Problème

Le théorème double-exponentiel d'Erdös-Renyi

Notations

Pour tout entier $n \geq 2$, $\Gamma(n)$ désigne l'ensemble des graphes non orientés et sans boucle dont l'ensemble des sommets est $\{1, \ldots, n\}$. L'ensemble des arêtes d'un élément G de $\Gamma(n)$, noté $A(G)$, est donc un sous-ensemble de l'ensemble $\mathcal{P}_2(n)$ des parties de cardinal 2 de $\{1, \ldots, n\}$.

Si $n \geq 2$ est un entier et p un élément de $[0,1]$, $G_{n,p}$ désigne une variable aléatoire à valeurs dans $\Gamma(n)$ qui vérifie la propriété suivante.

Pour $1 \leq i < j \leq n$, $X_{i,j}$ désignant la variable aléatoire indicatrice de l'événement « $\{i,j\}$ est une arête de $G_{n,p}$ », la famille $(X_{i,j})_{1 \leq i < j \leq n}$ est i.i.d. et chaque $X_{i,j}$ suit la loi de Bernoulli de paramètre p.

On considère une suite $(p_n)_{n \geq 2}$ d'éléments de $[0,1]$ et on pose, pour tout entier $n \geq 2$:

$$G_n = G_{n,p_n}.$$

Si $n \geq 2$ est un entier et p un élément de $[0,1]$, on note $I_{n,p}$ (resp. I_n) la variable aléatoire donnant le nombre de sommets isolés de $G_{n,p}$ (resp. G_n).

On note K_n (resp. N_n) le graphe complet (resp. le graphe vide) sur $\{1, \ldots, n\}$, défini par $A(K_n) = \mathcal{P}_2(n)$ (resp. $A(N_n) = \emptyset$).

Objectifs du problème

On se propose d'étudier le nombre de sommets isolés et la connexité de $G_{n,p}$, de manière à mettre en évidence un comportement limite poissonien et un phénomène de seuil, découverts en 1959 par Erdös et Renyi.

I. Généralités

1. Soit $n \geq 2$ un entier. Dénombrer $\Gamma(n)$.

2. Soient $n \geq 2$ un entier, $p \in (0,1]$. Exprimer les probabilités $P(G_{n,p} = K_n)$ et $P(G_{n,p} = N_n)$ en fonction de n et p.

3. Soit $\mathcal{P}$ une partie de $\Gamma(n)$. Montrer que l'application

$$\varphi_{\mathcal{P}} \ : p \in [0,1] \longmapsto P(G_{n,p} \in \mathcal{P})$$

est polynomiale.

II. Sommets isolés : première étude

4. Soient $n \geq 2$ un entier, $p \in [0,1]$. Déterminer $E(I_{n,p})$ et $E({I_{n,p}}^2)$.

5. Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. Montrer les inégalités

$$P(X \geq 1) \leq E(X) \quad ; \quad P(X \geq 1)\ E(X^2) \geq E(X)^2.$$

6. On suppose que

$$p_n = \frac{\ln(n)}{n} + \frac{a_n}{n}$$

où la suite $(a_n)_{n\geq 2}$ tend vers $+\infty$. Montrer que

$$P(I_n \geq 1) \longrightarrow 0.$$

7. On suppose que

$$p_n = \frac{\ln(n)}{n} + \frac{a_n}{n}$$

où la suite $(a_n)_{n\geq 2}$ tend vers $-\infty$. Montrer que :

$$P(I_n \geq 1) \longrightarrow 1.$$

Ce résultat sera précisé dans la question 20 (partie **VII***).*

III. Monotonie

Soit $n \geq 2$ un entier. Une partie $\mathcal{P}$ de $\Gamma(n)$ est dite *croissante* si, pour tout couple (G, G') d'éléments de $\Gamma(n)$, on a

$$A(G) \subset A(G'),\ G \in \mathcal{P} \Longrightarrow G' \in \mathcal{P}.$$

8. Soit $\mathcal{C}_n$ (resp. $\mathcal{I}_n$) l'ensemble des parties connexes (resp. sans point isolé) de $\Gamma(n)$. Montrer que $\mathcal{C}_n$ et $\mathcal{I}_n$ sont des parties croissantes de $\Gamma(n)$.

