l - t - k) + 2n(l + t + k - 1)(h - k)}{n^2} \\ &= \frac{2n^2 - 3n(l + k + t) + (l + k + t)^2 + 2n - (l + k + t)}{n^2} \\ &+ \frac{2n(l + k + t)(h - k) - 2n(h - k)}{n^2} \end{aligned}$$

observe que $2 \leq l + k + t \leq \frac{n}{4h} + h \leq \frac{n}{4} + \frac{\sqrt{n}}{4} \leq \frac{n}{2}$

$$\begin{aligned} &\leq \frac{2n^2 - 6n + (\frac{n}{2})^2 + 2n - 2 + 2n(\frac{n}{4h} + h)h - 2n}{n^2} \\ &\leq \frac{2\frac{3}{4}n^2 + 2n(\frac{\sqrt{n}}{4})^2}{n^2} \\ &= 2\frac{3}{4} + \frac{1}{8} \\ &\leq 3 \end{aligned}$$

ce qui reste vrai pour $B_{t-1} = 0 \Leftrightarrow Y_{\tau_h+t-1} = k \geq h$, donc formellement

$$\begin{aligned} & \mathbf{E}[B_t^2 - B_{t-1}^2 | Y_{m+t-1} = k \geq h, t-1 < \tau_0, \tau_h = l] \\ &= \mathbf{E}[B_t^2 | Y_{m+l} = j \geq h - 1, Y_{m+l-1} = k \geq h, t-1 < \tau_0, \tau_h = l] \\ &\leq 1 \end{aligned}$$

car Y_t décroît au maximum de 1 entre $t-1$ et t (voir (1d)). Tout ça fait de $B_{t \wedge \tau_0}^2 - 3(t \wedge \tau_0)$ une sur-martingale. Ensuite on utilise la notation suivante $\mathbf{P}_h(\star) = \mathbf{P}(\star | Y_{\tau_h} \geq h)$ et $\mathbf{E}_h(\star) = \mathbf{E}(\star | Y_{\tau_h} \geq h)$ et le théorème d'arrêt 25 pour obtenir

$$0 = \mathbf{E}_h[B_0^2] \geq \mathbf{E}_h[B_{\tau_0}^2 - 3\tau_0] \Rightarrow \mathbf{E}_h[B_{\tau_0}^2] \leq 3\mathbf{E}_h[\tau_0] \leq 3T_2$$

ce qui donne avec l'aide d'inégalité de Markov 15

$$\mathbf{P}_h(\tau_0 < T_2) \leq \mathbf{P}_h(B_{\tau_0} \geq h) \leq \frac{\mathbf{E}_h[B_{\tau_0}^2]}{h^2} \leq \frac{3T_2}{h^2}. \quad (16)$$

Partie 3 : En utilisant (13) et (16) on obtient

$$\begin{aligned}
& \mathbf{P}(\tau_0 < T_2) \\
&= \mathbf{P}(Y_{\tau_h} \geq h, \tau_0 < T_2) + \mathbf{P}(\tau_h = T_1, Y_{\tau_h} < h, \tau_0 < T_2) \\
&\leq \mathbf{P}(Y_{\tau_h} \geq h, \tau_0 < T_2) + \mathbf{P}(\tau_h = T_1) \\
&= \mathbf{P}(\tau_0 < T_2 | Y_{\tau_h} \geq h) \mathbf{P}(Y_{\tau_h} \geq h) + \mathbf{P}(\tau_h = T_1) \\
&\leq \mathbf{P}_h(\tau_0 < T_2) + \mathbf{P}(\tau_h = T_1) \\
&\leq \frac{3T_2}{h^2} + \frac{48h^3}{n}
\end{aligned}$$

On choisit maintenant $T_2 = \theta n^{2/3}$ et $h = \alpha \theta^{1/5} n^{1/3}$. On finit par

$$\mathbf{P}(|C_1| < T_2) \leq \mathbf{P}(\tau_0 < T_2) \leq \left(\frac{3}{\alpha^2} + 48\alpha^3 \right) \theta^{3/5} \leq (12 + 48)\theta^{3/5} = 60\theta^{3/5}$$

qui est minimisé pour un choix de $\alpha = 24^{-1/5} \in [\frac{1}{2}, 1]$ et $0 < \theta < \frac{1}{3}$, ce qui respecte toutes les conditions de la partie 2. Finalement, on voit qu'on doit avoir $3 \leq h = 24^{-1/5} \theta^{1/5} n^{1/3}$, qui est garantie quand $n \geq 8\theta^{-3/5}$. \square

3.5 Minoration sur-critique

Théorème 14 (Minoration sur-critique) *Dans le graphe $G(n, \frac{c}{n})$, où $c > 1$, on définit $\forall \varepsilon \in]0, 1[$: $B = \varepsilon \frac{c-1-\log c}{c}$ et $\forall \delta \in]0, B[$ on a*

$$\mathbf{P}(|C_1| < \delta n) \leq e^{-(B-\delta)^2 n / (2B)} \quad (17)$$

pour $n \geq N_{c,\varepsilon}$.

