

Méthodes de Couplage et Applications

Mémoire de Master 2
Université de Bourgogne

Vincent NOLOT

5 juillet 2010

Introduction

Lorsque μ et ν sont deux mesures de probabilité sur des espaces mesurables respectifs (E, \mathcal{E}) , (F, \mathcal{F}) , on aimerait les comparer, les lier. Un moyen de le faire est de considérer des couplages entre μ et ν , c'est-à-dire des mesures sur l'espace produit $(E \times F, \mathcal{E} \otimes \mathcal{F})$ dont les marginales (les mesures images par la projection Π_E resp. Π_F) sont μ resp. ν . L'objet de ce mémoire est d'illustrer cette notion de couplage, et de présenter plusieurs applications dans différents domaines. Notons qu'un couplage entre deux probabilités existe toujours : ne serait-ce que le couplage dit indépendant.

On introduit en Partie I, la notion de distance en variation totale entre deux mesures définies sur le même espace mesurable. Comme son nom l'indique, il s'agit d'une distance sur l'espace des probabilités. Rappelons que cet espace des probabilités sur (E, \mathcal{E}) peut être muni d'une topologie lorsque que E est lui-même topologique, grâce à la convergence faible, caractérisée par le fait qu'une suite $(\mu_n)_n$ converge faiblement vers μ si et seulement si :

$$\forall \varphi \text{ continue bornée sur } E \quad \int_E \varphi d\mu_n \longrightarrow \int_E \varphi d\mu.$$

Mais la convergence selon la distance en variation totale, entraîne la convergence faible. Ainsi cette distance en variation totale induit une topologie plus fine que celle induite par la convergence faible. Nous détaillons le point de vue intrinsèque de la distance en variation totale, mais également le point de vue extrinsèque donné par l'utilisation des couplages. L'utilisation de ces couplages, nous permettra d'obtenir une approximation effective de la variation totale, celle-ci étant généralement difficilement calculable. Cela constitue donc un bon outil pour les statisticiens. Dans cette même partie, nous définissons les p -distances de Wasserstein liées à un coût de transport, qui permettent également de comparer des mesures de probabilités définies sur le même espace et admettant des moments d'ordre p . Nous montrons que cette distance possède de bonnes propriétés dès lors que l'espace E est agréable. C'est pour cette raison que nous nous placerons dans des espaces Polonais. Dans ce cas, l'ensemble des probabilités admettant des moments d'ordre 2, muni de la 2-distance de Wasserstein (W_2) est un espace métrique complet. Nous donnons une caractérisation fondamentale de la convergence selon W_2 . A savoir entre autre, que la convergence selon W_2 entraîne la convergence faible des mesures. La démonstration que l'on propose ici est certainement plus agréable à lire que celle dans [1]. Soulignons que ces deux résultats restent vrais pour tout p entier et pas seulement pour $p = 2$. Enfin ces distances jouent un rôle essentiel en optimisation du coût de transport.

Dans la Partie II, nous nous intéressons à des exemples concrets de couplage dont une marginale est la loi de Poisson, qui fournit une bonne mesure de référence. Nous retrouvons grâce aux méthodes de couplage, des approximations de certaines lois, comme la loi Binomiale vers la Poisson. Les couplages nous permettent de contrôler l'erreur d'approximation, ce que ne nous fournit pas les théorèmes habituels de convergence en loi. Nous présentons également de manière détaillée la méthode dite de Stein, appliquée à la loi de Poisson. Elle utilise un autre point de vue de travail et fournit ses propres approximations. En combinant cette méthode et des méthodes de couplage, nous pouvons améliorer ces résultats d'approximation, notamment grâce à la notion de variables négativement liées. Nous explicitons cela sur divers exemples en fin de Partie, et le plus parlant est certainement le dernier, qui vise à approximer la loi Hypergéométrique par celle de Poisson. Ce résultat est souvent énoncé dans la littérature, mais rarement détaillé comme nous le présentons.

Dans la Partie III, nous introduisons la notion d'entropie d'une probabilité conditionnellement à une autre probabilité de référence. Cela propose une autre façon de lier deux probabilités entre elles, lorsque l'une est absolument continue par rapport à l'autre. L'outil important de cette partie est l'écart d_2 entre deux probabilités. Il généralise la distance de Wasserstein W_0 . Nous faisons alors

le lien entre l'entropie et cet écart. Cela nous permet d'obtenir des inégalités de concentration (dites de grandes déviations) et une inégalité de Poincaré : qui permet de contrôler les variances. Cette inégalité intervient notamment en statistiques et en systèmes dynamiques.

Enfin, le mémoire se termine en Partie IV, par une introduction à la théorie du transport. On adopte alors un nouveau point de vue sur l'espace des probabilités, dans le but d'y mettre une structure géométrique puis pseudo-différentiable. D'abord sur les mesures admettant un moment d'ordre 2, on relie la notion de dérivée d'un chemin géodésique reliant deux mesures à la distance W_2 . Nous verrons que le chemin géodésique correspond à se donner un couplage optimal entre ces deux mesures. Ensuite, cela nous permet d'établir une équation aux dérivées partielles que vérifient les probabilités admettant un moment d'ordre 2 sous des conditions d'absolue continuité. Nous proposons enfin une méthode d'itération sur ces mesures. Nous n'en exposons pas plus, mais cette méthode est un outil puissant pour affirmer l'existence et donne l'approximation de solution de certaines équations, au même titre que l'itération de Picard qui fournit l'existence de solutions dans les équations du type $y' = f(x, y)$. C'est le cas par exemple de l'équation de Fokker-Plank.

Remerciements

Mes remerciements s'adressent principalement à mes deux encadrants de ce mémoire : Patrick Gabriel et Shizan Fang. Patrick Gabriel qui depuis le mois d'octobre 2009 s'est largement investi pour m'orienter dans mon mémoire. Ses conseils et sa rigueur m'ont beaucoup apporté durant cette année, tant au niveau des méthodes mathématiques que de la rédaction. Shizan Fang qui a supervisé le travail et m'a initié à la théorie du transport, s'appuyant sur son cours d'été donné en Chine en 2007. Je les remercie tous deux, de façon très sincère, pour tout ce qu'ils m'ont apporté. Ils peuvent être fiers de m'avoir donné le goût et l'envie de continuer de pratiquer les mathématiques.

