DISTANCE DE LÉVY–PROHOROV ET COUPLAGES OPTIMAUX

L'objectif de ce devoir est de démontrer un résultat du à V. Strassen, qui relie la distance de Lévy–Prohorov $d_{\mathrm{LP}}(\mu,\nu)$ à la construction de couplages optimaux entre deux mesures de probabilités μ et ν (Strassen, The existence of probability measures with given marginals, *Ann. Math. Statist.* 36(2):423-439, 1965).

1. Séparation sous un couplage

On fixe un espace polonais $\mathfrak{X}$ muni d'une distance complète d, et on rappelle que la distance de Lévy–Prohorov $d_{\mathrm{LP}}(\mu,\nu)$ entre deux mesures de probabilités μ et ν sur $\mathfrak{X}$ est définie par

$$d_{\mathrm{LP}}(\mu,\nu) = \inf\{\varepsilon > 0 \mid \forall A \in \mathcal{B}(\mathfrak{X}),\ \mu(A) \leq \nu(A^\varepsilon) + \varepsilon\},$$

où $\mathcal{B}(\mathfrak{X})$ est l'ensemble des parties boréliennes de $\mathfrak{X}$, et $A^\varepsilon = \{x \in \mathfrak{X} \mid d(x,A) < \varepsilon\}$. Un *couplage* entre les deux mesures μ et ν est une mesure de probabilités M sur $\mathfrak{X} \times \mathfrak{X}$ dont les deux marginales sont $(\mathrm{pr}_1)_*(M) = \mu$ et $(\mathrm{pr}_2)_*(M) = \nu$, avec $\mathrm{pr}_1(x,y) = x$ et $\mathrm{pr}_2(x,y) = y$. Par exemple, $\mu \times \nu$ est un couplage de μ et ν, mais il existe en général de nombreux autres couplages (non indépendants) possibles entre μ et ν.

1. On note $\mathscr{C}(\mu,\nu)$ l'ensemble de tous les couplages entre μ et ν ; c'est une partie de l'ensemble des mesures de probabilités $\mathscr{M}^1(\mathfrak{X} \times \mathfrak{X})$. En utilisant le théorème de Prohorov, montrer que cette partie est compacte pour la topologie de la convergence faible sur $\mathscr{M}^1(\mathfrak{X} \times \mathfrak{X})$.

2. Si M est un couplage entre μ et ν, on dit que les lois μ et ν sont séparées de moins de ε sous le couplage M si $M(\{(x,y) \mid d(x,y) \geq \varepsilon\}) \leq \varepsilon$. La *séparation* entre μ et ν sous M est définie par :

   $$\mathrm{sep}(M) = \inf\{\varepsilon > 0 \mid \text{les lois } \mu \text{ et } \nu \text{ sont séparées de moins de } \varepsilon \text{ sous } M\}.$$

   Montrer que pour tout couplage $M \in \mathscr{C}(\mu,\nu)$, $\mathrm{sep}(M) \geq d_{\mathrm{LP}}(\mu,\nu)$.

2. Théorème de Strassen et réduction au cas fini

Le théorème de Strassen est l'énoncé suivant :

(S) Pour toutes mesures de probabilité μ et ν sur $\mathfrak{X}$, il existe un couplage optimal $M \in \mathscr{C}(\mu,\nu)$ tel que $\mathrm{sep}(M) = d_{\mathrm{LP}}(\mu,\nu)$.

Les questions de cette section donnent des énoncés équivalents au théorème, et montrent qu'il suffit de l'établir pour des mesures uniformes sur des ensembles finis.

3. On note $\Delta = d_{\mathrm{LP}}(\mu,\nu)$, et on considère les hypothèses suivantes : pour tout $\delta > 0$, il existe un couplage $M \in \mathscr{C}(\mu,\nu)$ dépendant *a priori* de δ et que l'on peut décomposer sous la forme $M = M_1 + M_2$ avec :

   (S1) M_1 et M_2 mesures boréliennes positives sur $\mathfrak{X}^2$;

   (S2) $M_1(\{(x,y) \mid d(x,y) > \Delta + \delta\}) = 0$;

Date: à rendre pour le 9 novembre 2018.

(S3) $M_2(\mathfrak{X}^2) \leq \Delta + \delta$.

