Corrigé : distance entre deux distributions de probabilité (Mines PSI 2023)

1 – Nombre de points fixes d'une permutation

1 ▷ On a card $S_n = n!$ donc $0 \leqslant \dfrac{d_n}{n!} \leqslant 1$. Le rayon de convergence de $\sum x^n$ est égal à 1 donc $R \geqslant 1$.

2 ▷ Il y a $\dbinom{n}{k}$ façons de choisir k points fixes et d_{n-k} permutations sans point fixe des $n-k$ points restants, donc il y a $\dbinom{n}{k}d_{n-k}$ permutation ayant exactement k points fixes. On a donc $\mathbb{P}_n(X_n = k) = \dfrac{1}{n!}\dbinom{n}{k}d_{n-k} = \dfrac{d_{n-k}}{k!(n-k)!}$.

3 ▷ Pour tout $x \in \mathbb{R}$, $e^x = \displaystyle\sum_{n=0}^{+\infty} \frac{x^n}{n!}$ donc on peut réaliser le produit de Cauchy des deux séries entières sur $]-1,1[$:

$$\forall x \in]-1,1[,\quad s(x)e^x = \sum_{n=0}^{+\infty}\Big(\sum_{k=0}^{n} \frac{d_{n-k}}{k!(n-k)!}\Big)x^n = \sum_{n=0}^{+\infty}\Big(\sum_{k=0}^{n} \mathbb{P}_n(X_n = k)\Big)x^n = \sum_{n=0}^{+\infty} x^n = \frac{1}{1-x}$$

car $\big([X_n = k] \,\big|\, k \in [\![0,n]\!]\big)$ est un système complet d'événements.
On en déduit que $\displaystyle\lim_{x\to 1^-} s(x) = +\infty$: s n'est pas de limite finie en 1 donc $R \leqslant 1$, et en définitive $R = 1$.

4 ▷ Pour tout $x \in]-1,1[$, $(1-x)s(x) = \displaystyle\sum_{n=0}^{+\infty} \frac{d_n}{n!}x^n - \sum_{n=1}^{+\infty} \frac{d_{n-1}}{(n-1)!}x^n = 1 + \sum_{n=1}^{+\infty}\Big(\frac{d_n}{n!} - \frac{d_{n-1}}{(n-1)!}\Big)x^n$.
Mais par ailleurs $(1-x)s(x) = e^{-x} = \displaystyle\sum_{n=0}^{+\infty} \frac{(-1)^n}{n!}x^n$ donc par unicité du développement en série entière on a pour tout $n \geqslant 1$, $\dfrac{d_n}{n!} - \dfrac{d_{n-1}}{(n-1)!} = \dfrac{(-1)^n}{n!}$. Par télescopage on en déduit $\dfrac{d_n}{n!} = d_0 + \displaystyle\sum_{k=1}^{n} \frac{(-1)^k}{k!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!}$.

Remarque. On peut aussi obtenir ce résultat en écrivant $s(x) = \dfrac{e^{-x}}{1-x}$ et en réalisant de nouveau un produit de Cauchy.

5 ▷ Sachant que $\mathbb{P}_n(X_n = k) = \dfrac{d_{n-k}}{k!(n-k)!}$ (question 2) on en déduit que $\mathbb{P}_n(X_n = k) = \dfrac{1}{k!}\displaystyle\sum_{i=0}^{n-k} \frac{(-1)^i}{i!}$.

6 ▷ Pour tout $i \in [\![1,n]\!]$, il y a $(n-1)!$ permutations de S_n qui admettent i pour point fixe donc $\mathbb{P}_n(U_i = 1) = \dfrac{(n-1)!}{n!} = \dfrac{1}{n}$, ce qui montre que $U_i \sim \mathscr{B}\Big(\dfrac{1}{n}\Big)$.
Si $i \neq j$, on a $U_i(\sigma)(U_j(\sigma) = 1$ si et seulement si i et j sont deux points fixes de σ. Or il y a $(n-2)!$ permutations de S_n qui laissent fixes i et j donc $\mathbb{P}_n(U_1U_j = 1) = \dfrac{(n-2)!}{n!} = \dfrac{1}{n(n-1)}$, et ainsi, $U_iU_j \sim \mathscr{B}\Big(\dfrac{1}{n(n-1)}\Big)$.