Je remercie également Pierre-André Zitt et Christian Bonatti pour avoir accepté de faire partie du jury à l'occasion de ma soutenance. Pierre-André Zitt qui durant cette année, m'a initié aux processus stochastiques et notamment, aux chaînes de Markov et à la modélisation sur ordinateur. Christian Bonatti qui, à travers son cours de recherche Dynamique des groupes, m'a donné une nouvelle vision sur d'autres domaines de la recherche actuelle.

Enfin je remercie Franck Gabriel, qui nous a permis de reprendre ses notes dans le cadre d'un travail sur les couplages, à l'Ecole Normale Supérieure.

Table des matières

I	Concepts de variation totale et de couplage	5
1	Quelques notations	5
2	Variation totale	5
3	Couplages	7
4	Couplage optimal pour un coût donné	8
4.1	Le couplage optimal existe	9
4.2	Construction d'un couplage optimal	10
4.3	Espaces et distances de Wasserstein	12
4.4	Distance W_2	13
II	Couplages avec une loi de Poisson	18
5	Quelques Couplages classiques	18
5.1	Couplages entre deux lois de Poisson $\mathcal{P}(\lambda)$ et $\mathcal{P}(\lambda')$ avec $\lambda > \lambda'$	18
5.2	Couplages d'une loi de Bernoulli avec une loi de Poisson	18
5.3	Couplages avec une somme de lois de Bernoulli	20
5.3.1	Cas où tout est indépendant	20
5.3.2	Cas où il existe de la dépendance	21
6	Méthode de Stein	22
6.1	Présentation	22
6.2	Cas de la loi de Poisson	23
6.3	Quelques applications	30
6.3.1	Un problème de la date d'anniversaire	30
6.3.2	Un problème classique d'occupation	30
6.3.3	Approximation de la loi hypergéométrique	32
III	Entropie conditionnelle	37
7	Liaison entropie et écart d_2	37
7.1	En dimension un	37
7.2	En dimension finie quelconque	40
8	Application : une inégalité de Poincaré	43
IV	Introduction à la théorie du transport	48
9	Mesure dérivée et application à W_2	48
10	Méthode itérative sur des mesures	51
	Appendices	55
A	Théorème de convergence L^p	55

Première partie

Concepts de variation totale et de couplage

1 Quelques notations

- Si μ est une mesure sur l'espace mesurable (E, \mathcal{E}) et $\psi : (E, \mathcal{E}) \rightarrow (F, \mathcal{F})$ une application mesurable, on note $\psi_*\mu$ la mesure sur l'espace mesurable (F, \mathcal{F}) , image de μ par ψ . C'est-à-dire : $\psi_*\mu = \mu \circ \psi^{-1}$. Elle est caractérisée par le fait que pour toute fonction numérique réelle borélienne positive, φ définie sur (F, \mathcal{F}) , on a :

$$\int_F \varphi d\psi_*\mu = \int_E \varphi \circ \psi d\mu$$

- Si μ et λ sont deux mesures sur l'espace mesurable (E, \mathcal{E}) et f une fonction borélienne positive définie sur (E, \mathcal{E}) , on dit que μ est à densité f relativement à λ ce que l'on note : $\mu = f \cdot \lambda$, si pour toute fonction numérique réelle borélienne positive, φ sur (E, \mathcal{E}) , on a :

$$\int_E \varphi d\mu = \int_E \varphi d(f \cdot \lambda) = \int_E \varphi f d\lambda.$$

- Si μ est une mesure sur l'espace mesurable (E, \mathcal{E}) et $\varphi : E \rightarrow \mathbb{R}^d$ une fonction borélienne, on note :

$$\|\varphi\|_{p,\mu} := \left(\int_E |\varphi|^p d\mu \right)^{\frac{1}{p}},$$

$$\|\varphi\|_{0,\mu} := \int_E |\varphi|^0 d\mu = \int_E \mathbb{1}_{\{\varphi \neq 0\}} d\mu = \mu[\varphi \neq 0].$$

où : $|v| = (v_1^2 + \dots + v_d^2)^{\frac{1}{2}}$ désigne la norme euclidienne du vecteur $v = (v_i)_{i=1,\dots,d} \in \mathbb{R}^d$

- Si φ est une fonction borélienne bornée sur l'espace mesurable (E, \mathcal{E}) , on note par $\|\varphi\|_u$ la norme uniforme de φ :

$$\|\varphi\|_u := \sup_{x \in E} |\varphi(x)|.$$

- On notera $\Delta_E := \{(x, x); x \in E\}$ la diagonale de E .

2 Variation totale

Pour comparer deux mesures de probabilité définies sur un même espace mesurable (E, \mathcal{E}) , un premier moyen est de considérer la distance en variation totale :

Définition 2.1 Pour μ et ν deux mesures sur (E, \mathcal{E}) . On définit la variation totale entre μ et ν par :

$$\|\mu - \nu\|_{VT} = \sup_{A \in \mathcal{E}} |\mu(A) - \nu(A)|.$$

La variation totale est une distance sur l'espace des probabilités sur (E, \mathcal{E}) . Elle peut se décrire d'un point de vue intrinsèque de plusieurs manières équivalentes.

LEMME 2.1 Si μ et ν sont deux mesures de probabilité sur (E, \mathcal{E}) , toutes deux absolument continues

par rapport à une même mesure λ avec densités respectives : $f := \frac{d\mu}{d\lambda}$ et $g := \frac{d\nu}{d\lambda}$, alors :

$$\|\mu - \nu\|_{VT} = \sup_{0 \leq \varphi \leq 1} \left| \int \varphi d\mu - \int \varphi d\nu \right| \quad (2.1)$$

$$= \frac{1}{2} \sup_{\|\varphi\|_u \leq 1} \left| \int \varphi d\mu - \int \varphi d\nu \right| \quad (2.2)$$

$$= \frac{1}{2} \int |f - g| d\lambda \quad (2.3)$$

$$= \int (f - g)^+ d\lambda. \quad (2.4)$$

Remarque : Il existe toujours une mesure de probabilité sur (E, \mathcal{E}) relativement à laquelle μ et ν sont toutes deux absolument continues : $\frac{1}{2}\mu + \frac{1}{2}\nu$ en est un exemple. Notons par ailleurs que les quantités intervenant dans (2.3) et (2.4) ne dépendent pas du choix de λ . En effet donnons nous λ_1 et λ_2 deux mesures de référence pour lesquelles : $\mu = f_1 \cdot \lambda_1 = f_2 \cdot \lambda_2$ et $\nu = g_1 \cdot \lambda_1 = g_2 \cdot \lambda_2$. Si $\lambda_1 \ll \lambda_2$ avec $\lambda_1 = h \cdot \lambda_2$, alors on a les égalités presque-sûres au sens de λ_2 : $f_1 h = f_2$ et $g_1 h = g_2$, avec ce faisant :

$$\begin{aligned} \int |f_1 - g_1| d\lambda_1 &= \int |f_1 - g_1| h d\lambda_2 = \int |f_2 - g_2| d\lambda_2 \\ \int (f_1 - g_1)^+ d\lambda_1 &= \int (f_1 - g_1)^+ h d\lambda_2 = \int (f_2 - g_2)^+ d\lambda_2 \end{aligned}$$

Dans le cas général, puisqu'il existe toujours une mesure de référence $\tilde{\lambda}$ par rapport à laquelle λ_1 et λ_2 sont absolument continues, et puisque deux quantités égales à une même troisième sont égales entre elles, on déduit du cas précédent, que les quantités considérées ne dépendent pas de la mesure de référence.