9. Soit $\mathcal{P}$ une partie croissante de $\Gamma(n)$.

a) Soient p et p' dans $[0,1]$, avec $p \leq p'$. Montrer qu'il existe deux variables de Bernoulli Y et Y', de paramètres respectifs p et p', définies sur le même univers et telles que $P(Y' \geq Y) = 1$.

b) Montrer que l'application $\varphi_{\mathcal{P}}$ est croissante sur $[0,1]$.

c) On suppose que $\mathcal{P}$ contient le graphe complet K_n et ne contient pas le graphe vide N_n. Montrer que l'application $\varphi_{\mathcal{P}}$ est strictement croissante sur $[0,1]$.

IV. Convergence en loi pour des variables discrètes

Soient I un ensemble dénombrable, $\mathcal{M}(I)$ l'ensemble des mesures de probabilités sur la tribu pleine $\mathcal{P}(I)$. Pour μ dans $\mathcal{M}(I)$, on pose

$$\forall i \in I, \qquad \mu_i = \mu(\{i\}).$$

On définit ainsi une bijection de $\mathcal{M}(I)$ sur l'ensemble des éléments de $\ell^1(I)$ à valeurs dans $\mathbb{R}^+$ et de somme 1.

On dit qu'une partie $\mathcal{K}$ de $\mathcal{M}(I)$ est *tendue* si, pour tout $\varepsilon > 0$, il existe une partie finie F de I telle que

$$\forall \mu \in \mathcal{K}, \qquad \mu(I \setminus F) \leq \varepsilon.$$

10. Soient μ et ν deux éléments de $\mathcal{M}(I)$. La *distance en variation totale* de μ et ν est
$$d(\mu, \nu) = \frac{1}{2} \sum_{i \in I} |\mu_i - \nu_i|.$$

a) Montrer que

$$d(\mu, \nu) = \max\{\mu(A) - \nu(A) \ ; \ A \in \mathcal{P}(I)\}.$$

b) Vérifier que l'application d est à valeurs dans $[0, 1]$ et définit une distance sur l'ensemble $\mathcal{M}(I)$. Caractériser l'égalité $d(\mu, \nu) = 1$.

11. Montrer que, si $(\mu_n)_{n \in \mathbb{N}}$ est une suite d'éléments de $\mathcal{M}(I)$ et μ un élément de $\mathcal{M}(I)$, on a

$$\forall i \in I, \quad \mu_n(\{i\}) \underset{n \to +\infty}{\longrightarrow} \mu(\{i\}) \iff d(\mu_n, \mu) \underset{n \to +\infty}{\longrightarrow} 0.$$

Si tel est le cas, on dira que $(\mu_n)_{n \in \mathbb{N}}$ *converge vers* μ.

Pour établir l'implication $\Rightarrow$, on montrera que la condition du membre de gauche de l'équivalence implique que $(\mu_n)_{n \in \mathbb{N}}$ est tendue.

12. Montrer qu'une partie $\mathcal{K}$ de $\mathcal{M}(I)$ est relativement compacte pour d si et seulement si elle est tendue.

V. Variables aléatoires à valeurs dans $\mathbb{N}$

Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. On rappelle que la fonction génératrice G_X de X est définie par

$$\forall t \in [-1, 1], \qquad G_X(t) = E(t^X) = \sum_{i=0}^{+\infty} P(X = i)\, t^i.$$

13. Montrer que G_X est de classe C^∞ sur $[-1,1]$ si et seulement si X admet des moments de tous ordres. Si tel est le cas, montrer que

$$\forall k \in \mathbb{N}, \qquad G^{(k)}(1) = E\left(\prod_{i=0}^{k-1}(X-i)\right).$$