7 ▷ On a $X_n = \displaystyle\sum_{i=1}^{n} U_i$ donc par linéarité de l'espérance, $\mathbb{E}(X_n) = \displaystyle\sum_{i=1}^{n} \mathbb{E}(U_i) = n \times \frac{1}{n} = 1$.
On a aussi $\mathbb{V}(X_n) = \displaystyle\sum_{i=1}^{n} \mathbb{V}(U_i) + 2\sum_{i<j} \mathrm{cov}(U_i, U_j)$, avec $\mathbb{V}(U_i) = \dfrac{1}{n}\Big(1 - \dfrac{1}{n}\Big) = \dfrac{n-1}{n^2}$ et $\mathrm{cov}(U_i, U_j) = \mathbb{E}(U_iU_j) - \mathbb{E}(U_i)\mathbb{E}(U_j) = \dfrac{1}{n^2(n-1)}$. Ainsi, $\mathbb{V}(X_n) = n \times \dfrac{n-1}{n^2} + n(n-1) \times \dfrac{1}{n^2(n-1)} = 1$.

8 ▷ On a $y_k = \dfrac{1}{k!}\displaystyle\sum_{i=0}^{+\infty} \frac{(-1)^i}{i!} = \frac{e^{-1}}{k!} = \frac{e^{-1}\,1^k}{k!}$ donc $Y \sim \mathscr{P}(1)$.

9 ▷ Par définition, $G_{X_n}(s)=\sum_{k=0}^{n}\mathbb{P}(X_n=k)s^k=\sum_{k=0}^{n}\sum_{i=0}^{n-k}\frac{(-1)^i}{i!}\frac{s^k}{k!}=\Big(\sum_{k=0}^{n}\frac{s^k}{k!}\Big)\Big(\sum_{i=0}^{n}\frac{(-1)^i}{i!}\Big)-\sum_{k=1}^{n}\sum_{i=n-k+1}^{n}\frac{(-1)^i}{i!}\frac{s^k}{k!}$.

On a $\lim\limits_{n\to+\infty}\Big(\sum_{k=0}^{n}\frac{s^k}{k!}\Big)\Big(\sum_{i=0}^{n}\frac{(-1)^i}{i!}\Big)=e^s\times e^{-1}=e^{s-1}=G_y(s)$ et :

$$\Big|\sum_{k=1}^{n}\sum_{i=n-k+1}^{n}\frac{(-1)^i}{i!}\frac{s^k}{k!}\Big|\leqslant\sum_{k=1}^{n}\sum_{i=n-k+1}^{n}\frac{|s|^k}{i!k!}\leqslant\sum_{k=1}^{n}\frac{k|s|^k}{(n-k+1)!k!}=\frac{|s|}{n!}\sum_{k=0}^{n-1}\binom{n}{k}|s|^k\leqslant\frac{|s|}{n!}\big(|s|+1\big)^n$$

D'après le critère de d'Alembert, $\lim\limits_{n\to+\infty}\frac{|s|}{n!}\big(|s|+1\big)^n=0$ donc $\lim\limits_{n\to+\infty}G_{X_n}(s)=G_Y(s)$.

2 – Convergence en variation totale

10 ▷ Pour tout $k\in\mathbb{N}$, $0\leqslant|x(k)-y(k)|\leqslant x(k)+y(k)$. les séries $\sum x(k)$ et $\sum y(k)$ convergent donc il en est de même de $\sum|x(k)-y(k)|$, et $0\leqslant d_{VT}(x,y)\leqslant\frac{1}{2}(1+1)=1$.
$d_{VT}(x,y)=0\iff\forall k\in\mathbb{N},\ |x(k)-y(k)|=0$, soit encore si et seulement si $x=y$.
Pour tout $k\in\mathbb{N}$, $|x(k)-y(k)|=|y(k)-x(k)|$ donc $d_{VT}(x,y)=d_{VT}(y,x)$.
Enfin, pour tout $k\in\mathbb{N}$, $|x(k)-z(k)|\leqslant|x(k)-y(k)|+|y(k)-z(k)|$ donc $d_{VT}(x,z)\leqslant d_{VT}(x,y)+d_{VT}(y,z)$.