Démonstration :

Démontrons maintenant la suite d'égalités annoncées.

* Vu que $(f - g)^+ = \frac{1}{2}(|f - g| + f - g)$ et comme $\int f d\lambda = \int g d\lambda = 1$, on a :

$$\int (f - g)^+ d\lambda = \frac{1}{2} \int |f - g| d\lambda = \int (g - f)^+ d\lambda.$$

Ce qui montre l'égalité des quantités intervenant dans (2.3) et (2.4).

* Pour $A \in \mathcal{E}$, en intégrant les inégalités fonctionnelles : $(f - g)\mathbb{1}_A \leq (f - g)^+\mathbb{1}_A \leq (f - g)^+$, on a :

$$\mu(A) - \nu(A) = \int_A (f - g) d\lambda \leq \int (f - g)^+ d\lambda$$

puis par symétrie :

$$\nu(A) - \mu(A) \leq \int (g - f)^+ d\lambda = \int (f - g)^+ d\lambda$$

et donc en passant au $\sup_{A \in \mathcal{E}}$:

$$\|\mu - \nu\|_{VT} = \sup_{A \in \mathcal{E}} |\mu(A) - \nu(A)| \leq \int (f - g)^+ d\lambda.$$

Par ailleurs, en considérant : $E^+ := \{x \in E, f(x) \geq g(x)\} \in \mathcal{E}$, on a :

$$\int (f - g)^+ d\lambda = \int_{E^+} (f - g) d\lambda = \mu(E^+) - \nu(E^+) \leq \|\mu - \nu\|_{VT}.$$

C'est donc que : $\|\mu - \nu\|_{VT} = \int (f - g)^+ d\lambda$, à savoir l'identification de la variation totale comme valant (2.4).

* Pour φ , \mathcal{E} -mesurable telle que $\|\varphi\|_u \leq 1$, considérons $\tilde{\varphi} := \frac{\varphi+1}{2}$ qui est \mathcal{E} -mesurable, qui vérifie $0 \leq \tilde{\varphi} \leq 1$ et pour laquelle on a :

$$\begin{aligned} \left| \int \tilde{\varphi} d\mu - \int \tilde{\varphi} d\nu \right| &= \left| \int \tilde{\varphi}(f - g) d\lambda \right| \\ &= \frac{1}{2} \left| \int \varphi(f - g) d\lambda + \int f d\lambda - \int g d\lambda \right| \\ &= \frac{1}{2} \left| \int \varphi d\mu - \int \varphi d\nu \right|. \end{aligned}$$

L'application $x \mapsto \frac{x+1}{2}$ établissant une bijection entre $[-1, 1]$ et $[0, 1]$, l'application $\varphi \mapsto \tilde{\varphi}$ échange fonctions mesurables bornées par 1 et fonctions mesurables à valeurs dans $[0, 1]$, de sorte que :

$$\sup_{0 \leq \varphi \leq 1} \left| \int \varphi d\mu - \int \varphi d\nu \right| = \frac{1}{2} \sup_{\|\varphi\|_u \leq 1} \left| \int \varphi d\mu - \int \varphi d\nu \right|.$$

Cela montre l'égalité des expressions (2.1) et (2.2).

* Soit φ une fonction \mathcal{E} -mesurable telle que $\|\varphi\|_u \leq 1$, alors :

$$\left| \int \varphi d\mu - \int \varphi d\nu \right| \leq \int |\varphi| |f - g| d\lambda \leq \int |f - g| d\lambda$$

et donc :

$$\sup_{\|\varphi\|_u \leq 1} \left| \int \varphi d\mu - \int \varphi d\nu \right| \leq \int |f - g| d\lambda.$$

On a de plus :

$$\begin{aligned} \|\mu - \nu\|_{VT} &= \sup_{\varphi = \mathbb{1}_A, A \in \mathcal{A}} \left| \int \varphi d\mu - \int \varphi d\nu \right| \\ &\leq \sup_{0 \leq \varphi \leq 1} \left| \int \varphi d\mu - \int \varphi d\nu \right|. \end{aligned}$$

Ce qui montre l'identification de la variation totale comme valant (2.1) ■

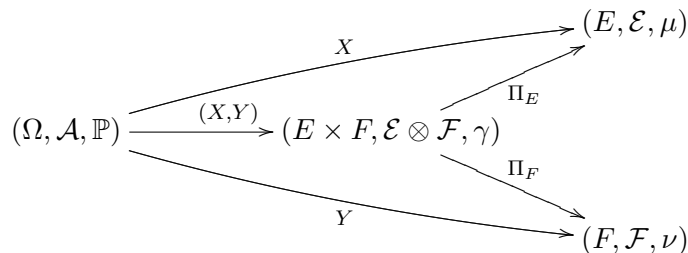
3 Couplages

Etant donnés deux espaces probabilisés (E, \mathcal{E}, μ) et (F, \mathcal{F}, ν) , une façon de comparer ces structures pour un certain point de vue, quantifié par ce que nous appellerons une fonction de coût, est de les faire apparaître comme sous-structures, d'un même espace probabilisé, en essayant d'en minimiser pour cette représentation, le coût moyen.

Pour des raisons techniques de mesurabilité, on supposera en général que E et F sont des espaces *Polonais* (i.e. espace topologique métrisable séparable dont la topologie peut être définie par une distance qui en fait un espace complet), dont on note \mathcal{E} et \mathcal{F} les tribus boréliennes. Etant données deux mesures de probabilité : μ sur (E, \mathcal{E}) et ν sur (F, \mathcal{F}) , introduisons la notion de *couplage entre μ et ν* .