14. On dit que X vérifie (U) si X admet des moments de tous ordres et si, pour toute variable aléatoire Y à valeurs dans $\mathbb{N}$, on a

$$\forall k \in \mathbb{N}, \qquad E(X^k) = E(Y^k) \Longrightarrow X \sim Y.$$

Montrer que, si le rayon de convergence de la série entière définissant G_X est strictement supérieur à 1, X vérifie (U). [1]

On pourra utiliser le fait suivant. Si f est la somme d'une série entière de rayon de convergence $R > 0$, alors

$$\forall x \in]-R, R[, \quad \forall h \in]|x|-R, R-|x|[, \qquad f(x+h) = \sum_{i=0}^{+\infty} \frac{f^{(i)}(x)}{i!} h^i.$$

VI. Suites de variables aléatoires à valeurs dans $\mathbb{N}$

Soient $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires à valeurs dans $\mathbb{N}$, X une variable aléatoire à valeurs dans $\mathbb{N}$. On dit que $(X_n)_{n\in\mathbb{N}}$ converge en loi vers X si $(P_{X_n})_{n\in\mathbb{N}}$ converge vers P_X au sens de la partie **III**, i.e. si

$$\forall i \in \mathbb{N}, \qquad P(X_n = i) \longrightarrow P(X = i).$$

15. On suppose que, pour tout $n \in \mathbb{N}$, X_n est d'espérance finie et que la suite $(E(X_n))_{n\in\mathbb{N}}$ est bornée. Montrer que la suite $(P_{X_n})_{n\in\mathbb{N}}$ est tendue.

16. Soit $k \in \mathbb{N}$. On suppose que, pour tout n, X_n^k est d'espérance finie, que la suite $\left(E(X_n^{k+1})\right)_{n\in\mathbb{N}}$ est bornée et que $(X_n)_{n\in\mathbb{N}}$ converge en loi vers X. Montrer que
$$E(X_n^k) \underset{n\to+\infty}{\longrightarrow} E(X^k).$$

17. On suppose que X vérifie (U). On suppose également que, pour tout $n \in \mathbb{N}$, X_n admet des moments de tous ordres et que

$$\forall k \in \mathbb{N}, \qquad E({X_n}^k) \underset{n\to+\infty}{\longrightarrow} E(X^k).$$

Montrer que $(X_n)_{n\in\mathbb{N}}$ converge en loi vers X. [2]

---

1. Avec un peu de travail, on peut construire des exemples de variables aléatoires à valeurs dans $\mathbb{N}$ ne vérifiant pas (U).
2. Ce résultat est le fondement de la ≪ méthode des moments ≫ en probabilités. Il s'étend aux variables aléatoires réelles. La méthode est la même, en remplaçant la fonction génératrice par la fonction caractéristique.

VII. Sommets isolés : comportement poissonien

18. a) Soient $\lambda \in \mathbb{R}^+$, Y une variable aléatoire suivant la loi de Poisson de paramètre λ. Calculer G_Y et montrer que Y vérifie (U).

    b) Soit $\lambda \in \mathbb{R}^+$. Que dire d'une suite $(X_n)_{n\geq 0}$ de variables aléatoire à valeurs dans $\mathbb{N}$ telle que [3]

$$\forall k \in \mathbb{N}, \qquad E\left(\binom{X_n}{k}\right) \longrightarrow \frac{\lambda^k}{k!} \text{ ?}$$

19. a) Soient $\ell \in \mathbb{N}^*$, $X_1, \dots, X_\ell$ des variables aléatoires de Bernoulli définies sur un même espace probabilisé. On pose

$$X = \sum_{i=1}^{\ell} X_i.$$

    Pour $k \in \mathbb{N}$, montrer que

$$\binom{X}{k} = \sum_{I \in \mathcal{P}_k(\{1,\dots,\ell\})} \prod_{i \in I} X_i.$$

    b) Soient $n \geq 2$ un entier, $k \in \mathbb{N}$, $p \in [0,1]$. Montrer que

$$E\left(\binom{I_{n,p}}{k}\right) = \binom{n}{k} (1-p)^{nk - \frac{k(k+1)}{2}}.$$

20. On suppose que c est un nombre réel, que

$$p_n = \frac{\ln(n)}{n} + \frac{c}{n} + o\left(\frac{1}{n}\right).$$

    Montrer que $(I_n)_{n\geq 2}$ converge en loi vers une variable aléatoire suivant la loi de Poisson de paramètre e^{-c}.