Montrer que le théorème de Strassen (S) est équivalent aux trois hypothèses (S1), (S2) et (S3). Montrer aussi que si ces hypothèses sont vérifiées, alors on peut en fait choisir le couplage M indépendamment du paramètre δ, avec ensuite $M_1 = M_{|\{(x,y) \mid d(x,y) \leq \Delta+\delta\}}$ et $M_2 = M_{|\{(x,y) \mid d(x,y) > \Delta+\delta\}}$.

4. Soit $\mathscr{S} \subset \mathscr{M}^1(\mathfrak{X}) \times \mathscr{M}^1(\mathfrak{X})$ l'ensemble des paires de mesures de probabilités (μ, ν) qui vérifient le théorème de Strassen :

$$\mathscr{S} = \{(\mu, \nu) \mid \forall \delta > 0,\ \exists M \in \mathscr{C}(\mu, \nu) \text{ satisfaisant les hypothèses (S1),(S2),(S3)}\}.$$

Montrer que $\mathscr{S}$ est une partie fermée de $\mathscr{M}^1(\mathfrak{X}) \times \mathscr{M}^1(\mathfrak{X})$.

5. Si $\mathfrak{X}$ est un espace polonais, montrer que l'ensemble $\mathscr{F}$ des mesures de probabilités de la forme

$$\mu = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i}$$

avec $N \geq 1$ et $x_1, \ldots, x_N \in \mathfrak{X}$ est dense dans $\mathscr{M}^1(\mathfrak{X})$.

6. Pour $N \geq 1$, on note $\mathscr{F}_N$ l'ensemble des mesures de probabilités uniformes sur N points $x_1, \ldots, x_N$ (avec éventuellement des identités entre ces points) ; ainsi, l'ensemble de mesures considéré à la question précédente est $\mathscr{F} = \bigcup_{N\geq 1} \mathscr{F}_N$. On note par ailleurs $\mathscr{S}_\delta$ l'ensemble des paires de mesures qui vérifient (S1), (S2) et (S3) pour un paramètre δ fixé ; ainsi, $\mathscr{S} = \bigcap_{\delta>0} \mathscr{S}_\delta$. Montrer que si $\mathscr{F}_N \times \mathscr{F}_N \subset \mathscr{S}_\delta$ pour tout $N \geq N_\delta$, alors le théorème de Strassen est vrai pour toute paire de mesures sur $\mathfrak{X}$.

3. Problème des mariages

Les questions précédentes ramènent la preuve du théorème de Strassen à un problème essentiellement combinatoire, lié au problème des mariages. Soit C et D deux ensembles finis de même cardinal $N \geq 1$, et $K \subset C \times D$ une relation entre éléments de C et de D. Un K-*mariage* est une application bijective $f : C \to D$ telle que $(c, f(c)) \in K$ pour tout $c \in C$. Le théorème des mariages de Hall est l'énoncé suivant :

(H) Étant donnée une relation K entre deux ensembles finis C et D de même cardinal N, il existe un K-mariage si et seulement si, pour toute partie $A \subset C$,

$$A^K = \{d \in D \mid \exists c \in A,\ (c, d) \in K\}$$

a un cardinal $\operatorname{card} A^K$ plus grand que $\operatorname{card} A$.

L'objectif de cette section est d'établir ce résultat ; on notera $\operatorname{card} A = |A|$ dans ce qui suit.

7. Montrer que $|A^K| \geq |A|$ pour toute partie $A \subset C$ est une condition nécessaire.

8. On raisonne par récurrence sur N, et on suppose que l'équivalence entre l'existence d'un K-mariage et la condition $|A^K| \geq |A|$ est démontrée pour toute paire d'ensembles (C, D) de cardinal inférieur à $N - 1$. On suppose par l'absurde qu'il existe C, D de cardinal N et une relation $K \subset C \times D$ vérifiant $|A^K| \geq |A|$ pour tout $A \subset C$, mais telle qu'il n'existe pas de K-mariage entre C et D.