PREUVE. Pour cette preuve on se place dans le cadre du graphe $G(n, \frac{c}{n})$. Pour $t = \alpha n \in O(n)$ on a $1 - (1 - \frac{c}{n})^{\alpha n} \approx 1 - e^{-c\alpha}$ pour n assez grand. Donc maintenant on calcule :

$$\begin{aligned}
& \mathbf{E}[X_{\alpha n}] \\
&= \mathbf{E}[\text{Bin}(n-1, 1 - (1 - \frac{c}{n})^{\alpha n}) + 1 - \alpha n] \\
&= (n-1)[1 - (1 - \frac{c}{n})^{\alpha n}] + 1 - \alpha n \\
&= (n-1)[1 - (1 - \frac{c}{n})^{\alpha n} - \alpha] + (1 - \alpha) \\
&\geq \varepsilon n(1 - e^{-c\alpha} - \alpha) \quad \forall n \geq N_{\varepsilon,c,\alpha} \\
&= n\varepsilon f(\alpha)
\end{aligned}$$

En étudiant la fonction $f(x) = 1 - e^{-cx} - x$ on découvre qu'elle a les propriétés suivantes : elle vaut 0 deux fois (une fois en 0 et une fois en $\gamma = \gamma(c) \in]0, 1[$) et elle a un maximum global en $\beta = \beta(c) = \frac{\log c}{c} \in]0, \gamma[$, la valeur de ce maximum étant $m = m(c) = f(\beta) = 1 - \frac{1+\log c}{c} = \frac{c-1-\log c}{c}$. On remarque que la même expression apparaît dans la preuve de 10 et est croissante en c . Maintenant on

definit $B = \varepsilon m$ et $\delta \in]0, B[$ et on obtient

$$\begin{aligned} \mathbf{P}(|C_1| \geq \delta n) &\geq \mathbf{P}(Y_{\beta n} \geq \delta n) \geq \mathbf{P}(X_{\beta n} \geq \delta n) \\ &= 1 - \mathbf{P}(X_{\beta n} < \delta n) = 1 - \mathbf{P}(X_{\beta n} < (1 - \frac{B - \delta}{B})Bn) \\ &\geq 1 - e^{-\left(\frac{B - \delta}{B}\right)^2 \frac{Bn}{2}} = 1 - e^{-\frac{(B - \delta)^2 n}{2B}} \end{aligned}$$

ce qui tend vers 1 quand $n \rightarrow \infty$. Finalement, pose $n \geq N_{c, \varepsilon} = N_{\varepsilon, c, \beta}$. \square

4 Notations et théorèmes de base

4.1 Énoncés de la probabilité

Dans cette section je vais énoncer les définitions et théorèmes qui sont liés à la probabilité. On commence avec 2 lemmes de base, voir [4] page 55, et les énoncés reliés à la relation d'ordre stochastique, voir [4] page 107 :

Lemme 15 (Inégalité de Markov) Soit $p \in \mathbb{N}^*$. Si $f \in L^p(\mu)$ alors pour tout $\varepsilon > 0$,

$$\mu(|f| > \varepsilon) \leq \frac{1}{\varepsilon^p} \|f\|_p^p.$$

Lemme 16 (Inégalité de Hölder, Cauchy-Schwarz pour $p = 2$) Soient $1 \leq p, q \leq \infty$ tels que $\frac{1}{p} + \frac{1}{q} = 1$. Si $f \in L^p(\mu)$ et $g \in L^q(\mu)$ alors

$$\|fg\|_1 \leq \|f\|_p \|g\|_q.$$

Définition 17 (L'ordre stochastique) Soient X et Y deux variables aléatoires réelles. La variable X est dite stochastiquement plus grande que la variable Y , et nous noterons $X \stackrel{st}{\geq} Y$, si

$$\mathbf{P}(X > c) \geq \mathbf{P}(Y > c), \quad \forall c \in \mathbb{R}.$$

Lemme 18 On a $X \stackrel{st}{\geq} Y$ si et seulement si

$$\mathbf{E}[f(X)] \geq \mathbf{E}[f(Y)]$$

pour toute fonction borélienne f croissante.