Définition 3.1 *Un couplage entre μ et ν est une mesure γ sur l'espace produit $(E \times F, \mathcal{E} \otimes \mathcal{F})$ qui a pour mesures marginales $\Pi_{E^*} \gamma = \mu$ et $\Pi_{F^*} \gamma = \nu$.*

Remarque : D'un point de vue probabiliste, trouver un *couplage* γ entre μ et ν consiste à trouver un couple (X, Y) de variables aléatoires, toutes deux définies sur un même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$, à valeurs dans respectivement (E, \mathcal{E}) et (F, \mathcal{F}) , et dont les lois respectives sont : $\mathcal{L}(X) = \mu$ et $\mathcal{L}(Y) = \nu$, puis à considérer pour γ la loi du couple (X, Y) . On dira dans ce cas que le couple (X, Y) réalise le couplage γ .



Remarque : Il n'y a pas en général unicité du couplage. En voici deux exemples :

Exemple 3.1 *Le couplage indépendant est le produit tensoriel $\mu \otimes \nu$, caractérisé par le fait que pour tout rectangle $A \times B$, avec $(A, B) \in \mathcal{E} \times \mathcal{F}$, on a : $(\mu \otimes \nu)(A \times B) = \mu(A) \cdot \nu(B)$.*

Le couplage porté par un graphe : dans le cas où $\nu = \psi_ \mu$ avec $\psi : (E, \mathcal{E}) \rightarrow (F, \mathcal{F})$, une application mesurable, la mesure $\mu^\psi := (id_E, \psi)_* \mu$, sur $(E \times F, \mathcal{E} \otimes \mathcal{F})$ est un couplage entre μ et ν . Il est caractérisé par le fait que pour tout rectangle $(A, B) \in \mathcal{E} \times \mathcal{F}$, on a : $\mu^\psi(A \times B) = \mu[A \cap \psi^{-1}(B)]$. Il est porté par $(\psi \circ \Pi_E, \Pi_F)^{-1}(\Delta_F)$, le graphe de ψ .*

Dans le cas où E et F sont discrets, les couplages $\mu \otimes \nu$ et μ^ψ sont caractérisés par :

$$\forall (i, j) \in E \times F, \quad (\mu \otimes \nu)_{i,j} = \mu_i \nu_j, \text{ et } : \mu_{i,j}^\psi = \begin{cases} \mu_i & \text{si } j = \psi(i) \\ 0 & \text{sinon} \end{cases}$$

Notations : Pour μ et ν deux mesures définies sur un même espace Polonais (E, \mathcal{E}) , $\mathcal{M}(\mu, \nu)$ désigne l'ensemble des couplages entre μ et ν .

La relative compacité de $\mathcal{M}(\mu, \nu)$, pour la convergence étroite des mesures de probabilité sera bien souvent un argument clé :

LEMME 3.1 *Dans le cas où μ et ν sont des mesures de probabilité, $\mathcal{M}(\mu, \nu)$ est un ensemble tendu de probabilités.*

Démonstration :

Comme toute mesure de probabilité sur un espace polonais muni de sa tribu borélienne est tendue, pour tout $\epsilon > 0$ fixé, il existe deux compacts K_1, K_2 de l'espace polonais E tels que :

$$\mu(K_1) > 1 - \frac{\epsilon}{2}, \quad \nu(K_2) > 1 - \frac{\epsilon}{2}.$$

Ainsi $K = K_1 \cup K_2$ est un compact de E pour lequel $\mu(K^c) + \nu(K^c) < \frac{\epsilon}{2} + \frac{\epsilon}{2}$. Par ailleurs, $K \times K$ est un compact de $E \times E$, qui vérifie : $(K \times K)^c \subset (K^c \times E) \cup (E \times K^c)$ de sorte que pour tout $\gamma \in \mathcal{M}(\mu, \nu)$, on a :

$$\gamma\left((K \times K)^c\right) \leq \gamma(K^c \times E) + \gamma(E \times K^c) = \mu(K^c) + \nu(K^c) < \epsilon.$$

On en déduit le caractère tendu de $\mathcal{M}(\mu, \nu)$ ■

4 Couplage optimal pour un coût donné

On se place sur (E, \mathcal{E}) un espace Polonais mesurable.

Définition 4.1 *On appellera fonction de coût positive toute application $c : E \rightarrow \mathbb{R}_+$ \mathcal{E} -mesurable.*

Toutes les fonctions de coût que nous envisagerons seront au moins semi-continues inférieurement. La notion de couplage optimal, est relative à une fonction de coût donnée.

Définition 4.2 Soit c une fonction de coût positive. On dira que $\gamma \in \mathcal{M}(\mu, \nu)$ est un couplage optimal entre μ et ν relativement à c si :

$$\int c(x, y) d\gamma(x, y) = \inf_{\Pi \in \mathcal{M}(\mu, \nu)} \int c(x, y) d\Pi(x, y).$$

Remarque :

- Si $c : (x, y) \mapsto \mathbb{1}_{\Delta_E^c}(x, y)$. Comme Δ_E^c est ouvert, la fonction de coût c est *semi-continue inférieurement*. Pour cette fonction de coût, on parlera de *couplage optimal*. Dans ce cas si γ est un couplage optimal entre μ et ν alors :

$$\gamma(\{x \neq y\}) = \inf_{\Pi \in \mathcal{M}(\mu, \nu)} \Pi(\{x \neq y\}).$$

- On aura affaire également aux fonctions de coût de type $c : (x, y) \mapsto |x - y|^p$ sur $\mathbb{R}^d \times \mathbb{R}^d$, où p est un entier plus grand que 1. Ces fonctions de coût sont alors continues et on parlera de *couplage optimal relativement à c* .

4.1 Le couplage optimal existe

Une question naturelle se pose : pour deux mesures données et une fonction de coût fixée, existe-t-il toujours un couplage optimal ? La réponse est oui lorsque la fonction de coût est suffisamment régulière.

PROPOSITION 4.1 Si μ et ν sont deux probabilités sur E et $c : E \times E \rightarrow \mathbb{R}^+$ est une fonction de coût, semi-continue inférieurement et bornée, alors la fonction :

$$\gamma \in \mathcal{M}(\mu, \nu) \mapsto \int_{E \times E} c(x, y) d\gamma(x, y)$$

atteint son minimum sur $\mathcal{M}(\mu, \nu)$.

