

On traitera **ou bien** le problème 1 et l'un des deux exercices au choix, **ou bien** le problème 2. Les panachages des deux sujets ne seront pas corrigés.

Sujet 1 — Problème CCMP 2023, PSI (3h)

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 $\mathfrak{S}_n$ l'ensemble des permutations de l'intervalle entier $\llbracket 1, n \rrbracket$, c'est-à-dire des bijections de $\llbracket 1, n \rrbracket$ sur lui-même. Si $\sigma \in \mathfrak{S}_n$ est une permutation, on appelle *point fixe* de σ tout entier $i \in \llbracket 1, n \rrbracket$ tel que $\sigma(i) = i$.

Une permutation $\sigma \in \mathfrak{S}_n$ est appelée *dérangement* si elle n'a aucun point fixe. Pour tout $n \geq 1$, on note d_n le nombre de dérangements de $\mathfrak{S}_n$. Par convention, on pose $d_0 = 1$.

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

On introduit enfin la série entière $\sum_{n \geq 0} \frac{d_n}{n!} x^n$, dont le rayon de convergence est noté R , et dont la somme sur l'intervalle de convergence $] -R, R[$ est notée $s(x)$.

Q 1. Rappeler sans justification le cardinal de $\mathfrak{S}_n$. Qu'en déduit-on quant à R ?

Q 2. Pour $k \in \llbracket 0, n \rrbracket$, montrer que le nombre d'éléments de $\mathfrak{S}_n$ ayant exactement k points fixes est $\binom{n}{k} d_{n-k}$. En déduire que $\mathbf{P}_n(X_n = k) = \frac{d_{n-k}}{k! (n-k)!}$.

Q 3. Montrer que $s(x) e^x = \frac{1}{1-x}$ pour tout $x \in]-1, 1[$. En déduire que $R = 1$.

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

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

$$\forall k \in \llbracket 0, n \rrbracket : \mathbf{P}_n(X_n = k) = \frac{1}{k!} \sum_{i=0}^{n-k} \frac{(-1)^i}{i!}.$$

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

Montrer que $U_i \sim \mathcal{B}(n^{-1})$. Montrer que, si $i \neq j$, la variable $U_i U_j$ suit une loi de Bernoulli dont on précisera le paramètre.

Q 7. Exprimer X_n à l'aide des U_i , $1 \leq i \leq n$. En déduire son espérance $\mathbf{E}(X_n)$ et sa variance $\mathbf{V}(X_n)$.

Q 8. Dans cette question, on fixe un entier naturel k . Déterminer $y_k = \lim_{n \rightarrow +\infty} \mathbf{P}_n(X_n = k)$.

Soit Y une variable aléatoire sur un espace probabilisé à valeurs dans $\mathbb{N}$, et vérifiant $\mathbf{P}(Y = k) = y_k$ pour tout $n \in \mathbb{N}$. Reconnaître la loi de Y .

Q 9. On note G_{X_n} et G_Y les fonctions génératrices respectives des variables X_n et Y de la question précédente. Exprimer $G_{X_n}(s)$ sous forme de somme, pour s réel, et vérifier que

$$\forall s \in \mathbb{R}: \lim_{n \rightarrow +\infty} G_{X_n}(s) = G_Y(s).$$

2. Convergence en variation totale

Dans la suite du problème, on appelle *distribution de probabilités sur $\mathbb{N}$* toute application $x: \mathbb{N} \rightarrow \mathbb{R}_+$ 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 éléments de $\mathcal{D}_{\mathbb{N}}$, on définit la *distance en variation totale* entre x et y par

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

Q 10. Soient x, y, z trois distributions de probabilités sur $\mathbb{N}$. Prouver les propriétés ci-dessous :

- i) $0 \leq d_{VT}(x, y) \leq 1$;
- ii) $d_{VT}(x, y) = 0 \iff x = y$ (séparation) ;
- iii) $d_{VT}(y, x) = d_{VT}(x, y)$ (symétrie) ;
- iv) $d_{VT}(x, z) \leq d_{VT}(x, y) + d_{VT}(y, z)$ (inégalité triangulaire).