8.1. Montrer qu'il existe une paire $(c, d) \in K$ et une partie $A \subset C \setminus \{c\}$ telle que $|A^K| = |A|$.

8.2. Montrer qu'il existe une partie $B \subset C \setminus A$ telle que $|B^K \setminus A^K| < |B|$.

8.3. Obtenir une contradiction en considérant $A \cup B$.

4. Preuve du théorème de Strassen

On fixe dans cette dernière section $\delta > 0$, $N \geq N_\delta = \lceil \frac{1}{\delta} \rceil$ et $\mu = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}$, $\nu = \frac{1}{N} \sum_{i=1}^N \delta_{y_i}$. L'objectif est de montrer que $(\mu, \nu) \in \mathscr{S}_\delta$. On note comme précédemment $\Delta = d_{\mathrm{LP}}(\mu, \nu)$, et $M = \lceil \Delta N \rceil$. On pose $C = D = [\![1, M+N]\!]$, et on définit une relation $K \subset C \times D$ par :

$$(c, d) \in K \iff c > N \text{ ou } d > N \text{ ou } d(x_c, y_d) \leq \Delta.$$

9. Montrer qu'il existe un K-mariage f entre C et D, qu'on fixe dans la suite.

10. Soit $W = \{c \in [\![1, N]\!] \mid f(c) \in [\![1, N]\!]\}$. Estimer $|W|$, et montrer qu'il existe une application bijective $g : [\![1, N]\!] \to [\![1, N]\!]$ telle que $g_{|W} = f_{|W}$. On pose

$$M_1 = \frac{1}{N} \sum_{c \in W} \delta_{(x_c, y_{g(c)})};$$
$$M_2 = \frac{1}{N} \sum_{c \in [\![1,N]\!] \setminus W} \delta_{(x_c, y_{g(c)})};$$
$$M = M_1 + M_2.$$

Montrer que $M \in \mathscr{C}(\mu, \nu)$ et vérifie les hypothèses (S1), (S2) et (S3) vis-à-vis du paramètre δ. Conclure.

Corrigé

1. Comme pr_1 et pr_2 sont des applications continues entre $\mathfrak{X} \times \mathfrak{X}$ et $\mathfrak{X}$, elles induisent des applications continues $(\mathrm{pr}_1)_*$ et $(\mathrm{pr}_2)_*$ entre $\mathscr{M}^1(\mathfrak{X} \times \mathfrak{X})$ et $\mathscr{M}^1(\mathfrak{X})$, ces espaces étant munis de la toplogie de la convergence faible. Alors,
$$\mathscr{C}(\mu,\nu) = \left(((\mathrm{pr}_1)_*)^{-1}(\mu) \right) \cap \left(((\mathrm{pr}_2)_*)^{-1}(\nu) \right)$$
est l'intersection de deux images réciproques par des applications continues de parties fermées, donc est une partie fermée de $\mathscr{M}^1(\mathfrak{X})$. Il reste à montrer la relative compacité, qui est équivalente à la tension par le théorème de Prohorov. Soit $\varepsilon > 0$ et une partie compacte $K \subset \mathfrak{X}$ telle que $\mu(K) \geq 1-\varepsilon$ et $\nu(K) \geq 1-\varepsilon$. Notons que
$$\mathfrak{X}^2 = K^2 \cup ((\mathfrak{X} \setminus K) \times \mathfrak{X}) \cup (\mathfrak{X} \times (\mathfrak{X} \setminus K)).$$
Si M est un couplage entre μ et ν, on a donc
$$\begin{aligned} 1 = M(\mathfrak{X}^2) &\leq M(K^2) + M((\mathfrak{X} \setminus K) \times \mathfrak{X}) \cup M(\mathfrak{X} \times (\mathfrak{X} \setminus K)) \\ &\leq M(K^2) + \mu(\mathfrak{X} \setminus K) + \nu(\mathfrak{X} \setminus K) \leq M(K^2) + 2\varepsilon \end{aligned}$$
et K^2 est un compact de $\mathfrak{X}^2$ tel que $M(K^2) \geq 1 - 2\varepsilon$ pour tout M dans $\mathscr{C}(\mu,\nu)$. Cette partie de $\mathscr{M}^1(\mathfrak{X}^2)$ est donc fermée et relativement compacte, donc compacte.

2. Soit M un couplage entre μ et ν, et $\varepsilon > 0$ plus grand que la séparation entre μ et ν sous M. Si A est un borélien de $\mathfrak{X}$, alors
$$\begin{aligned} \mu(A) = M(A \times \mathfrak{X}) &\leq M(A \times A^\varepsilon) + M(\{(x,y) \mid x \in A \text{ et } d(x,y) \geq \varepsilon\}) \\ &\leq M(\mathfrak{X} \times A^\varepsilon) + \varepsilon = \nu(A^\varepsilon) + \varepsilon. \end{aligned}$$
Par conséquent, $\varepsilon \geq d_{\mathrm{LP}}(\mu,\nu)$, et par passage à la borne inférieure, $\mathrm{sep}(M) \geq d_{\mathrm{LP}}(\mu,\nu)$.