Démonstration :

Notons $m = \inf_{\gamma \in \mathcal{M}(\mu, \nu)} \left\{ \int_{E \times E} c(x, y) d\gamma(x, y) \right\}$. Comme c est bornée, on a $m < +\infty$. D'après la caractérisation de la borne inférieure, pour chaque $n \in \mathbb{N}^*$, il existe un couplage $\gamma_n \in \mathcal{M}(\mu, \nu)$ tel que :

$$\int_{E \times E} c(x, y) d\gamma_n(x, y) \leq m + \frac{1}{n}. \quad (4.5)$$

Le lemme (3.1) nous indique que la suite de couplages $(\gamma_n)_{n \geq 1}$ est tendue. Il en existe une sous-suite $(\gamma_{n_k})_k$ qui converge étroitement vers une probabilité $\gamma \in P(E \times E)$. Montrons que γ est un couplage optimal relativement à c .

* Soit alors $\varphi \in \mathcal{C}_b(E \times E)$. La convergence étroite se traduit par :

$$\int_{E \times E} \varphi d\gamma_{n_k} \longrightarrow \int_{E \times E} \varphi d\gamma.$$

Ainsi pour $\psi \in \mathcal{C}_b(E)$, si l'on pose $\varphi = \psi \circ \pi_1 \in \mathcal{C}_b(E \times E)$, et comme $\gamma_{n_k} \in \mathcal{M}(\mu, \nu)$ on a :

$$\int_E \psi d\mu = \int_{E \times E} \varphi d\gamma_{n_k} \longrightarrow \int_{E \times E} \varphi d\gamma = \int_E \psi d\gamma.$$

Cela signifie que $(\pi_1)_*\gamma = \mu$. De la même façon, on montre que $(\pi_2)_*\gamma = \nu$. Autrement dit on a déjà que $\gamma \in \mathcal{M}(\mu, \nu)$.

* Grâce à la représentation de Skorohod, il existe un espace $(\Omega, \mathcal{F}, \mathbb{P})$ probabilisé, des variables aléatoires $(X_k)_k$, X sur cet espace telles que :

$$\begin{aligned} \forall k \geq 1 \quad \mathcal{L}(X_k) &= \gamma_{n_k}, & \mathcal{L}(X) &= \gamma \\ (X_k)_k &\longrightarrow X & \text{p.s.} \end{aligned}$$

Par les lemmes du Transfert et de Fatou on a donc :

$$\liminf_k \int_{E \times E} c(x, y) d\gamma_{n_k}(x, y) = \liminf_k \mathbb{E}[c(X_k)] \geq \mathbb{E}[\liminf_k c(X_k)],$$

et la fonction $c : E \times E \longrightarrow \mathbb{R}_+$ étant semi-continue inférieurement, on a :

$$\liminf_k \int_{E \times E} c(x, y) d\gamma_{n_k}(x, y) \geq \int_{E \times E} c(x, y) d\gamma(x, y).$$

Et grâce à (4.5), en faisant tendre $k \longrightarrow +\infty$ on obtient :

$$\int_{E \times E} c(x, y) d\gamma(x, y) \leq m.$$

Finalement grâce à la définition de m , on obtient que γ est un couplage optimal relativement à c ■

Remarque : La proposition (4.1) reste vraie si on ne suppose plus que c est bornée, mais qu'il existe un couplage $\rho \in \mathcal{M}(\mu, \nu)$ telle que $\int_{E \times E} c(x, y) d\rho(x, y) < +\infty$.

4.2 Construction d'un couplage optimal

On donne ici un moyen explicite de construction d'un couplage optimal pour le coût $(x, y) \mapsto \mathbb{1}_{\Delta_E^c}(x, y)$.

PROPOSITION 4.2 *Si μ et ν sont deux mesures de probabilité sur (E, \mathcal{E}) à densités respectives f et g par rapport à une même mesure ρ , la mesure de probabilité sur $(E \times E, \mathcal{E} \otimes \mathcal{E})$, donné par :*

$$\tilde{\gamma} = i_{\Delta_*} \left(\min(f, g) \cdot \rho \right) + \frac{(f - g)^+ \otimes (g - f)^+}{\|\mu - \nu\|_{VT}} \cdot (\rho \otimes \rho)$$

où i_{Δ} est l'application $E \rightarrow E \times E$ qui $x \mapsto (x, x)$, est un couplage optimal entre μ et ν .

Remarque : La mesure $\tilde{\gamma}$ décrite ci-dessus, est caractérisée par la façon dont elle intègre les fonctions boréliennes positives sur $E \times E$. Pour $\psi : E \times E \rightarrow \mathbb{R}_+$, fonction borélienne positive, on a :

$$\begin{aligned} \iint_{E \times E} \psi(x, y) d\tilde{\gamma}(x, y) &:= \int_E \psi(x, x) \min\{f(x), g(x)\} d\rho(x) \\ &+ \frac{2}{\int_E |f(x) - g(x)| d\rho(x)} \iint_{E \times E} \psi(x, y) (f(x) - g(x))^+ (g(y) - f(y))^+ d\rho(x) d\rho(y) \end{aligned}$$

Dans le cas discret, où E est l'ensemble dénombrable discret I , avec les mesures μ et ν données par l'intermédiaire des vecteurs de probabilités $(\mu_i)_{i \in I}$ et $(\nu_i)_{i \in I}$, correspondant aux densités respectives de μ et ν relativement à la mesure de décompte sur I le couplage optimal $\tilde{\gamma}$ est donné par l'intermédiaire de sa densité relativement à la mesure de décompte sur $I \times I$: $(\tilde{\gamma}_{i,j})_{(i,j) \in I \times I}$ sous la forme :

$$\tilde{\gamma}_{i,j} = \min(\mu_i, \nu_j) \delta_{ij} + \frac{(\mu_i - \nu_i)^+ (\nu_j - \mu_j)^+}{\|\mu - \nu\|_{VT}}.$$

Démonstration :