Voici les énoncés reliés aux filtrations, espérances conditionnelles et martingales, voir [4], chapitre 5. Dans la suite $(\Omega, \mathcal{F}, \mathbf{P})$ désigne l'espace de probabilité et \mathcal{G} une sous-tribu de \mathcal{F} .

Définition 19 (Probabilité conditionnelle) Soit A un évènement dans $(\Omega, \mathcal{F}, \mathbf{P})$. La probabilité conditionnelle de A sachant \mathcal{G} est une variable aléatoire \mathcal{G} -mesurable (que nous noterons $\mathbf{P}(A|\mathcal{G})$) vérifiant

$$\int_B \mathbf{P}(A|\mathcal{G}) d\mathbf{P} = \mathbf{P}(A \cap B), \quad B \in \mathcal{G}$$

Si $\{B_i\}_{i \in I}$ est un système de constituents fini ou dénombrable de Ω et si $\mathcal{G} = \sigma(B_i : i \in I)$, alors la relation suivante est vérifiée :

$$\mathbf{P}(A|\mathcal{G}) = \sum_{i \in I} \frac{\mathbf{P}(A \cap B_i)}{\mathbf{P}(B_i)} \mathbf{1}_{\{B_i\}}.$$

Définition 20 (Espérance conditionnelle) Soit X une variable aléatoire définie sur $(\Omega, \mathcal{F}, \mathbf{P})$. L'espérance conditionnelle d'une variable aléatoire X positive sachant \mathcal{G} est une variable aléatoire \mathcal{G} -mesurable, à valeurs dans $\overline{\mathbb{R}}_+$, notée $\mathbf{E}(X|\mathcal{G})$, et telle que

$$\int_C X d\mathbf{P} = \int_C \mathbf{E}(X|\mathcal{G}) d\mathbf{P}.$$

Si X est une variable aléatoire réelle et

$$\min[\mathbf{E}(X^+|\mathcal{G}), \mathbf{E}(X^-|\mathcal{G})] < \infty$$

l'espérance conditionnelle de X sachant \mathcal{G} est définie par

$$\mathbf{E}(X|\mathcal{G}) = \mathbf{E}(X^+|\mathcal{G}) - \mathbf{E}(X^-|\mathcal{G}).$$

Théorème 21 (Indépendance de variables aléatoires) Si \mathcal{G}_1 et \mathcal{G}_2 sont 2 sous-tribus de \mathcal{F} indépendantes et X est une variable aléatoire intégrable telle que $\sigma(X), \mathcal{G}_1$ et \mathcal{G}_2 sont indépendantes, alors

$$\mathbf{E}[X|\sigma(\mathcal{G}_1, \mathcal{G}_2)] = \mathbf{E}[X|\mathcal{G}_1] \quad ps.$$

Définition 22 (Filtration à temps discret) Une filtration \mathbf{F} dans $(\Omega, \mathcal{F}, \mathbf{P})$ est une suite croissante $(\mathcal{F}_n)_{n \geq 0}$ de sous-tribus de \mathcal{F} .

Définition 23 (Temps d'arrêt à temps discret) Soit $\mathbf{F} = (\mathcal{F}_n)$ une filtration de \mathcal{F} . Un temps d'arrêt pour \mathbf{F} (ou adapté à \mathbf{F} ou un \mathbf{F} -temps d'arrêt) est une variable aléatoire T à valeurs dans $\overline{\mathbb{N}}$ vérifiant une des conditions (équivalentes) suivantes :

1. $(T \leq n) \in \mathcal{F}_n, \quad \forall n \in \mathbb{N};$
2. $(T = n) \in \mathcal{F}_n, \quad \forall n \in \mathbb{N};$
3. $\mathbf{1}_{\{T=n\}}$ est \mathcal{F}_n -mesurable $\forall n \in \mathbb{N}$.