3. On note dans ce qui suit $\mathrm{diag}(L) = \{(x,y) \mid d(x,y) \leq L\}$ l'ensemble des paires de points à distance inférieure à un niveau L. Si le théorème de Strassen est vrai, soit M un couplage optimal, $\delta > 0$, $M_1 = M_{|\mathrm{diag}(\Delta+\delta)}$ et $M_2 = M_{|\mathfrak{X}^2 \setminus \mathrm{diag}(\Delta+\delta)}$. On a $M = M_1 + M_2$ et
$$M_1(\{(x,y) \mid d(x,y) > \Delta + \delta\}) = 0$$
par construction, et par définition de la séparation entre μ et ν sous M,
$$M_2(\mathfrak{X}^2) = M(\{(x,y) \mid d(x,y) > \Delta + \delta\}) \leq M(\{(x,y) \mid d(x,y) \geq \Delta + \delta\}) \leq \Delta + \delta.$$
Réciproquement, supposons donné pour tout $\delta > 0$ un couplage $M_\delta = M_{1,\delta} + M_{2,\delta}$ vérifiant les hypothèses. On choisit une suite $(\delta_n)_{n\in\mathbb{N}}$ tendant vers 0 et telle que $(M_{\delta_n})_{n\in\mathbb{N}}$ converge faiblement vers un couplage M ; c'est possible par compacité de $\mathscr{C}(\mu,\nu)$. Montrons que la séparation entre μ et ν sous M est inférieure à Δ. Si $\delta > 0$, alors par Portmanteau appliqué à l'ouvert $\mathfrak{X}^2 \setminus \mathrm{diag}(\Delta + \frac{\delta}{2})$,
$$\begin{aligned} M(\{(x,y) \mid d(x,y) \geq \Delta + \delta\}) &\leq M\left(\left\{(x,y) \mid d(x,y) > \Delta + \frac{\delta}{2}\right\}\right) \\ &\leq \liminf_{n\to\infty} M_{\delta_n}\left(\left\{(x,y) \mid d(x,y) > \Delta + \frac{\delta}{2}\right\}\right) \\ &\leq \liminf_{n\to\infty} M_{\delta_n}(\{(x,y) \mid d(x,y) > \Delta + \delta_n\}) \\ &\leq \liminf_{n\to\infty} M_{2,\delta_n}(\mathfrak{X}) \leq \liminf_{n\in\mathbb{N}} (\Delta + \delta_n) \\ &\leq \Delta \leq \Delta + \delta. \end{aligned}$$
En faisant tendre δ vers 0, on obtient $\mathrm{sep}(M) \leq \Delta$, et il y a égalité d'après la seconde question.

4. Énonçons pour commencer une généralisation facile du résultat de la première question : si $\mathscr{K}_1$ et $\mathscr{K}_2$ sont des parties (relativement) compactes de $\mathscr{M}^1(\mathfrak{X})$, alors l'ensemble $\mathscr{C}(\mathscr{K}_1, \mathscr{K}_2)$ des couplages de lois $\mu_1 \in \mathscr{K}_1$ et $\mu_2 \in \mathscr{K}_2$ est également (relativement) compacte dans $\mathscr{M}^1(\mathfrak{X}^2)$. En effet, dans la preuve de la relative compacité équivalente à la tension, on peut choisir un compact $K \subset \mathfrak{X}$ commun à toutes les mesures μ_1 et μ_2 tel que $\mu_1(K) \geq 1-\varepsilon$ et $\mu_2(K) \geq 1-\varepsilon$; alors, tout couplage M dans $\mathscr{C}(\mathscr{K}_1, \mathscr{K}_2)$ vérifie $M(K^2) \geq 1 - 2\varepsilon$.

