Distance entre deux distributions de probabilité (Mines PSI 2023)

Durée : libre

1 – Nombre de points fixes d'une permutation

Soit n un entier naturel non nul. On note $\mathcal{S}_n$ l'ensemble des permutations de l'intervalle entier $[\![1,n]\!] = \{1,2,\dots,n\}$, c'est-à-dire des bijections de $[\![1,n]\!]$ vers lui-même. Si $\sigma \in \mathcal{S}_n$ est une permutation, on appelle *point fixe* de σ tout entier $i \in [\![1,n]\!]$ tel que : $\sigma(i) = i$.

Une permutation $\sigma \in \mathcal{S}_n$ est appelée un **dérangement** si elle n'a aucun point fixe. Pour tout $n \geqslant 1$, on note d_n le nombre de dérangements de l'intervalle entier $[\![1,n]\!]$. Par convention, on pose $d_0 = 1$.

On munit l'ensemble fini $\mathcal{S}_n$ de la probabilité uniforme notée $\mathbb{P}_n$. Sur l'espace probabilisé fini $(\mathcal{S}_n, \mathbb{P}_n)$, on définit la variable aléatoire X_n telle que, pour tout $\sigma \in \mathcal{S}_n$, $\mathrm{X}_n(\sigma)$ est le nombre de points fixes de la permutation σ.

On introduit enfin la série entière $\displaystyle\sum_{n\geqslant 1} \frac{d_n}{n!} x^n$, dont le rayon de convergence est noté R, et dont la somme sur l'intervalle de convergence $]-\mathrm{R},\mathrm{R}[$ est notée s :

$$\forall x \in]-\mathrm{R},\mathrm{R}[, \quad s(x) = \sum_{n=0}^{+\infty} \frac{d_n}{n!} x^n.$$

1 ▷ Rappeler le cardinal de $\mathcal{S}_n$. En déduire que $\mathrm{R} \geqslant 1$.

2 ▷ Pour tout $k \in [\![0,n]\!]$, montrer que le nombre de permutations de $[\![1,n]\!]$ ayant exactement k points fixes est $\binom{n}{k} d_{n-k}$. En déduire que $\mathbb{P}_n(\mathrm{X}_n = k) = \dfrac{d_{n-k}}{k!(n-k)!}$.

3 ▷ Montrer que :

$$\forall x \in]-1,1[, \quad s(x)\mathrm{e}^x = \frac{1}{1-x}.$$

En déduire que $\mathrm{R} = 1$.

4 ▷ En partant de la relation $(1-x)s(x) = \mathrm{e}^{-x}$ pour tout $x \in]-1,1[$, exprimer $\dfrac{d_n}{n!}$ pour n entier naturel, sous la forme d'une somme.

5 ▷ Montrer que la loi de la variable aléatoire X_n est donnée par

$$\forall k \in [\![0,n]\!], \quad \mathbb{P}_n(\mathrm{X}_n = k) = \frac{1}{k!} \sum_{i=0}^{n-k} \frac{(-1)^i}{i!}.$$

6 ▷ Sur l'espace probabilisé fini $(\mathcal{S}_n, \mathbb{P}_n)$, on définit, pour tout $i \in [\![1,n]\!]$, la variable aléatoire U_i telle que, pour tout $\sigma \in \mathcal{S}_n$, on ait $\mathrm{U}_i(\sigma) = 1$ si $\sigma(i) = i$, et $\mathrm{U}_i(\sigma) = 0$ sinon.

Montrer que U_i suit une loi de Bernoulli de paramètre $\dfrac{1}{n}$.

Montrer que, si $i \neq j$, la variable $\mathrm{U}_i\mathrm{U}_j$ suit une loi de Bernoulli dont on précisera le paramètre.

7 ▷ Exprimer X_n à l'aide des U_i pour $i \in [\![1,n]\!]$. En déduire l'espérance $\mathbb{E}(\mathrm{X}_n)$ et la variance $\mathbb{V}(\mathrm{X}_n)$.