Définition 24 (Martingale à temps discret) Une suite aléatoire (X_n) définie sur une base stochastique $B = (\Omega, \mathcal{F}, \mathbf{F}, \mathbf{P})$ est une

1. \mathbf{F} -martingale (ou martingale pour \mathbf{F}) si les 3 conditions suivantes sont vérifiées :
 - (a) X_n est intégrable $\forall n \in \mathbb{N};$
 - (b) X_n est \mathbf{F} -adaptée $\forall n \in \mathbb{N};$
 - (c) $X_n = \mathbf{E}(X_{n+1}|\mathbf{F}_n), \forall n \in \mathbb{N}$.
2. \mathbf{F} -sous-martingale (ou sous-martingale pour \mathbf{F}) si elle vérifie a., b. et \leq dans c.
3. \mathbf{F} -sur-martingale (ou sur-martingale pour \mathbf{F}) si elle vérifie a., b. et \geq dans c.

Théorème 25 (Théorème d'arrêt) Soit (X_n) une martingale (resp. sous-martingale) pour $\mathbf{F} = (\mathcal{F}_n)$. Si S et T sont deux temps d'arrêt adaptés à \mathbf{F} tels que

1. $\max(\mathbf{E}(|X_T|), \mathbf{E}(|X_S|)) < \infty$,
2. $\liminf \mathbf{E}[|X_n| \mathbf{1}_{\{T > n\}}] = 0$ et $\liminf \mathbf{E}[|X_n| \mathbf{1}_{\{S > n\}}] = 0$,

alors

$$X_{S \wedge T} = \mathbf{E}(X_T | \mathcal{F}_S) \quad (\text{resp. } \geq) \quad \mathbf{P}\text{-p.s..}$$

Dans le cas où on a une sur-martingale bornée et finie le théorème d'arrêt est aussi applicable.

4.2 Énoncés de la distribution binomiale

La proposition suivante est essentiellement prise de [1], page 239, théorèmes A.8 - A.13 :

Proposition 26 (Déviations large d'une loi binomiale) Soit $Y \sim \text{Bin}(n, p)$ et $\beta > 0$ alors

$$\mathbf{P}(Y < (1 - \beta)np) \leq e^{-\beta^2 np/2} \quad (18a)$$

et

$$\mathbf{P}(Y > (1 + \beta)np) \leq [e^\beta(1 + \beta)^{-(1+\beta)}]^{np} \quad (18b)$$

PREUVE. Soit $X \sim \text{Bin}(n, p) - np$, $a > 0$ et $\lambda > 0$, alors

$$\begin{aligned} & \mathbf{P}(X < -a) \\ &= \mathbf{P}(e^{-\lambda X} > e^{\lambda a}) \\ &\leq \mathbf{E}[e^{-\lambda X}] e^{-\lambda a} && \text{par l'inégalité de Markov 15} \\ &= [pe^{-\lambda(1-p)} + (1-p)e^{\lambda p}]^n e^{-\lambda a} \\ &= e^{\lambda np} [1 + (e^{-\lambda} - 1)p]^n e^{-\lambda a} \\ &\leq e^{\lambda np + np(e^{-\lambda} - 1) - \lambda a} && \text{par } 1 + u \leq e^u \forall u \in \mathbb{R} \\ &= e^{np(e^{-\lambda} + \lambda - 1) - \lambda a} \\ &< e^{np\lambda^2/2 - \lambda a} && \text{par } e^{-\lambda} < 1 - \lambda + \frac{\lambda^2}{2} \forall \lambda > 0 \\ &= e^{npa^2/2n^2p^2 - (a/np)a} && \text{en prenant } \lambda = \frac{a}{np} \\ &= e^{a^2/2np - a^2/np} \\ &= e^{-a^2/2np} \end{aligned}$$

ce qui donne

$$\begin{aligned} \mathbf{P}(Y < (1 - \beta)np) &= \mathbf{P}(Y - np < -\beta np) = \mathbf{P}(X < -\beta np) \\ &< e^{-(\beta np)^2/(2np)} = e^{-\beta^2 np/2} \end{aligned}$$