VIII. Arbres et formule de Cayley

Un graphe (V, E) est un *arbre* s'il est connexe et ne contient pas de cycle. Si $n \geq 2$ est un entier, on note $\mathcal{A}_n$ l'ensemble des arbres dont l'ensemble des sommets est $\{1, \dots, n\}$.

21. a) Calculer $|\mathcal{A}_n|$ pour $n = 2$, $n = 3$, $n = 4$.

    b) Soient $n \geq 2$ un entier, $A \in \mathcal{A}_n$. Pour $1 \leq i \leq n$, on note d_i le degré de i. Calculer

$$\sum_{i=1}^{n} d_i.$$

3. Ce cas particulier de la « méthode des moments » peut être établi de façon plus directe et élémentaire que ce qui précède, via une méthode de crible.

22. Soient $n \geq 2$ un entier, $d_1, \ldots, d_n$ des éléments de $\mathbb{N}^*$ dont la somme vaut $2n-2$. Montrer que le nombre de $A \in \mathcal{A}_n$ dont l'ensemble des sommets est $\{1, \ldots, n\}$ et tels que, pour tout $i \in \{1, \ldots, n\}$, le degré du sommet i soit d_i est

$$\binom{n-2}{d_1 - 1, \ldots, d_n - 1} = \frac{(n-2)!}{\prod_{i=1}^{n}(d_i - 1)!}.$$

On raisonnera par récurrence. Le résultat étant établi à l'ordre $n-1$, on adoptera les notations ci-dessus. On montrera d'abord que l'un des d_i vaut 1. Quitte à renuméroter les sommets, on pourra supposer $d_n = 1$. On remarquera alors que le graphe obtenu en ôtant à A le sommet n et l'unique arête qui le contient est un élément de $\mathcal{A}_{n-1}$.

23. Montrer, si $n \geq 2$ est un entier, la formule [4]

$$|\mathcal{A}_n| = n^{n-2}.$$

On utilisera la formule du multinôme.

IX. Seuil de connexité pour $G_{n,p}$

Pour $n \in \mathbb{N}^*$ tel que $n \geq 4$, $p \in [0,1]$ et r entier tel que $2 \leq r \leq \dfrac{n}{2}$, soit $C_{n,p,r}$ la variable aléatoire donnant le nombre de composantes connexes de cardinal r de $G_{n,p}$.

24. Soient $n \in \mathbb{N}^*$ tel que $n \geq 4$, $p \in [0,1]$. Montrer que

$$P(C_{n,p,2} \geq 1) \leq P(G_{n,p} \text{ connexe}) \leq \sum_{r=2}^{\lfloor n/2 \rfloor} P(C_{n,p,r} \geq 1).$$

25. a) Soient $n \in \mathbb{N}^*$, $p \in [0,1]$. Calculer $E(C_{n,p,2})$.

b) Soient $n \geq 6$ un entier, $p \in [0,1]$, r un entier tel que $3 \leq r \leq \dfrac{n}{2}$. Montrer que

$$E(C_{n,p,r}) \leq r^{r-2} \binom{n}{r} p^{r-1} (1-p)^{r(n-r)}.$$

26. On suppose que c est un nombre réel, que

$$p_n = \frac{\ln(n)}{n} + \frac{c}{n} + o\left(\frac{1}{n}\right).$$

Montrer que

$$P(G_n \text{ connexe}) \longrightarrow e^{-e^{-c}}.$$

4. Établie par Cayley en 1889.