Le fait que $\tilde{\gamma}$ soit un couplage entre μ et ν est de vérification immédiate. Pour toute fonction positive, borélienne sur E , φ , on a :

$$\begin{aligned} \int \varphi d(\pi_{1*}\tilde{\gamma}) &= \int \varphi \min\{f, g\} d\rho + 2 \frac{\int (g-f)^+ d\rho}{\int |f-g| d\rho} \int \varphi (f-g)^+ d\rho \\ &= \int \varphi \left(\min\{f, g\} + (f-g)^+ \right) d\rho = \int \varphi f d\rho = \int \varphi d\mu \end{aligned}$$

$$\begin{aligned} \int \varphi d(\pi_{2*}\tilde{\gamma}) &= \int \varphi \min\{f, g\} d\rho + 2 \frac{\int (f-g)^+ d\rho}{\int |f-g| d\rho} \int \varphi (g-f)^+ d\rho \\ &= \int \varphi \left(\min\{f, g\} + (g-f)^+ \right) d\rho = \int \varphi g d\rho = \int \varphi d\nu \end{aligned}$$

de sorte que : $\tilde{\gamma} \in \mathcal{M}(\mu, \nu)$. Par ailleurs, vu que $(f-g)^+(g-f)^+ = 0$, on a :

$$\tilde{\gamma}(\Delta_E^c) = 1 - \tilde{\gamma}(\Delta_E) = 1 - \int \min\{f, g\} d\rho = 1 - \int f d\rho + \int (f-g)^+ d\rho = \int_E (f-g)^+ d\rho = \|\mu - \nu\|_{VT}.$$

Enfin pour tout $\gamma \in \mathcal{M}(\mu, \nu)$, couplage entre les deux mesures μ et ν sur l'espace probabilisé (E, \mathcal{E}) , et pour toute fonction borélienne $\varphi : E \rightarrow [0, 1]$, on a :

$$\begin{aligned} \left| \int \varphi d\mu - \int \varphi d\nu \right| &= \left| \iint \varphi(x) d\gamma(x, y) - \iint \varphi(y) d\gamma(x, y) \right| \\ &= \left| \iint (\varphi(x) - \varphi(y)) d\gamma(x, y) \right| \\ &= \left| \iint_{\Delta_E^c} \varphi(x) - \varphi(y) d\gamma(x, y) \right| \leq \iint_{\Delta_E^c} |\varphi(x) - \varphi(y)| d\gamma(x, y) \leq \gamma(\Delta_E^c). \end{aligned}$$

Ceci étant vrai pour toute fonction borélienne $0 \leq \varphi \leq 1$, on en déduit que :

$$\|\mu - \nu\|_{VT} = \sup_{0 \leq \varphi \leq 1} \left| \int \varphi d\mu - \int \varphi d\nu \right| \leq \gamma(\Delta_E^c).$$

Et donc que :

$$\sup_{0 \leq \varphi \leq 1} \left| \int \varphi d\mu - \int \varphi d\nu \right| = \|\mu - \nu\|_{VT} = \tilde{\gamma}(\Delta_E^c) = \inf_{\gamma \in \mathcal{M}(\mu, \nu)} \gamma(\Delta_E^c).$$

D'où le fait que $\tilde{\gamma}$ soit un couplage entre μ et ν , optimal pour le coût : $(x, y) \mapsto \mathbb{1}_{\Delta_E^c}(x, y)$ ■

Remarque : Lors de cette proposition, nous avons obtenu une nouvelle caractérisation, cette fois extrinsèque, de la *variation totale* :

$$\|\mu - \nu\|_{VT} = \inf_{\gamma \in \mathcal{M}(\mu, \nu)} \left(\int \mathbb{1}_{\{x \neq y\}} d\gamma(x, y) \right) = \inf_{\gamma \in \mathcal{M}(\mu, \nu)} \gamma(\Delta_E^c).$$

Remarque : D'un point de vue probabiliste, notons que si un couple de variables aléatoires (X, Y) sur $(\Omega, \mathcal{A}, \mathbb{P})$, réalise un couplage $\gamma \in \mathcal{M}(\mu, \nu)$, on a :

$$\mathbb{P}(X \neq Y) = \gamma(\Delta_E^c).$$

Pour que le couple (X, Y) réalise un couplage optimal $\tilde{\gamma}$, il faut que :

$$\mathbb{P}(X \neq Y) = \tilde{\gamma}(\Delta_E^c) = \|\mu - \nu\|_{VT} = \int (f - g)^+ d\rho$$

et donc que :

$$\mathbb{P}(X = Y) = \tilde{\gamma}(\Delta_E) = 1 - \|\mu - \nu\|_{VT} = \int f - (f - g)^+ d\rho = \int \min(f, g) d\rho$$

Cette condition se trouve réalisée si l'on suppose que pour $A, B \in \mathcal{E}$:

$$\mathbb{P}(X \in A, Y \in B, X = Y) = \mathbb{P}(X \in A \cap B, X = Y) = \mathbb{P}(Y \in A \cap B, X = Y) = \int_{A \cap B} \min(f, g) d\rho$$

avec en particulier :

$$\mathbb{P}(X \in A, X \neq Y) = \mathbb{P}(X \in A) - \mathbb{P}(X \in A, X = Y) = \int_A f - \min(f, g) d\rho = \int_A (f - g)^+ d\rho$$

$$\mathbb{P}(Y \in B, X \neq Y) = \mathbb{P}(Y \in B) - \mathbb{P}(Y \in B, X = Y) = \int_B g - \min(f, g) d\rho = \int_B (g - f)^+ d\rho$$

et que par ailleurs, X et Y sont indépendants sachant $\{X \neq Y\}$, à savoir que pour $A, B \in \mathcal{E}$:

$$\mathbb{P}(X \in A, Y \in B \mid X \neq Y) = \mathbb{P}(X \in A \mid X \neq Y) \cdot \mathbb{P}(Y \in B \mid X \neq Y)$$

$$\begin{aligned} \mathbb{P}(X \in A, Y \in B, X \neq Y) &= \frac{1}{\mathbb{P}(X \neq Y)} \mathbb{P}(X \in A, Y \neq X) \cdot \mathbb{P}(Y \in B, X \neq Y) \\ &= \frac{1}{\|\mu - \nu\|_{VT}} \left(\int_A (f(x) - g(x))^+ d\rho(x) \right) \left(\int_B (g(y) - f(y))^+ d\rho(y) \right) \\ &= \frac{1}{\|\mu - \nu\|_{VT}} \left(\iint_{A \times B} (f(x) - g(x))^+ (g(y) - f(y))^+ d(\rho \otimes \rho)(x, y) \right). \end{aligned}$$

ce qui en bilan donne la condition :

$$\mathbb{P}(X \in A, Y \in B) = \int_{A \cap B} \min(f, g) d\rho + \frac{1}{\|\mu - \nu\|_{VT}} \left(\iint_{A \times B} (f - g)^+ \otimes (g - f)^+ d(\rho \otimes \rho) \right).$$

On retrouve que la loi du couple (X, Y) est le couplage $\tilde{\gamma}$ donné dans la proposition.