Si X est une variable aléatoire à valeurs dans $\mathbb{N}$, définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$, on note p_X la distribution de probabilités de X . Ainsi, p_X est l'application de $\mathbb{N}$ vers $\mathbb{R}_+$ définie par $p_X(k) = \mathbf{P}(X = k)$ pour tout $k \in \mathbb{N}$. Il est clair que $p_X \in \mathcal{D}_{\mathbb{N}}$. En particulier, si λ est un réel strictement positif, on appelle *distribution de Poisson de paramètre λ* l'application $\pi_\lambda: \mathbb{N} \rightarrow \mathbb{R}_+$ telle que

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

Q 11. Soient $X \sim \mathcal{B}(\lambda)$ et $Y \sim \mathcal{B}(\mu)$ deux variables de Bernoulli avec $(\lambda, \mu) \in]0, 1]^2$. Calculer $d_{VT}(p_X, p_Y)$.

Q 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 **Q 8** et **Q 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 $\lim_{n \rightarrow \infty} d_{VT}(p_{X_n}, \pi_1) = 0$ et que cette convergence est assez rapide.

Q 13. Vérifier la relation

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

Q 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 $+\infty$.

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

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

On pourra faire intervenir des coefficients binomiaux.

3. Autres estimations en variation totale

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

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

Q 16. Montrer que $x * y \in \mathcal{D}_{\mathbb{N}}$.

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

$$p_{X+Y} = p_X * p_Y$$

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

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

Q 19. Avec les notations de la question précédente, établir l'inégalité

$$d_{VT}(x * y, u * v) \leq d_{VT}(x, u) + d_{VT}(y, v).$$

Q 20. Soit U une variable binomiale de paramètres $n \in \mathbb{N}^*$ et $\lambda \in]0, 1[$. Montrer que $d_{VT}(p_U, \pi_{n\lambda}) \leq n\lambda^2$.

Q 21. Soit α un réel strictement positif. Pour tout entier naturel n tel que $n > \lfloor \alpha \rfloor$, on note $B_n \sim \mathcal{B}\left(n, \frac{\alpha}{n}\right)$. Pour tout k entier naturel, déterminer $\lim_{n \rightarrow +\infty} \mathbf{P}(B_n = k)$. On pourra utiliser la question précédente.

Q 22. Soit $(\alpha, \beta) \in (\mathbb{R}_+^*)^2$. En utilisant les résultats et les méthodes qui précèdent, montrer que

$$d_{VT}(\pi_\alpha, \pi_\beta) \leq |\beta - \alpha|.$$

Sujet 1 — Exercice 1 (au choix) — Oral X

Soit $\theta > 0$ un paramètre. Des individus numérotés à partir de 1 arrivent dans un restaurant abritant une infinité de tables infiniment longues. L'installation des convives se fait de la manière suivante :

- les individus arrivent un à un, le numéro 1, puis le numéro 2, etc ;
- lorsque le $k + 1$ -ème individu se présente, il choisit au hasard l'un des k convives déjà présents et s'installe à sa table avec la probabilité $\frac{1}{k + \theta}$. Il s'installe seul à une nouvelle table avec la probabilité $\frac{\theta}{k + \theta}$.

On note K_n le nombre de tables occupées après que n personnes se sont présentées et G_n sa fonction génératrice.

Q 23. Déterminer $\mathbf{P}(K_n = 1)$.

Q 24. Montrer que $G_{n+1}(x) = \frac{L_n(\theta x)}{L_n(\theta)}$, où $L_n(x) = \prod_{i=0}^n (x + i)$.

Q 25. Calculer $\mathbf{E}(K_{n+1})$ et $\mathbf{V}(K_{n+1})$. En donner des équivalents quand n tend vers l'infini.

Q 26. Étudier le comportement de $\frac{K_n}{\ln n}$ (convergence en probabilité vers 1).