8 ▷ Dans cette question, on fixe un entier naturel k. Déterminer

$$y_k = \lim_{n\to+\infty} \mathbb{P}_n(\mathrm{X}_n = k).$$

Soit Y une variable aléatoire sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, à valeurs dans $\mathbb{N}$, et vérifiant

$$\forall k \in \mathbb{N}, \quad \mathbb{P}(\mathrm{Y} = k) = y_k.$$

Reconnaître la loi de Y.

9 ▷ On note $\mathrm{G}_{\mathrm{X}_n}$ et G_Y les fonctions génératrices respectives des variables X_n et Y de la question précédente. Exprimer $\mathrm{G}_{\mathrm{X}_n}(s)$ sous forme de somme, pour s réel, et vérifier que

$$\forall s \in \mathbb{R}, \quad \lim_{n\to+\infty} \mathrm{G}_{\mathrm{X}_n}(s) = \mathrm{G}_\mathrm{Y}(s).$$

2 – Convergence en variation totale

Dans la suite du problème, on appelle *distribution (de probabilités) sur* $\mathbb{N}$ toute application $x : \mathbb{N} \to \mathbb{R}_+$ telle que

$$\sum_{k=0}^{+\infty} x(k) = 1.$$

On note $\mathcal{D}_{\mathbb{N}}$ l'ensemble des distributions de probabilités sur $\mathbb{N}$.
Si x et y sont deux distributions sur $\mathbb{N}$, on définit la *distance en variation totale* entre x et y par

$$\mathrm{d}_{\mathrm{VT}}(x,y) = \frac{1}{2}\sum_{k=0}^{+\infty} |x(k) - y(k)|.$$

10 ▷ Soient x, y, z trois distributions sur $\mathbb{N}$. Prouver les propriétés :

$$0 \leqslant \mathrm{d}_{\mathrm{VT}}(x,y) \leqslant 1\,;$$
$$\mathrm{d}_{\mathrm{VT}}(x,y) = 0 \iff x = y\,;$$
$$\mathrm{d}_{\mathrm{VT}}(y,x) = \mathrm{d}_{\mathrm{VT}}(x,y)\,;$$
$$\mathrm{d}_{\mathrm{VT}}(x,z) \leqslant \mathrm{d}_{\mathrm{VT}}(x,y) + \mathrm{d}_{\mathrm{VT}}(y,z).$$

Si X est une variable aléatoire à valeurs dans $\mathbb{N}$, définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, on note p_{X} la distribution de probabilités de X. Ainsi, p_{X} est l'application de X dans $\mathbb{R}_+$ définie par

$$\forall k \in \mathbb{N}, \quad p_{\mathrm{X}}(k) = \mathbb{P}(\mathrm{X} = k).$$

Il est clair que $p_{\mathrm{X}} \in \mathcal{D}_{\mathbb{N}}$.
En particulier, si λ est un réel strictement positif, on appelle *distribution de Poisson de paramètre* λ l'application $\pi_\lambda : \mathbb{N} \to \mathbb{R}_+$ telle que

$$\forall k \in \mathbb{N}, \quad \pi_\lambda(k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

11 ▷ Soient X et Y deux variables de Bernoulli, ayant respectivement pour paramètres $\lambda \in]0,1[$ et $\mu \in]0,1[$. Calculer $\mathrm{d}_{\mathrm{VT}}(p_{\mathrm{X}}, p_{\mathrm{Y}})$.

12 ▷ Soit X une variable de Bernoulli de paramètre $\lambda \in]0,1[$. Montrer que

$$\mathrm{d}_{\mathrm{VT}}(p_{\mathrm{X}}, \pi_\lambda) = \lambda(1 - e^{-\lambda}).$$

En déduire que

$$\mathrm{d}_{\mathrm{VT}}(p_{\mathrm{X}}, \pi_\lambda) \leqslant \lambda^2.$$

On considère de nouveau les variables aléatoires X_n introduites dans la partie I. Les questions 8 et 9 semblent montrer une certaine « convergence » des lois des variables X_n vers la loi de Poisson de paramètre 1. Le but de la fin de cette partie est de montrer que

$$\mathrm{d}_{\mathrm{VT}}(p_{\mathrm{X}_n}, \pi_1) \underset{n\to+\infty}{\longrightarrow} 0,$$

et que cette convergence est assez rapide.