et pour le deuxième cas

$$\begin{aligned}
& \mathbf{P}(X > a) \\
&= \mathbf{P}(e^{\lambda X} > e^{\lambda a}) \\
&\leq \mathbf{E}[e^{\lambda X}]e^{-\lambda a} && \text{par l'inégalité de Markov 15} \\
&= e^{-\lambda np}[pe^{\lambda} + (1-p)]^n e^{-\lambda a} \\
&= e^{-\log(1+a/np)(np+a)}[pe^{\log(1+a/np)} + 1-p]^n && \text{prenant } \lambda = \log\left(1 + \frac{a}{np}\right) \\
&= e^{-\log(1+a/np)(np+a)}\left[p\left(1 + \frac{a}{np}\right) + 1-p\right]^n \\
&= e^{-\log(1+a/np)(np+a)}\left(1 + \frac{a}{n}\right)^n \\
&\leq e^{a-\log(1+a/np)(np+a)} && \left(1 + \frac{a}{n}\right)^n \leq e^a \quad \forall a > 0 \\
&\leq e^{a-(a/np-a^2/2p^2n^2)(np+a)} && \log(1+u) \geq u - \frac{u^2}{2} \quad \forall u > 0 \\
&= e^{a-a-a^2/pn+a^2/2pn+a^3/2p^2n^2} \\
&= e^{-a^2/2pn+a^3/2p^2n^2}
\end{aligned}$$

or

$$\begin{aligned}
& \mathbf{P}(Y > (1+\beta)np) \\
&= \mathbf{P}(Y - np > \beta np) \\
&= \mathbf{P}(X > \beta np) \\
&\leq e^{\beta np - \log(1+\beta np/np)(np+\beta np)} \\
&= e^{\beta np - \log(1+\beta)(1+\beta)np} \\
&= [e^{\beta}(1+\beta)^{-(1+\beta)}]^{np}
\end{aligned}$$

□

Voilà une petite proposition sur la distribution d'un variable binomiale itérée :

Proposition 27 (Distribution binomiale itérée) *Soit $Y \sim \text{Bin}(m, \tilde{p})$ et X une loi telle que toutes ses lois conditionnelles sachant Y sont de la forme $X\mathbf{1}_{\{Y=k\}} \sim \text{Bin}(Y, p)\mathbf{1}_{\{Y=k\}}$. Alors la probabilité conditionnelle de X sachant Y est définie par*

$$\mathbf{P}(X = k|Y) = \binom{Y}{k} p^k (1-p)^{Y-k} \mathbf{1}_{\{Y \geq k\}} \text{ et } X \sim \text{Bin}(m, p\tilde{p}).$$

PREUVE. Il y a deux facons de regarder cet énoncé. La première est intuitive : on a m éléments, dont chacun est choisi avec probabilité \tilde{p} et après avec probabilité p — donc avec probabilité $p\tilde{p}$. La deuxième est par le calcul. En notant $q = 1-p$ et $\tilde{q} = 1-\tilde{p}$ on a pour $\forall k \in \{0, \dots, m\}$:

$$\begin{aligned}
& \mathbf{E} [\mathbf{P}(X = k|Y)] \\
&= \mathbf{E} \left[\sum_{l=0}^m \frac{\mathbf{P}(X = k, Y = l)}{\mathbf{P}(Y = l)} \mathbf{1}_{\{Y=l\}} \right] \\
&= \sum_{l=0}^m \frac{\binom{l}{k} p^k q^{l-k} \binom{m}{l} \tilde{p}^l \tilde{q}^{m-l}}{\binom{m}{l} \tilde{p}^l \tilde{p}^{m-l}} \mathbf{E} [\mathbf{1}_{\{Y=l\}}] \\
&= \sum_{l \geq k}^m \binom{l}{k} p^k q^{l-k} \mathbf{P}(Y = l) \\
&= \sum_{l \geq k}^m \binom{l}{k} p^k q^{l-k} \binom{m}{l} \tilde{p}^l \tilde{q}^{m-l} \\
&= \binom{m}{k} (p\tilde{p})^k \sum_{l \geq k}^m \binom{m-k}{l-k} q^{l-k} \tilde{p}^{l-k} \tilde{q}^{m-l} \\
&= \binom{m}{k} (p\tilde{p})^k \underbrace{\sum_{l \geq k}^m \binom{m-k}{l-k} \tilde{p}^{l-k} \sum_{i=0}^{l-k} \binom{l-k}{i} (-1)^i p^i \sum_{j=0}^{m-l} \binom{m-l}{j} (-1)^j \tilde{p}^j}_{(\star)}
\end{aligned}$$