11 ▷ Si $X\sim\mathscr{B}(\lambda)$ et $Y\sim\mathscr{B}(\mu)$ alors

$$d_{VT}(p_X,p_y)=\frac{1}{2}\sum_{k=0}^{+\infty}\Big|\mathbb{P}(X=k)-\mathbb{P}(y=k)\Big|=\Big|\mathbb{P}(X=0)-\mathbb{P}(Y=0)\Big|+\Big|\mathbb{P}(X=1)-\mathbb{P}(Y=1)\Big|=|\lambda-\mu|$$

12 ▷ Si $X\sim\mathscr{B}(\lambda)$ alors, en utilisant l'inégalité de convexité $1-\lambda\leqslant e^{-\lambda}$:

$$\begin{aligned}d_{VT}(p_X,\pi_\lambda)&=\frac{1}{2}\sum_{k=0}^{+\infty}\Big|\mathbb{P}(X=k)-\pi_\lambda(k)\Big|=\frac{1}{2}\Big(\Big|(1-\lambda)-e^{-\lambda}\Big|+\Big|\lambda-\lambda e^{-\lambda}\Big|+\sum_{k=2}^{+\infty}e^{-\lambda}\frac{\lambda^k}{k!}\Big)\\&=\frac{1}{2}\Big((e^{-\lambda}-1-\lambda)+(\lambda-\lambda e^{-\lambda})+(1-e^{-\lambda}-\lambda e^{-\lambda})\Big)=\lambda(1-e^{-\lambda})\end{aligned}$$

et en utilisant de nouveau l'inégalité $1-\lambda\leqslant e^{-\lambda}$ on obtient $d_{VT}(p_X,\pi_\lambda)\leqslant\lambda^2$.

13 ▷ D'après la question 5, $2\,d_{VT}(p_{X_n},\pi_1)=\sum_{k=0}^{+\infty}\Big|\mathbb{P}(X_n=k)-\pi_1(k)\Big|=\sum_{k=0}^{n}\frac{1}{k!}\Big|\sum_{i=0}^{n-k}\frac{(-1)^i}{i!}-e^{-1}\Big|+e^{-1}\sum_{k=n+1}^{+\infty}\frac{1}{k!}$.
Or $e^{-1}=\sum_{i=0}^{+\infty}\frac{(-1)^i}{i!}$ donc $2\,d_{VT}(p_{X_n},\pi_1)=\sum_{k=0}^{n}\frac{1}{k!}\Big|\sum_{i=n-k+1}^{+\infty}\frac{(-1)^i}{i!}\Big|+e^{-1}\sum_{k=n+1}^{+\infty}\frac{1}{k!}$.

14 ▷ Pour tout $k\geqslant n+2$, $k!=(n+1)!\prod_{i=n+2}^{k}i\geqslant(n+1)!\prod_{i=n+2}^{k}(n+2)=(n+1)!(n+2)^{k-n-1}$ donc

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

En calculant cette somme géométrique on obtient $r_n\leqslant\frac{n+2}{(n+1)(n+1)!}\sim\frac{1}{(n+1)!}$. Mais par ailleurs on a de façon évidente $\frac{1}{(n+1)!}\leqslant r_n$ donc $r_n\sim\frac{1}{(n+1)!}$.

15 ▷ D'après le critère spécial relatif aux séries alternées, $\Big|\sum_{i=n-k+1}^{+\infty}\frac{(-1)^i}{i!}\Big|\leqslant\Big|\frac{(-1)^{n-k+1}}{(n-k+1)!}\Big|=\frac{1}{(n-k+1)!}$ donc

$$\sum_{k=0}^{n}\frac{1}{k!}\Big|\sum_{i=n-k+1}^{+\infty}\frac{(-1)^i}{i!}\Big|\leqslant\sum_{k=0}^{n}\frac{1}{k!(n-k+1)!}=\frac{1}{(n+1)!}\sum_{k=0}^{n}\binom{n+1}{k}\leqslant\frac{1}{(n+1)!}\sum_{k=0}^{n+1}\binom{n+1}{k}=\frac{2^{n+1}}{(n+1)!}$$

Ainsi, $d_{VT}(p_{X_n},\pi_1)=O\Big(\frac{2^n}{(n+1)!}\Big)+O\Big(\frac{1}{(n+1)!}\Big)=O\Big(\frac{2^n}{(n+1)!}\Big)$.