   Soit $(\mu_n, \nu_n)_{n\in\mathbb{N}}$ une suite de paires de mesures dans $\mathscr{S}$ qui tend vers une limite (μ, ν) ; on veut montrer que $(\mu, \nu) \in \mathscr{S}$. On choisit des couplages optimaux $M_n \in \mathscr{C}(\mu_n, \nu_n)$ qui vérifient le théorème de Strassen. Comme $\{\mu_n,\ n \in \mathbb{N}\}$ et $\{\nu_n,\ n \in \mathbb{N}\}$ sont deux parties relativement compactes de $\mathscr{M}^1(\mathfrak{X})$, d'après ce qui précède, les couplages M_n forment aussi une suite relativement compacte, et à extraction près, on peut supposer que $M_n \rightharpoonup M$, où M est une certaine mesure de probabilités sur $\mathfrak{X}^2$. Comme $(\mathrm{pr}_1)_*$ et $(\mathrm{pr}_2)_*$ sont des applications continues, M est un couplage entre μ et ν. Notons $\Delta_n = d_{\mathrm{LP}}(\mu_n, \nu_n)$ et $\Delta = d_{\mathrm{LP}}(\mu, \nu)$. La distance de Lévy–Prohorov métrise la topologie de la convergence faible, donc c'est une application continue vis-à-vis de cette topologie, et $\Delta = \lim_{n\to\infty} \Delta_n$. Fixons $\delta > 0$, et notons δ_n la suite définie par
$$\Delta + \delta = \Delta_n + \delta_n.$$
   Quitte à enlever quelques termes de la suite, on peut supposer $\delta_n > 0$ pour tout n, car on a $\delta = \lim_{n\to\infty} \delta_n$. On note $M_{n,1} = (M_n)_{|\mathrm{diag}(\Delta+\delta)}$ et $M_{n,2} = (M_n)_{|\mathfrak{X}^2\setminus\mathrm{diag}(\Delta+\delta)}$; pour tout n, la décomposition $M_n = M_{n,1} + M_{n,2}$ vérifie les hypothèses (S1), (S2) et (S3) vis-à-vis des mesures μ_n et ν_n et du paramètre δ_n. Posons $M_1 = M_{|\mathrm{diag}(\Delta+\delta)}$ et $M_2 = M_{\mathfrak{X}^2\setminus\mathrm{diag}(\Delta+\delta)}$. Les hypothèses (S1) et (S2) sont clairement vérifiées, et pour (S3), on a par Portmanteau
$$\begin{aligned} M_2(\mathfrak{X}^2) = M(\mathfrak{X}^2 \setminus \mathrm{diag}(\Delta+\delta)) &\leq \liminf_{n\to\infty} M_n(\mathfrak{X}^2 \setminus \mathrm{diag}(\Delta_n + \delta_n)) \\ &\leq \liminf_{n\to\infty} M_{n,2}(\mathfrak{X}^2) = \liminf_{n\to\infty} (\Delta_n + \delta_n) = \Delta + \delta. \end{aligned}$$
   Ainsi, la paire de mesures (μ, ν) appartient bien à la classe $\mathscr{S}$.

5. Fixons une mesure de probabilités μ sur $\mathfrak{X}$, et considérons des variables $X_1, X_2, \ldots, X_N, \ldots$ indépendantes et de loi μ, et les mesures empiriques $\mu_N = \frac{1}{N}\sum_{i=1}^N \delta_{X_i}$. On rappelle le résultat suivant : presque sûrement, $\mu_N \rightharpoonup_{N\to\infty} \mu$. En effet, fixons une base dénombrable d'ouverts $(U_i)_{i\in I}$ de $\mathfrak{X}$: tout ouvert de $\mathfrak{X}$ s'écrit sous la forme $U = \bigcup_{j\in J} U_j$ avec $J \subset I$. On note dans ce qui suit $U_J = \bigcup_{j\in J} U_j$ si $J \subset I$. Par la loi forte des grands nombres, pour toute partie finie $J \subset I$, $\mu_N(U_J) \to_{N\to\infty} \mu(U_J)$ presque sûrement, et comme l'ensemble de ces parties finies est dénombrable, on a avec probabilité 1 :
$$\text{pour toute partie finie } J \subset I,\ \lim_{N\to\infty} \mu_N(U_J) = \mu(U_J).$$
   Alors, si U est un ouvert arbitraire de $\mathfrak{X}$, pour tout $\varepsilon > 0$, il existe U_J avec J partie finie tel que $\mu(U_J) \geq \mu(U) - \varepsilon$, donc
$$\liminf_{N\to\infty} \mu_N(U) \geq \liminf_{N\to\infty} \mu_N(U_J) = \mu(U_J) \geq \mu(U) - \varepsilon,$$
   et $\liminf_{N\to\infty} \mu_N(U) \geq \mu(U)$ en faisant tendre ε vers 0. Ceci est vrai avec probabilité 1, et par le théorème de Portmanteau, c'est équivalent à $\mu_N \rightharpoonup_{N\to\infty} \mu$. En particulier, choisissant un $\omega \in \Omega$ tel qu'on ait cette convergence, la suite $(\mu_N(\omega))_{N\geq 1}$ converge vers μ et est une suite de mesures uniformes sur des ensembles finis, donc la densité de cet ensemble de mesures $\mathscr{F}$ dans $\mathscr{M}^1(\mathfrak{X})$ est bien démontrée.