Ainsi on va montrer que que $(\star) = (1 - p\tilde{p})^{n-k}$. Pour ça on montre l'égalité de tous les coefficients de $p^s \tilde{p}^t$:

$$\begin{aligned}
& [p^s \tilde{p}^t](\star) \\
&= [p^s \tilde{p}^t] \sum_{l \geq k}^m \binom{m-k}{l-k} \tilde{p}^{l-k} \sum_{i=0}^{l-k} \binom{l-k}{i} (-1)^i p^i \sum_{j=0}^{m-l} \binom{m-l}{j} (-1)^j \tilde{p}^j \\
&= \sum_{l \geq k}^m \binom{m-k}{l-k} \binom{l-k}{s} (-1)^s \binom{m-l}{t-(l-k)} (-1)^{t-(l-k)} \\
&= \sum_{l \geq k}^m \frac{(m-k)!}{(m-l)!(l-k)!} \frac{(l-k)!}{s!(l-k-s)!} \frac{(m-l)!}{(t+k-l)!(m-t+k)!} (-1)^{s+t+k-l}
\end{aligned}$$

On remarque qu'on a les restrictions $0 \leq s \leq l-k \leq t \leq m-l$.

$$= \frac{(m-k)!}{s!(m-k-t)!} \frac{(-1)^{s+t+k}}{(t-s)!} \sum_{l \geq k}^m \frac{(t-s)!}{(l-k-s)!(t+k-l)!} (-1)^l$$

On substitue $y = l-k$ et observe que $0 \leq y \leq t$ mais aussi $y \leq s$

$$= \frac{(m-k)!}{s!(m-k-t)!} \frac{(-1)^{s+t+k}}{(t-s)!} \sum_{y=0}^s \binom{t-s}{t-y} (-1)^t$$

Cas $s = t$:

$$\begin{aligned}
&= \binom{m-k}{s} \frac{(-1)^{2s-k}}{0!} \underbrace{\sum_{y=0}^s \binom{0}{s-y} (-1)^y}_{=1} \\
&= \binom{m-k}{s} (-1)^k \\
&= [(p\tilde{p})^s] \sum_{i=0}^{m-k} \binom{m-k}{i} (-p\tilde{p})^i \\
&= [(p\tilde{p})^s] (1 - p\tilde{p})^{m-k}
\end{aligned}$$

Cas $s < t$:

$$\begin{aligned}
&= \frac{(m-k)!}{s!(m-k-t)!} \frac{(-1)^{s+t+k}}{(t-s)!} \underbrace{(1-1)^{t-s}}_{=0} \\
&= 0
\end{aligned}$$

d'où

$$\begin{aligned}
\mathbf{P}(X = k) &= \mathbf{E}[\mathbf{1}_{\{X=k\}}] = \mathbf{E}[\mathbf{E}[\mathbf{1}_{\{X=k\}}|Y]] = \mathbf{E}[\mathbf{E}[\mathbf{1}_{\{X=k\}}|Y]] \\
&= \mathbf{E}[\mathbf{P}(X = k|Y)] = \binom{m}{k} (p\tilde{p})^k (1 - p\tilde{p})^{m-k} = \mathbf{P}(Bin(m, p\tilde{p}) = k)
\end{aligned}$$

□

4.3 Le dépassement d'un marche aléatoire

Cette section est dédiée à quantifier le dépassement d'une marche aléatoire et le majorer (en terme d'ordre stochastique) par une loi binomiale.

Proposition 28 (Dépassement d'une variable binomiale) Soient $r \geq 0$ fixé, $\xi \sim Bin(n, p)$ et $X \sim Bin(n, p)$. Alors on a $\forall c \in \mathbb{R}$:

$$\mathbf{P}(\xi - r > c | \xi \geq r) < \mathbf{P}(X > c).$$

PREUVE. Si on a la condition $\{\xi \geq r\}$, on réécrit $\xi = \sum_{i=1}^n I_i$ où $I_i \sim B(p)$. Puis on définit $J(r) = \min\{j : \sum_{i=1}^j I_i = r\}$ et on constate qu'on a $\{\xi \geq r\}$ si et seulement si $\{r \leq J(r) \leq n\}$ (c'est-à-dire que $J(r)$ existe). Maintenant on a

