
Distance entre deux distributions de probabilités sur $\mathbb{N}$

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 $\llbracket 1, n \rrbracket = \{1, 2, \dots, n\}$, c'est-à-dire des bijections de $\llbracket 1, n \rrbracket$ vers lui-même. Si $\sigma \in \mathcal{S}_n$ est une permutation, on appelle **point fixe** de σ tout entier $i \in \llbracket 1, n \rrbracket$ 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 $n \geq 1$, on note d_n le nombre de dérangements de l'intervalle entier $\llbracket 1, n \rrbracket$. Par convention, on pose $d_0 = 1$.

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

2 ▷ Pour $k \in \llbracket 0, n \rrbracket$, montrer que le nombre de permutations de $\llbracket 1, n \rrbracket$ ayant exactement k points fixes est $\binom{n}{k} d_{n-k}$.

En déduire que $P_n(X_n = k) = \frac{d_{n-k}}{k!(n-k)!}$.

- 6** ▷ Sur l'espace probabilisé fini $(\mathcal{S}_n, P_n)$, on définit, pour tout $i \in \llbracket 1, n \rrbracket$, la variable aléatoire U_i telle que, pour tout $\sigma \in \mathcal{S}_n$, on ait $U_i(\sigma) = 1$ si $\sigma(i) = i$, et $U_i(\sigma) = 0$ sinon.

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

Montrer que, si $i \neq j$, la variable $U_i U_j$ suit une loi de Bernoulli dont on précisera le paramètre.

- 7** ▷ Exprimer X_n à l'aide des U_i , $1 \leq i \leq n$. En déduire l'espérance $E(X_n)$ et la variance $V(X_n)$.

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

$$y_k = \lim_{n \rightarrow +\infty} P_n(X_n = k) .$$

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

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

Reconnaître la loi de Y .

2 Convergence en variation totale

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

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

On note $\mathcal{D}_{\mathbf{N}}$ l'ensemble des distributions de probabilités sur $\mathbf{N}$.

Si x et y sont deux distributions sur $\mathbf{N}$, on définit la **distance en variation totale** entre x et y par

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

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

$$0 \leq d_{VT}(x, y) \leq 1 ;$$

$$d_{VT}(x, y) = 0 \iff x = y ;$$

$$d_{VT}(y, x) = d_{VT}(x, y) ;$$

$$d_{VT}(x, z) \leq d_{VT}(x, y) + d_{VT}(y, z) .$$

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

$$\forall k \in \mathbf{N} \quad p_X(k) = P(X = k) .$$

Il est clair que $p_X \in \mathcal{D}_{\mathbf{N}}$.

En particulier, si λ est un réel strictement positif, on appelle **distribution de Poisson de paramètre λ** l'application $\pi_\lambda: \mathbf{N} \rightarrow \mathbf{R}_+$ telle que

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

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

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

$$d_{VT}(p_X, \pi_\lambda) = \lambda (1 - e^{-\lambda}) .$$

En déduire que

$$d_{VT}(p_X, \pi_\lambda) \leq \lambda^2 .$$

On considère de nouveau les variables aléatoires X_n introduites dans la partie 1. 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

$$d_{VT}(p_{X_n}, \pi_1) \xrightarrow{n \rightarrow +\infty} 0,$$

et que cette convergence est assez rapide.

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

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

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

$$r_n \leq \frac{1}{(n+1)!} \sum_{k=0}^{+\infty} \frac{1}{(n+2)^k} .$$

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

- 15** ▷ En continuant de majorer le second membre de l'égalité de la question **13.**, établir l'estimation

$$d_{VT}(p_{X_n}, \pi_1) \underset{n \rightarrow +\infty}{=} O\left(\frac{2^n}{(n+1)!}\right).$$

On pourra faire intervenir des coefficients binomiaux.