6. Comme $\mathscr{F}$ est une partie dense de $\mathscr{M}^1(\mathfrak{X})$ et $\mathscr{S}$ est une partie fermée de $\mathscr{M}^1(\mathfrak{X}) \times \mathscr{M}^1(\mathfrak{X})$, si $\mathscr{F} \times \mathscr{F} \subset \mathscr{S}$, alors
$$\mathscr{M}^1(\mathfrak{X}) \times \mathscr{M}^1(\mathfrak{X}) = \overline{\mathscr{F} \times \mathscr{F}} \subset \overline{\mathscr{S}} = \mathscr{S}.$$
Comme $\mathscr{S} = \bigcap_{\delta>0} \mathscr{S}_\delta$, pour établir le théorème de Strassen, il suffit donc de montrer que $\mathscr{F} \times \mathscr{F} \subset \mathscr{S}_\delta$ pour tout δ.

   Notons que si N divise M, alors $\mathscr{F}_N \subset \mathscr{F}_M$. En effet, avec $M = kN$ et $\mu = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$ dans $\mathscr{F}_N$, on a
$$\mu = \frac{1}{kN}\sum_{i=1}^N \sum_{j=1}^k \delta_{x_i},$$
donc μ est aussi dans $\mathscr{F}_M$. On en déduit que $\mathscr{F}_N \times \mathscr{F}_M \subset \mathscr{F}_{NM} \times \mathscr{F}_{NM}$, et donc que
$$\mathscr{F} \times \mathscr{F} = \bigcup_{N,M\geq 1} \mathscr{F}_N \times \mathscr{F}_M = \bigcup_{N\geq 1} \mathscr{F}_N \times \mathscr{F}_N = \bigcup_{N\geq N_\delta} \mathscr{F}_N \times \mathscr{F}_N$$
pour n'importe quel niveau N_δ. Ainsi, $\mathscr{F} \times \mathscr{F} \subset \mathscr{S}_\delta$ est équivalent au fait que $\mathscr{F}_N \times \mathscr{F}_N \subset \mathscr{S}_\delta$ pour tout $N \geq N_\delta$ suffisamment grand.

7. La condition $|A^K| \geq |A|$ est clairement nécessaire à l'existence d'un K-mariage : si f est un K-mariage, alors A^K contient $f(A)$ pour toute partie A, et $|A^K| \geq |f(A)| = |A|$.

8. Sous les hypothèses de la question, fixons $c \in C$ et $d \in \{c\}^K$, qui existe puisque $\mathrm{card}\,\{c\}^K \geq \mathrm{card}\,\{c\} = 1$. S'il existait un K-mariage $f : C \setminus \{c\} \to D \setminus \{d\}$, alors on pourrait l'étendre en un K-mariage $f : C \to D$ en posant $f(c) = d$. Donc, il n'y a pas de K-mariage entre $C \setminus \{c\}$ et $D \setminus \{d\}$, et par hypothèse de récurrence, il existe une partie $A \subset C \setminus \{c\}$ telle que
$$\big|\{e \in D \mid e \neq d \text{ et } \exists a \in A \text{ avec } (a,e) \in K\}\big| < |A|.$$
Notons par ailleurs que par hypothèse,
$$|A^K| = \big|\{e \in D \mid \exists a \in A \text{ avec } (a,e) \in K\}\big| \geq |A|.$$
Ces deux estimées ne sont possibles que si $|A^K| = |A|$. Alors, par hypothèse de récurrence, puisque $|A| < |C| = N$, il existe un K-mariage f entre A et A^K. S'il existe aussi un K-mariage entre $C \setminus A$ et $D \setminus A^K$, alors on peut l'utiliser pour compléter f en un K-mariage global entre C et D. Sinon, de nouveau par hypothèse de récurrence, il existe une partie $B \subset C \setminus A$ telle que
$$\big|\{e \in D \setminus A^K \mid \exists b \in B \text{ avec } (b,e) \in K\}\big| < |B|.$$
Notons que le terme de gauche ci-dessus est $|B^K \setminus A^K|$. Alors,
$$|(A \cup B)^K| = |A^K| + |B^K \setminus A^K| < |A| + |B| = |A \cup B|,$$
ce qui contredit l'hypothèse sur la relation K. Le théorème des mariages est donc encore vrai au rang N, et par récurrence il est vrai pour toute taille N d'ensembles.