$\forall c \in \mathbb{R} :$

$$\begin{aligned}
& \mathbf{P}(\xi - r > c | \xi \geq r) \\
&= \mathbf{P}\left(\left(\sum_{i=1}^n I_i\right) - r > c | \xi \geq r\right) \\
&= \mathbf{P}\left(\left(\sum_{i=1}^{J(r)} I_i\right) - r + \sum_{i>J(r)}^n I_i > c | \xi \geq r\right) && \text{voir déf de } J(r) \\
&= \mathbf{P}\left(\sum_{i>J(r)}^n I_i > c | \xi \geq r\right) \\
&= \mathbf{P}\left(\sum_{i>J(r)}^n I_i > c | r \leq J(r) \leq n\right) \\
&= \sum_{j=r}^n \frac{\mathbf{P}\left(\sum_{i=j+1}^n I_i > c, J(r) = j\right)}{\sum_{j=r}^n \mathbf{P}(J(r) = j)} \\
&= \sum_{j=r}^n \frac{\mathbf{P}\left(\sum_{i=j+1}^n I_i > c\right) \mathbf{P}(J(r) = j)}{\sum_{j=r}^n \mathbf{P}(J(r) = j)} && \text{car les } \{I_i\}_{i>j} \text{ indépendants de } J(r) \\
&\leq \sum_{j=r}^n \frac{\mathbf{P}(X > c) \mathbf{P}(J(r) = j)}{\sum_{j=r}^n \mathbf{P}(J(r) = j)} && \text{car } \text{Bin}(n-j, p) \stackrel{\text{st}}{\leq} \text{Bin}(n, p) \\
&= \mathbf{P}(X > c)
\end{aligned}$$

ce qui donne le resultat souhaite en utilisant la définition 17. \square

Lemme 29 (Dépassement d'un processus stochastique) *Soit Z_t un processus à valeurs entières en temps discret et ζ un temps d'arrêt tel que pour tout t et un H fixé on a $\{Z_t \geq H\} \Rightarrow \{\zeta = t\}$. Si on a*

$$\mathbf{P}(Z_\zeta - H > c | \zeta = l, Z_{l-1} = H - r, Z_\zeta \geq H) \leq \mathbf{P}(X > c)$$

où $X \sim \text{Bin}(n, p)$, $c \in \mathbb{R}$, $r > 1$, $l > 1$ alors

$$\mathbf{P}(Z_\zeta - H > c | Z_\zeta \geq H) \leq \mathbf{P}(X > c).$$

PREUVE. En utilisant la notation suivante pour la probabilité et l'espérance sous $\{Z_\zeta \geq H\}$:

$$\mathbf{E}_{Z_\zeta \geq H}[\star] = \mathbf{E}[\star | Z_\zeta \geq H] \text{ et } \mathbf{P}_{Z_\zeta \geq H}(\star) = \mathbf{P}(\star | Z_\zeta \geq H)$$

on obtient $\forall c \in \mathbb{R}$ et $H > 0$ fixé :

$$\begin{aligned}
& \mathbf{P}_{Z_\zeta \geq H} [S_\gamma - H > c | \zeta = l, Z_{\zeta-1} = H - r] \\
&= \frac{\mathbf{P}_{Z_\zeta \geq H} (Z_\zeta - H > c, \zeta = l, Z_{\zeta-1} = H - r)}{\mathbf{P}_{Z_\zeta \geq H} (\zeta = l, Z_{\zeta-1} = H - r)} \\
&= \frac{\mathbf{P} (Z_\zeta - H > c, \zeta = l, Z_{\zeta-1} = H - r | Z_\zeta \geq H)}{\mathbf{P} (\zeta = l, Z_{\zeta-1} = H - r | Z_\zeta \geq H)} \\
&= \frac{\mathbf{P} (Z_\zeta - H > c, \zeta = l, Z_{\zeta-1} = H - r, Z_\zeta \geq H) / \mathbf{P} (Z_\zeta \geq H)}{\mathbf{P} (\zeta = l, Z_{\zeta-1} = H - r, Z_\zeta \geq H) / \mathbf{P} (Z_\zeta \geq H)} \\
&= \frac{\mathbf{P} (Z_\zeta - H > c, \zeta = l, Z_{\zeta-1} = H - r, Z_\zeta \geq H)}{\mathbf{P} (\zeta = l, Z_{\zeta-1} = H - r, Z_\zeta \geq H)} \\
&= \mathbf{P} (Z_\zeta - H > c | \zeta = l, Z_{\zeta-1} = H - r, Z_\zeta \geq H) \\
&\stackrel{\text{st}}{\leq} \mathbf{P} (X > c)
\end{aligned}$$