3 – Autres estimations de distances en variation totale

16 ▷ Si x et y sont deux distributions, $x*y$ est bien à valeurs positives, et par produit de Cauchy de deux séries positives convergentes,

$$\Big(\sum_{i=0}^{+\infty} x(i)\Big)\Big(\sum_{j=0}^{+\infty} x(j)\Big) = \sum_{k=0}^{+\infty}\Big(\sum_{i+j=k} x(i)y(j)\Big) = \sum_{k=0}^{+\infty} x*y(k)$$

donc $\displaystyle\sum_{k=0}^{+\infty} x*y(k) = 1\times 1 = 1$; $x*y$ est bien une distribution sur $\mathbb{N}$.

17 ▷ Puisque $X \perp\!\!\!\perp Y$, pour tout $k\in\mathbb{N}$,

$$\mathbb{P}(X+Y=k) = \sum_{i=0}^{k}\mathbb{P}\big(X=i \text{ et } Y=k-i\big) = \sum_{i=0}^{k}\mathbb{P}(X-i)\mathbb{P}(Y=k-i) = \sum_{i=0}^{k} p_X(i)p_Y(k-i) = p_X * p_Y(k)$$

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

18 ▷ Pour tout $k\in\mathbb{N}$,

$$\begin{aligned}\big|(x*y)(k)-(u*v)(k)\big| &\leqslant \sum_{i+j=k}\big|x(i)y(j)-u(i)v(j)\big| = \sum_{i+j=k}\big|(x(i)-u(i))y(j)+u(i)(y(j)-v(j))\big|\\ &\leqslant \sum_{i+j=k}\big|x(i)-u(i)\big|y(j) + \sum_{i+j=k} u(i)\big|y(j)-v(j)\big|\end{aligned}$$

19 ▷ Il s'agit d'une somme de termes positifs donc

$$\sum_{k=0}^{+\infty}\sum_{i+j=k}\big|x(i)-u(i)\big|y(j) = \sum_{k=0}^{+\infty}\sum_{i=0}^{k}\big|x(i)-u(i)\big|y(k-i) = \sum_{i=0}^{+\infty}\sum_{k=i}^{+\infty}\big|x(i)-u(i)\big|y(k-i) = \sum_{i=0}^{+\infty}\big|x(i)-u(i)\big| \quad \text{car } \sum_{k=0}^{+\infty} y(k) = 1.$$

De même, $\displaystyle\sum_{k=0}^{+\infty}\sum_{i+j=k} u(i)\big|y(j)-v(j)\big| = \sum_{j=0}^{+\infty}\big|y(j)-v(j)\big|$ donc d'après l'inégalité précédente, $d_{VT}(x*y,u*v)\leqslant d_{VT}(x,u)+d_{VT}(y,v)$.

20 ▷ Raisonnons par récurrence sur n :

– si $n=1$, le résultat a déjà été montré à la question 12.

– si $n>1$, supposons le résultat acquis au rang $n-1$. Soit $(X_1,\dots,X_n)$ une suite de variables aléatoires indépendantes et de même loi binomiale de paramètre λ. Posons $U=\displaystyle\sum_{k=1}^{n} X_k$ et $V=\displaystyle\sum_{k=1}^{n-1} X_k$. On sait que $U\sim\mathscr{B}(n,\lambda)$, $V\sim\mathscr{B}(n-1,\lambda)$, et d'après le lemme des coalitions, $V\perp\!\!\!\perp X_n$.
D'après la question 17, $p_U = p_{X_n} * p_V$, et par hypothèse de récurrence, $d_{VT}(p_V,\pi_{(n-1)\lambda})\leqslant(n-1)\lambda^2$.
D'après la question 19, $d_{VT}(p_U,\pi_\lambda*\pi_{(n-1)\lambda})\leqslant d_{VT}(p_V,\pi_{(n-1)\lambda})+d_{VT}(p_{X_n},\pi_\lambda)\leqslant(n-1)\lambda^2+\lambda^2 = n\lambda^2$.
Pour conclure, il reste à établir que $\pi_\lambda*\pi_{(n-1)\lambda}=\pi_{n\lambda}$; pour ce faire, on montre plus généralement que $\pi_\lambda*\pi_\mu=\pi_{\lambda+\mu}$:

$$\forall k\in\mathbb{N},\quad \pi_\lambda*\pi_\mu(k) = e^{-(\lambda+\mu)}\sum_{i+j=k}\frac{\lambda^i\mu^j}{i!j!} = \frac{e^{-(\lambda+\mu)}}{k!}\sum_{i=0}^{n}\binom{k}{i}\lambda^i\mu^{k-i} = e^{-(\lambda+\mu)}\frac{(\lambda+\mu)^k}{k!} = \pi_{\lambda+\mu}(k)$$

La récurrence se propage.

21 ▷ D'après la question précédente appliquée à $\lambda=\dfrac{\alpha}{n}$ on a $d_{VT}(p_{B_n},\pi_\alpha)\leqslant\dfrac{\alpha^2}{n}$ donc $\displaystyle\lim_{n\to+\infty} d_{VT}(p_{B_n},\pi_\alpha)=0$.

Pour tout $k\in\mathbb{N}$ on a $\left|\mathbb{P}(B_n=k)-e^{-\alpha}\dfrac{\alpha^k}{k!}\right|\leqslant d_{VT}(p_{B_n},\pi_\alpha)$ donc $\displaystyle\lim_{n\to+\infty}\mathbb{P}(B_n=k)=e^{-\alpha}\frac{\alpha^k}{k!}$.

22 ▷ Considérons deux suites de variables aléatoires indépendantes (X_n) et (Y_n) telles que $\mathrm{X}_k \sim \mathscr{B}\left(\frac{\alpha}{n}\right)$ et $\mathrm{Y}_k \sim \mathscr{B}\left(\frac{\beta}{n}\right)$, ainsi que $\mathrm{U}_n = \mathrm{X}_1 + \cdots + \mathrm{X}_n$ et $\mathrm{V}_n = \mathrm{Y}_1 + \cdots + \mathrm{Y}_n$. Alors $\mathrm{U}_n \sim \mathscr{B}\left(n, \frac{\alpha}{n}\right)$ et $\mathrm{V}_n \sim \mathscr{B}\left(n, \frac{\beta}{n}\right)$.
À l'aide de la question 17 on établit par récurrence que $p_{\mathrm{U}_n} = p_{\mathrm{X}_1} * \cdots * p_{\mathrm{X}_n}$ et $p_{\mathrm{V}_n} = p_{\mathrm{Y}_1} * \cdots * p_{\mathrm{Y}_n}$.
À l'aide de la question 19 on prouve par récurrence que $\mathrm{d}_{\mathrm{VT}}(p_{\mathrm{U}_n}, p_{\mathrm{V}_n}) \leqslant \sum_{k=1}^{n} \mathrm{d}_{\mathrm{VT}}(p_{\mathrm{X}_k}, p_{\mathrm{Y}_k})$ puis, grâce à la question 11, que $\mathrm{d}_{\mathrm{VT}}(p_{\mathrm{U}_n}, p_{\mathrm{V}_n}) \leqslant \sum_{k=1}^{n} \left|\frac{\alpha}{n} - \frac{\beta}{n}\right| = |\alpha - \beta|$.
D'après la question 10, $\mathrm{d}_{\mathrm{VT}}(\pi_\alpha, \pi_\beta) \leqslant \mathrm{d}_{\mathrm{VT}}(\pi_\alpha, p_{\mathrm{U}_n}) + \mathrm{d}_{\mathrm{VT}}(p_{\mathrm{U}_n}, p_{\mathrm{V}_n}) + \mathrm{d}_{\mathrm{VT}}(p_{\mathrm{V}_n}, \pi_\beta)$, et comme on l'a vu à la question 21, $\mathrm{d}_{\mathrm{VT}}(\pi_\alpha, \pi_\beta) \leqslant |\alpha - \beta| + \frac{\alpha^2}{n} + \frac{\beta^2}{n}$. Il reste à faire tendre n vers $+\infty$ pour obtenir $\mathrm{d}_{\mathrm{VT}}(\pi_\alpha, \pi_\beta) \leqslant |\alpha - \beta|$.