4.3 Espaces et distances de Wasserstein

Notations : Dans toute la suite, on désigne par $P(\mathbb{R}^d)$ l'ensemble des probabilités sur \mathbb{R}^d .

On appelle *p-espace de Wasserstein* (pour $p \in \mathbb{N}^*$) l'espace des probabilités, défini comme suit :

$$P_p(\mathbb{R}^d) := \left\{ \mu \in P(\mathbb{R}^d); m_p(\mu) := \int_{\mathbb{R}^d} |x|^p d\mu(x) < +\infty \right\}$$

muni de la *p-distance de Wasserstein* associée W_p , telle que pour $\mu, \nu \in P_p(\mathbb{R}^d)$:

$$W_p(\mu, \nu) := \inf_{\gamma \in \mathcal{M}(\mu, \nu)} \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p d\gamma(x, y) \right\}^{\frac{1}{p}} = \inf_{\gamma \in \mathcal{M}(\mu, \nu)} \{ \|\pi_1 - \pi_2\|_{p, \gamma} \},$$

où on rappelle que $\mathcal{M}(\mu, \nu) = \{ \gamma \in P(\mathbb{R}^d \times \mathbb{R}^d); (\pi_1)_* \gamma = \mu, (\pi_2)_* \gamma = \nu \}$, avec $\pi_i : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ est la projection sur la i^{eme} composante.

Remarque : W_0 est la distance correspondant au couplage optimal défini plus haut :

$$W_0(\mu, \nu) = \inf_{\gamma \in \mathcal{M}(\mu, \nu)} \gamma[\pi_1 \neq \pi_2] = \inf_{\gamma \in \mathcal{M}(\mu, \nu)} \gamma(\Delta_{\mathbb{R}^d}^c).$$

Nous nous intéresserons ici plus spécifiquement à l'espace de Wasserstein $(P_2(\mathbb{R}^d), W_2)$.

4.4 Distance W_2

Remarquons en premier lieu que W_2 est toujours finie. En effet, pour $x, y \in \mathbb{R}^d$ on a :

$$|x + y|^2 = \sum_{i=1}^d (x_i + y_i)^2 \leq 2 \sum_{i=1}^d x_i^2 + 2 \sum_{i=1}^d y_i^2 = 2|x|^2 + 2|y|^2.$$

Ainsi pour $\mu, \nu \in P_2(\mathbb{R}^d)$, et $\gamma \in \mathcal{M}(\mu, \nu)$, il vient que :

$$\begin{aligned} W_2^2(\mu, \nu) &\leq \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\gamma(x, y) \leq 2 \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x|^2 d\gamma(x, y) + \int_{\mathbb{R}^d \times \mathbb{R}^d} |y|^2 d\gamma(x, y) \right) \\ &= 2(m_2(\mu) + m_2(\nu)) < +\infty. \end{aligned}$$

Un cas particulier de la proposition 4.1, est la proposition suivante :

PROPOSITION 4.3 *Pour $\mu, \nu \in P_2(\mathbb{R}^d)$, il existe un couplage $\gamma_0 \in \mathcal{M}(\mu, \nu)$ tel que :*

$$W_2^2(\mu, \nu) = \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\gamma_0(x, y). \quad (4.6)$$

Un tel couplage est dit optimal entre μ et ν pour le coût quadratique : $(x, y) \mapsto |x - y|^2$.

Notations : On désignera l'ensemble des *couplages optimaux pour W_2* entre μ et ν par

$$\mathcal{M}_0^2(\mu, \nu) = \{\gamma \in \mathcal{M}(\mu, \nu); \text{ qui vérifie (4.6)}\}.$$

Justifions maintenant la terminologie de *distance* pour W_2 .

PROPOSITION 4.4 *W_2 est bien une distance sur $P_2(\mathbb{R}^d)$.*

Démonstration :

On vérifie les trois axiomes de la définition de distance.

i) Considérons l'application $i_\Delta : \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{R}^d$ qui : $x \mapsto (x, x)$. Soit $\gamma := i_{\Delta*}\mu$. Vu que : $\pi_1 \circ i_\Delta = id_E = \pi_2 \circ i_\Delta$, on a : $\gamma \in \mathcal{M}(\mu, \mu)$ et :

$$W_2^2(\mu, \mu) \leq \int_{\mathbb{R}^d \times \mathbb{R}^d} |\pi_1 - \pi_2|^2 d\gamma = \int_{\mathbb{R}^d} |\pi_1 \circ i_\Delta - \pi_2 \circ i_\Delta|^2 d\mu = 0.$$

Si maintenant μ et ν sont telles que $W_2(\mu, \nu) = 0$. On peut, d'après la proposition précédente, considérer un couplage optimal $\gamma \in \mathcal{M}_0^2(\mu, \nu)$, pour lequel on a :

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} |\pi_1 - \pi_2|^2 d\gamma = 0.$$

C'est donc que $\pi_1 = \pi_2$, γ -p.p, à savoir que γ est portée par la diagonale $\Delta_{\mathbb{R}^d}$. Ainsi pour toute fonction φ borélienne bornée définie sur \mathbb{R}^d , on a :

$$\int_{\mathbb{R}^d} \varphi d\mu = \int_{\mathbb{R}^d \times \mathbb{R}^d} \varphi \circ \pi_1 d\gamma = \int_{\mathbb{R}^d \times \mathbb{R}^d} \varphi \circ \pi_2 d\gamma = \int_{\mathbb{R}^d} \varphi d\nu.$$

On en déduit que $\mu = \nu$.

ii) Considérons l'application $S : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ qui : $(x, y) \mapsto (y, x)$. Soit $\gamma \in \mathcal{M}_0^2(\mu, \nu)$ et $\hat{\gamma} := S_*\gamma$.