ce qui donne

$$\begin{aligned}
& \mathbf{P} (Z_\zeta - H > c | Z_\zeta \geq H) \\
&= \mathbf{P}_{Z_\zeta \geq H} (Z_\zeta - H > c) \\
&= \mathbf{E}_{Z_\zeta \geq H} [\mathbf{1}_{\{Z_\zeta - H > c\}}] \\
&= \mathbf{E}_{Z_\zeta \geq H} [\mathbf{E}_{Z_\zeta \geq H} [\mathbf{1}_{\{Z_\zeta - H > c\}} | \sigma(\zeta, Z_{\zeta-1})]] \\
&= \mathbf{E}_{Z_\zeta \geq H} \left[\sum_{l,r} \mathbf{E}_{Z_\zeta \geq H} [\mathbf{1}_{\{Z_\zeta - H > c\}} | \zeta = l, Z_{\zeta-1} = H - r] \mathbf{1}_{\{\zeta=l, Z_{\zeta-1}=H-r\}} \right] \\
&= \sum_{l,r} \mathbf{E}_{Z_\zeta \geq H} [\mathbf{P}_{Z_\zeta \geq H} (Z_\zeta - H > c | \zeta = l, Z_{\zeta-1} = H - r) \mathbf{1}_{\{\zeta=l, Z_{\zeta-1}=H-r\}}] \\
&\leq \sum_{l,r} \mathbf{E}_{Z_\zeta \geq H} [\mathbf{P} (X > c) \mathbf{1}_{\{\zeta=l, Z_{\zeta-1}=H-r\}}] \quad \text{par lemme 18} \\
&= \mathbf{P} (X > c) \sum_{l,r} \mathbf{P}_{Z_\zeta \geq H} (\zeta = l, Z_{\zeta-1} = H - r) \\
&= \mathbf{P} (X > c)
\end{aligned}$$

□

Théorème 30 (Dépassement d'une marche aléatoire) Soient $\{\xi_i\}_{i \geq 1}$ des variables i.i.d. avec $\text{Bin}(n, p)$ où $0 \leq p \leq 1$. En prenant le processus

$$S_t := 1 + \sum_{i=1}^t (\xi_i - 1) \quad (19)$$

et en fixant un entier $H > 0$, on définit

$$\gamma := \min\{t \geq 1 : S_t \geq H \vee S_t = 0\}. \quad (20)$$

Or on a pour un $X \sim \text{Bin}(n, p)$: $(S_\gamma - H) \mathbf{1}_{\{S_\gamma \geq H\}} \stackrel{\text{st}}{\leq} X$.

PREUVE. Maintenant on conditionne avec l'événement

$$\{\gamma = l\} \cap \{S_{l-1} = H - r\} \cap \{S_\gamma \geq H\} = \{\xi_l \geq r + 1\} \cap \{S_{l-1} = H - r\}.$$

Ainsi pour $H > 0$ fixé on a :

$$\begin{aligned} & \mathbf{P}(S_\gamma - H > c | \gamma = l, S_{l-1} = H - r, S_\gamma \geq H) \\ &= \mathbf{P}(S_l - H > c | \xi_l \geq r + 1, S_{l-1} = H - r) \\ &= \mathbf{P}(S_{l-1} + \xi_l - 1 - H > c | \xi_l \geq r + 1, S_{l-1} = H - r) \\ &= \mathbf{P}(H - r + \xi_l - 1 - H > c | \xi_l \geq r + 1) && \text{par théorème 21} \\ &= \mathbf{P}(\xi_l - (r + 1) > c | \xi_l \geq r + 1) \\ &\stackrel{\text{st}}{\leq} \mathbf{P}(X > c) && \text{par la proposition 28} \end{aligned}$$

Finalement on applique le lemme précédent 29. \square

Corollaire 31 *Soit $X \sim \text{Bin}(n, p)$ et f une fonction croissante, borelienne sur \mathbb{R} . En prenant la notation du théorème 30 on a :*

$$\mathbf{E}[f(S_\gamma - H) | S_\gamma \geq H] \leq \mathbf{E}f(X).$$

PREUVE. On combine le resultat du théorème 30 avec le lemme 18 pour la fonction f . \square

Références

- [1] N. Alon, H. J. Spencer, and P. Erdős. *The Probabilistic Method*. John Wiley & Sons, 1992.
- [2] C. Berge. *Graphes et hypergraphes*. Dunod, Paris, 1970.
- [3] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5 :17–61, 1960.
- [4] V. Girardin and N. Limnios. *Probabilités*. Librairie Vuibert, 2001.
- [5] R. M. Karp. The transitive closure of a random digraph. *Random Structures Algorithms*, 1 :73–93, 1990.
- [6] A. Martin-Löf. Symmetric sampling procedures, general epidemic processes and their threshold limit theorems. *J. Appl. Probab.*, 23 :265–282, 1986.
- [7] A. Nachmias and Y. Peres. The critical random graph, with martingales. *arXiv*, 2006.