Sujet 1 — Exercice 2 (au choix) — Oral CCINP

On dispose d'une urne contenant N boules, dont r boules blanches et $N - r$ boules noires. On tire les boules sans remise. Soit X la variable aléatoire donnant le nombre de tirages nécessaires pour avoir tiré toutes les boules blanches.

Q 27. Dans le cas $r = 1$, quelle loi X suit-elle ? Donner son espérance. Mêmes questions si $r = N$. On suppose désormais que $1 < r < N$.

Q 28. Déterminer $X(\Omega)$, puis montrer que pour tout $k \in X(\Omega)$, $\mathbf{P}(X = k) = \binom{k-1}{r-1} \binom{N}{r}^{-1}$.

Q 29. Après avoir exprimé le lien entre $\binom{p}{q}$ et $\binom{p-1}{q-1}$, donner une expression de $\mathbf{E}(X)$ sans factorielle.

Sujet 2 — Problème CCMP 2020, MP (4h)

Nombre de sites visités par une marche aléatoire

Dans tout le texte, d est un élément de $\mathbb{N}^*$. On note 0_d le d -uplet dont toutes les coordonnées valent 0, c'est-à-dire le vecteur nul de $\mathbb{R}^d$.

On considère une variable aléatoire X à valeurs dans $\mathbb{Z}^d$, $(X_k)_{k \in \mathbb{N}^*}$ une suite de variables aléatoires mutuellement indépendantes suivant chacune la loi de X et définies sur un même espace probabilisé. La suite de variables aléatoires $(S_n)_{n \in \mathbb{N}}$ est définie par

$$S_0 = 0_d \quad \& \quad \forall n \in \mathbb{N}^*: S_n = \sum_{k=1}^n X_k.$$

La suite $(S_n)_{n \in \mathbb{N}}$ est une *marche aléatoire de pas X* , à valeurs dans $\mathbb{Z}^d$.

On note R la variable aléatoire à valeurs dans $\mathbb{N}^* \cup \{+\infty\}$ définie par

$$R = \begin{cases} \min\{n \in \mathbb{N}^*; S_n = 0_d\} & \text{si } \{n \in \mathbb{N}^*, S_n = 0_d\} \neq \emptyset, \\ +\infty & \text{sinon.} \end{cases}$$

Autrement dit, R est égal à $+\infty$ si la marche aléatoire $(S_n)_{n \in \mathbb{N}^*}$ ne revient jamais en 0_d zr au premier instant auquel cette marche aléatoire revient en 0_d sinon.

Pour $n \in \mathbb{N}$, soit $N_n = \#\{S_k, k \in \{0, \dots, n\}\}$. Le nombre N_n est donc le nombre de points de $\mathbb{Z}^d$ visités par la marche aléatoire $(S_n)_{n \in \mathbb{N}}$ après n pas.

Le but du problème est d'étudier asymptotiquement l'espérance $\mathbf{E}(N_n)$ de la variable aléatoire N_n . La partie D est indépendante des parties précédentes.

A. PRÉLIMINAIRES

Les cinq questions de cette partie sont indépendantes et utilisées dans les parties C et E.

Q 1. Soit $n \in \mathbb{N}$. En utilisant la factorisation $(X+1)^{2n} = (X+1)^n (X+1)^n$, montrer que $\sum_{k=0}^n \binom{n}{k}^2 = \binom{2n}{n}$.

Q 2. Rappeler la formule de Stirling, puis déterminer un nombre réel $c > 0$ tel que $\binom{2n}{n} \underset{n \rightarrow +\infty}{\sim} c \frac{4^n}{\sqrt{n}}$.

Q 3. Si $\alpha \in]0, 1[$, montrer, par exemple en utilisant une comparaison série-intégrale, que

$$\sum_{k=1}^n \frac{1}{k^\alpha} \underset{n \rightarrow +\infty}{\sim} \frac{n^{1-\alpha}}{1-\alpha}.$$

Si $\alpha > 1$, montrer de même que $\sum_{k=n+1}^{+\infty} \frac{1}{k^\alpha} \underset{n \rightarrow +\infty}{\sim} \frac{1}{(\alpha-1)n^{\alpha-1}}$.