9. On va vérifier la condition de Hall pour la relation K. Soit $A \subset C$. Si $A = \emptyset$, on a évidemment $|A^K| \geq |A|$. Par ailleurs, si $A \cap [\![N+1, N+M]\!] \neq \emptyset$, alors $A^K = D$ et de nouveau $|A^K| \geq |A|$. Il reste à traiter le cas où $A \subset [\![1,N]\!]$ et est non vide. Alors, A^K contient tout $[\![N+1, N+M]\!]$, donc est au moins de cardinal M. Les autres éléments de A^K sont les indices $b \in [\![1,N]\!]$ tels

que $d(x_a, y_b) \leq \Delta$ pour un certain $a \in A$. Pour tout $\eta > 0$, par définition de la distance de Lévy–Prohorov,

$$\frac{|A|}{N} = \mu(\{x_a,\ a \in A\}) \leq \nu(\{y_b \mid d(y_b, \{x_a,\ a \in A\}) < \Delta + \eta\}) + \Delta + \eta$$
$$\leq \frac{|\{b \mid d(y_b, \{x_a,\ a \in A\}) < \Delta + \eta\}|}{N} + \Delta + \eta.$$

Comme ceci est vrai pour tout $\eta > 0$, on obtient

$$\frac{|A|}{N} \leq \frac{|\{b \mid d(y_b, \{x_a,\ a \in A\}) \leq \Delta\}|}{N} + \Delta \leq \frac{|\{b \mid d(y_b, \{x_a,\ a \in A\}) \leq D\}| + M}{N},$$

et le numérateur du terme de droite est $|A^K|$. Ceci achève la vérification.

10. Les éléments de $[\![1, N]\!]$ qui ne sont pas mariés par f à d'autres éléments de $[\![1, N]\!]$ sont envoyés dans $[\![N+1, N+M]\!]$, donc $|W| \geq N - M$. Le mariage partiel $f_{|W} : W \to f(W)$ peut bien sûr être complété en une application bijective $g : [\![1, N]\!] \to [\![1, N]\!]$, en choisissant aribtrairement les images d'éléments $c \in [\![1, N]\!] \setminus W$. Considérons alors les mesures M_1 et M_2 proposées par l'énoncé. On a $M = \frac{1}{N} \sum_{c=1}^{N} \delta_{(x_c, y_{g(c)})}$, et donc

$$(\mathrm{pr}_1)_*(M) = \frac{1}{N} \sum_{c=1}^{N} \delta_{x_c} = \mu$$
$$(\mathrm{pr}_2)_*(M) = \frac{1}{N} \sum_{c=1}^{N} \delta_{y_{g(c)}} = \frac{1}{N} \sum_{d=1}^{N} \delta_{y_d} = \nu$$

en utilisant le caractère bijectif de g pour la seconde identité. La mesure M est donc bien un couplage entre μ et ν. Par construction, $M = M_1 + M_2$ et M_1 ne charge que $\mathrm{diag}(\Delta)$, donc

$$M_1(\{(x, y) \mid d(x, y) > \Delta + \delta\}) \leq M_1(\mathfrak{X}^2 \setminus \mathrm{diag}(\Delta)) = 0.$$

Finalement, $M_2(\mathfrak{X}^2) \leq \frac{M}{N} \leq \frac{\Delta N + 1}{N} \leq \Delta + \delta$ puisque $N \geq N_\delta$. On a donc montré que la paire (μ, ν) appartenait à $\mathscr{S}_\delta$, et que $\mathscr{F}_N \times \mathscr{F}_N \subset \mathscr{S}_\delta$ pour $N \geq N_\delta$. D'après les questions de la seconde partie, ceci implique le théorème de Strassen.