13 ▷ Vérifier la relation, pour tout n entier naturel non nul :

$$2\,\mathrm{d}_{\mathrm{VT}}(p_{\mathrm{X}_n}, \pi_1) = \sum_{k=0}^{n} \frac{1}{k!}\left|\sum_{i=n-k+1}^{+\infty} \frac{(-1)^i}{i!}\right| + \mathrm{e}^{-1}\sum_{k=n+1}^{+\infty}\frac{1}{k!}.$$

14 ▷ Pour tout n entier naturel, on pose $r_n = \displaystyle\sum_{k=n+1}^{+\infty}\frac{1}{k!}$. Prouver la majoration

$$\forall n \in \mathbb{N}, \quad r_n \leqslant \frac{1}{(n+1)!}\sum_{k=0}^{+\infty}\frac{1}{(n+2)^k}.$$

En déduire un équivalent simple de r_n lorsque n tend vers $+\infty$.

15 $\triangleright$ En continuant de majorer le second membre de l'égalité de la question 13, établir l'estimation

$$\mathrm{d}_{\mathrm{VT}}(p_{\mathrm{X}_n}, \pi_1) \underset{n\to+\infty}{=} \mathrm{O}\left(\frac{2^n}{(n+1)!}\right).$$

On pourra faire intervenir les coefficients binomiaux.

3 – Autres estimations de distances en variation totale

Si x et y sont deux distributions de probabilités sur $\mathbb{N}$, on définit l'application $x * y : \mathbb{N} \to \mathbb{R}_+$ par

$$\forall k \in \mathbb{N}, \quad (x*y)(k) = \sum_{i=0}^{k} x(i)y(k-i) = \sum_{i+j=k} x(i)y(j).$$

16 $\triangleright$ Montrer que $x * y$ est une distribution sur $\mathbb{N}$.

17 $\triangleright$ Soient X et Y des variables aléatoires indépendantes, à valeurs dans $\mathbb{N}$, définies sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. Prouver la relation

$$p_{\mathrm{X+Y}} = p_{\mathrm{X}} * p_{\mathrm{Y}}.$$

18 $\triangleright$ Soit $(x, y, u, v) \in (\mathcal{D}_{\mathbb{N}})^4$. Montrer que, pour tout k entier naturel,

$$\left|(x*y)(k) - (u*v)(k)\right| \leqslant \sum_{i+j=k} y(j)\left|x(i) - u(i)\right| + \sum_{i+j=k} u(i)\left|y(j) - v(j)\right|.$$

19 $\triangleright$ Avec les notations de la question précédente, établir l'inégalité

$$\mathrm{d}_{\mathrm{VT}}(x*y, u*v) \leqslant \mathrm{d}_{\mathrm{VT}}(x,u) + \mathrm{d}_{\mathrm{VT}}(y,v).$$

20 $\triangleright$ Soit U une variable binomiale de paramètres $n \in \mathbb{N}^*$ et $\lambda \in]0,1[$. Prouver l'inégalité

$$\mathrm{d}_{\mathrm{VT}}(p_{\mathrm{U}}, \pi_{n\lambda}) \leqslant n\lambda^2.$$

21 $\triangleright$ Soit α un réel strictement positif. Pour tout entier naturel n tel que $n > \lfloor \alpha \rfloor$, on note B_n une variable binomiale de paramètres n et $\dfrac{\alpha}{n}$. Pour tout k entier naturel, déterminer

$$\lim_{n\to+\infty} \mathbb{P}(\mathrm{B}_n = k).$$

On pourra utiliser la question précédente.

22 $\triangleright$ Soient α et β deux réels strictement positifs. En utilisant les résultats et les méthodes qui précèdent, montrer que

$$\mathrm{d}_{\mathrm{VT}}(\pi_\alpha, \pi_\beta) \leqslant |\beta - \alpha|.$$