Q 4. Pour $x \in [2, +\infty[$, on pose $I(x) = \int_2^x \frac{dt}{\ln(t)}$. Justifier, pour $x \in [2, +\infty[$, la relation

$$I(x) = \frac{x}{\ln(x)} - \frac{2}{\ln(2)} + \int_2^x \frac{dt}{(\ln(t))^2}.$$

Q 5. Montrer que $\int_2^x \frac{dt}{(\ln(t))^2} \underset{x \rightarrow +\infty}{\sim} \mathfrak{o}(I(x))$. En déduire un équivalent de $I(x)$ lorsque $x \rightarrow +\infty$.

Q 6. En utilisant un développement en série entière usuel, justifier la formule

$$\forall x \in]-1, 1[: \frac{1}{\sqrt{1-x}} = \sum_{n=0}^{+\infty} \frac{1}{4^n} \binom{2n}{n} x^n.$$

B. MARCHES ALÉATOIRES, RÉCURRENCE

On considère les fonctions F et G définies par les formules

$$\forall x \in]-1, 1[: F(x) = \sum_{n=0}^{+\infty} \mathbf{P}(S_n = 0_d) x^n \quad \& \quad \forall x \in]-1, 1[: G(x) = \sum_{n=0}^{+\infty} \mathbf{P}(R = n) x^n.$$

Q 7. Montrer que les séries entières définissant F et G ont un rayon de convergence supérieur ou égal à 1. Justifier alors que les fonctions F et G sont définies et de classe $\mathcal{C}^\infty$ sur $] -1, 1[$.

Q 8. Montrer que G est définie et continue sur $[-1, 1]$ et que $G(1) = \mathbf{P}(R \neq +\infty)$.

Q 9. Si k et n sont des entiers naturels tels que $k \leq n$, montrer que

$$\mathbf{P}((S_n = 0_d) \cap (R = k)) = \mathbf{P}(R = k) \mathbf{P}(S_{n-k} = 0_d)$$

En déduire que

$$\forall n \in \mathbb{N}^*, \mathbf{P}(S_n = 0_d) = \sum_{k=1}^n \mathbf{P}(R = k) \mathbf{P}(S_{n-k} = 0_d).$$

Q 10. Montrer que $F(x) = 1 + F(x)G(x)$ pour tout $x \in]-1, 1[$.

Q 11. En discutant selon la valeur de $\mathbf{P}(R \neq +\infty)$, déterminer la limite de $F(x)$ lorsque $x \rightarrow 1^-$.

Q 12. Soit $(c_k)_{k \in \mathbb{N}}$ une suite d'éléments de $\mathbb{R}^+$ telle que la série entière $\sum c_k x^k$ ait un rayon de convergence égal à 1 et que la série $\sum c_k$ diverge. Montrer que $\lim_{x \rightarrow 1^-} \sum_{k=0}^{+\infty} c_k x^k = +\infty$.

Q 13. Montrer que la série $\sum \mathbf{P}(S_n = 0_d)$ est divergente si, et seulement si, $\mathbf{P}(R \neq +\infty) = 1$.

Q 14. Pour $i \in \mathbb{N}^*$, soit Y_i la variable de Bernoulli indicatrice de l'événement $(S_i \notin \{S_k, 0 \leq k \leq i-1\})$. Montrer que $\mathbf{P}(Y_i = 1) = \mathbf{P}(R > i)$ pour tout $i \in \mathbb{N}^*$.

Q 15. En déduire que $\mathbf{E}(N_n) = 1 + \sum_{i=1}^n \mathbf{P}(R > i)$ pour tout $n \in \mathbb{N}^*$.

Q 16. Conclure que $\lim_{n \rightarrow +\infty} \frac{\mathbf{E}(N_n)}{n} = \mathbf{P}(R = +\infty)$. On pourra admettre et utiliser le théorème de Cesàro : si

$(u_n)_{n \in \mathbb{N}^*}$ est une suite réelle convergeant vers le nombre réel ℓ , alors $\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=1}^n u_k = \ell$.

C. LES MARCHES DE BERNOULLI SUR $\mathbb{Z}$

Dans cette question, $d = 1$ et on note donc simplement $0_d = 0$. Par ailleurs, $p \in]0, 1[$, $q = 1 - p$ et la loi de X est donnée par $\mathbf{P}(X = 1) = p$ et $\mathbf{P}(X = -1) = q$.

Q 17. Pour $n \in \mathbb{N}$, déterminer $\mathbf{P}(S_{2n+1} = 0)$ et justifier l'égalité $\mathbf{P}(S_{2n} = 0) = \binom{2n}{n} (pq)^n$.

Q 18. Pour $x \in]-1, 1[$, donner une expression simple de $G(x)$.

Q 19. Exprimer $\mathbf{P}(R = +\infty)$ en fonction de $|p - q|$ et déterminer la loi de R .

Q 20. On suppose que $p = q = 1/2$. Donner un équivalent simple de $\mathbf{P}(R = 2n)$ lorsque n tend vers $+\infty$. En déduire un équivalent simple de $\mathbf{E}(N_n)$ lorsque n tend vers $+\infty$.

D. UN RÉSULTAT ASYMPTOTIQUE

Soient $(a_n)_{n \in \mathbb{N}}$ et $(b_n)_{n \in \mathbb{N}}$ deux suites d'éléments de $\mathbb{R}_+^*$. On suppose que $(a_n)_{n \in \mathbb{N}}$ est décroissante et que $\sum_{k=0}^n a_k b_{n-k} = 1$ pour tout $n \in \mathbb{N}$. Pour $n \in \mathbb{N}$, on pose $B_n = \sum_{k=0}^n b_k$.

Q 21. Soient m et n deux entiers naturels tels que $m > n$. Montrer que

$$a_n \leq \frac{1}{B_n} \quad \text{et} \quad 1 \leq a_n B_{m-n} + a_0 (B_m - B_{m-n}).$$

Q 22. On suppose dans cette question qu'il existe $(m_n)_{n \in \mathbb{N}}$ vérifiant $m_n > n$ pour n assez grand ainsi que

$$B_{m_n-n} \underset{n \rightarrow +\infty}{\sim} B_n \quad \text{et} \quad B_{m_n} - B_{m_n-n} \xrightarrow[n \rightarrow +\infty]{} 0.$$

Montrer que $a_n \underset{n \rightarrow +\infty}{\sim} \frac{1}{B_n}$.

Q 23. On suppose dans cette question qu'il existe $C > 0$ tel que $b_n \underset{n \rightarrow +\infty}{\sim} \frac{C}{n}$. En utilisant la question précédente pour une suite $(m_n)_{n \in \mathbb{N}}$ bien choisie, montrer que

$$a_n \underset{n \rightarrow +\infty}{\sim} \frac{1}{C \ln(n)}.$$

E. LA MARCHE ALÉATOIRE SIMPLE SUR $\mathbb{Z}^2$: UN THÉORÈME D'ERDŐS ET DVORESTSKY

Q 24. Soit $n \in \mathbb{N}^*$. Montrer que

$$\sum_{k=0}^n \mathbf{P}(S_k = 0_d) \mathbf{P}(R > n - k) = 1.$$

Dans les deux dernières questions, on suppose que $d = 2$ et que la loi de X est donnée par

$$\mathbf{P}(X = (0, 1)) = \mathbf{P}(X = (0, -1)) = \mathbf{P}(X = (1, 0)) = \mathbf{P}(X = (-1, 0)) = \frac{1}{4}.$$

Q 25. Soit $n \in \mathbb{N}$. Établir l'égalité

$$\mathbf{P}(S_{2n} = 0_2) = \frac{1}{4^{2n}} \binom{2n}{n}^2.$$

Q 26. Donner un équivalent simple de $\mathbf{E}(N_n)$ lorsque n tend vers $+\infty$.

Si vous avez terminé, vous pouvez traiter l'exercice d'oral de l'X du premier sujet...