Vu que $\pi_1 \circ S = \pi_2$, on a : $\pi_{1*}\hat{\gamma} = \pi_{1*}(S_*\gamma) = (\pi_1 \circ S)_*\gamma = \pi_{2*}\gamma = \nu$. On obtient de même $(\pi_2)_*\hat{\gamma} = \mu$, de sorte que $\hat{\gamma} \in \mathcal{M}(\nu, \mu)$. Comme par ailleurs, γ est un couplage optimal, on a :

$$W_2^2(\nu, \mu) \leq \int_{\mathbb{R}^d \times \mathbb{R}^d} |\pi_1 - \pi_2|^2 d\hat{\gamma} = \int_{\mathbb{R}^d \times \mathbb{R}^d} |\pi_2 - \pi_1|^2 d\gamma = W_2^2(\mu, \nu).$$

D'où en échangeant les rôles de μ et ν , la symétrie de W_2 .

iii) Soient $\mu_1, \mu_2, \mu_3 \in P_2(\mathbb{R}^d)$, $\gamma_1 \in \mathcal{M}_0^2(\mu_1, \mu_2)$ et $\gamma_2 \in \mathcal{M}_0^2(\mu_2, \mu_3)$. Les marginales $(\pi_2)_*\gamma_1 = \mu_2 = (\pi_1)_*\gamma_2$ sont identiques et si on note $\hat{\pi}_i : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ la i^{ieme} projection, il existe une probabilité $\lambda \in P_2(\mathbb{R}^2 \times \mathbb{R}^d \times \mathbb{R}^d)$ telle que :

$$(\hat{\pi}_1, \hat{\pi}_2)_*\lambda = \gamma_1 \quad \text{et} \quad (\hat{\pi}_2, \hat{\pi}_3)_*\lambda = \gamma_2.$$

De plus on a les égalités :

$$\begin{aligned} (\hat{\pi}_1)_*\lambda &= (\pi_1)_*\gamma_1 = (\pi_1)_*(\hat{\pi}_1, \hat{\pi}_3)_*\lambda = \mu_1, \\ (\hat{\pi}_3)_*\lambda &= (\pi_2)_*\gamma_2 = (\pi_2)_*(\hat{\pi}_1, \hat{\pi}_3)_*\lambda = \mu_3, \end{aligned}$$

qui impliquent que $(\hat{\pi}_1, \hat{\pi}_3)_*\lambda \in \mathcal{M}(\mu_1, \mu_3)$. Si on note $\|\cdot\|_{L^2(\lambda)}$ la norme habituelle sur $L^2(\lambda)$, on a, par successions d'inégalités triangulaires :

$$\begin{aligned} W_2(\mu_1, \mu_3) &\leq \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x_1 - x_3|^2 d((\hat{\pi}_1, \hat{\pi}_3)_*\lambda)(x_1, x_3) \right)^{\frac{1}{2}} = \| \hat{\pi}_1 - \hat{\pi}_3 \|_{L^2(\lambda)} \\ &\leq \| \hat{\pi}_1 - \hat{\pi}_2 \|_{L^2(\lambda)} + \| \hat{\pi}_2 - \hat{\pi}_3 \|_{L^2(\lambda)} = W_2(\mu_1, \mu_2) + W_2(\mu_2, \mu_3). \end{aligned}$$

Donc W_2 vérifie bien l'inégalité triangulaire ■

Donnons à présent une caractérisation de la convergence dans $P_2(\mathbb{R}^d)$ au sens de W_2 .

PROPOSITION 4.5 Soient $(\mu_n)_n, \mu \in P_2(\mathbb{R}^d)$. On a équivalence entre :

1. $(\mu_n)_n$ converge vers μ au sens de W_2
2. $(\mu_n)_n$ converge faiblement vers μ et

$$\lim_{R \rightarrow +\infty} \sup_n \left(\int_{|x| \geq R} |x|^2 d\mu_n(x) \right) = 0.$$

Notations : Dans la suite, on notera \int pour $\int_{\mathbb{R}^d}$ ou $\int_{\mathbb{R}^d \times \mathbb{R}^d}$.

Remarque : Chercher $\alpha, \beta > 0$ tels que $(a+b)^2 \leq (1+\alpha^2)a^2 + (1+\beta^2)b^2$, revient à trouver $\alpha, \beta > 0$ tels que

$$2ab \leq \alpha^2 a^2 + \beta^2 b^2 = (\alpha a - \beta b)^2 + 2\alpha\beta ab.$$

Ce qui est le cas si $\alpha\beta = 1$. Ainsi pour tout $\epsilon > 0$, on peut prendre $\alpha = \sqrt{\epsilon}$, $\beta = 1/\sqrt{\epsilon}$ et obtenir :

$$(a+b)^2 \leq (1+\epsilon)a^2 + \left(1 + \frac{1}{\epsilon}\right)b^2.$$

Il résulte de cette remarque, le lemme suivant :

LEMME 4.1 Soient $(\mu_n)_n$ est une suite de mesures et μ dans $P_2(\mathbb{R}^d)$.

1. Si $\sup_{n \geq 1} W_2(\mu_n, \mu) < +\infty$ alors :

$$\sup_{n \geq 1} m_2(\mu_n) < +\infty,$$

et $(\mu_n)_n$ est tendue.

2. Si de plus $(\mu_n)_n$ qui converge vers μ au sens de W_2 , alors pour tout $x_0 \in \mathbb{R}^d$:

$$\limsup_{n \rightarrow +\infty} \int |x - x_0|^2 d\mu_n(x) \leq \int |x - x_0|^2 d\mu(x).$$

Démonstration :

On considère pour chaque $n \geq 1$, $\gamma_n \in \mathcal{M}_0^2(\mu_n, \mu)$ un couplage optimal entre μ_n et μ . Fixons $x_0 \in \mathbb{R}^d$. En appliquant successivement l'inégalité triangulaire de la norme euclidienne, et la remarque précédente on a :

$$\begin{aligned} \int |x - x_0|^2 d\mu_n(x) &\leq \int (|x - y| + |y - x_0|)^2 d\gamma_n(x, y) \\ &\leq \left(1 + \frac{1}{\epsilon}\right) \int |x - y|^2 d\gamma_n(x, y) + (1 + \epsilon) \int |y - x_0|^2 d\mu(y). \end{aligned}$$

On reconnaît dans le membre de droite, le premier terme $W_2^2(\mu_n, \mu)$.

1. Si $\sup_{n \geq 1} W_2(\mu_n, \mu) < +\infty$, on a en spécialisant $x_0 = 0$: $K := \sup_{n \geq 1} m_2(\mu_n) < +\infty$. Soit alors $\epsilon > 0$. Pour $R > \sqrt{K}/\sqrt{\epsilon}$, on a pour tout $n \in \mathbb{N}^*$ et grâce à l'inégalité de Markov :

$$\mu_n(|x| > R) \leq \frac{1}{R^2} \int |x|^2 d\mu_n(x) \leq \frac{K}{R^2} < \epsilon.$$

Il en découle que $(\mu_n)_n$ est tendue.

2. Si $W_2(\mu_n, \mu)$ tend vers 0 quand $n \rightarrow +\infty$, alors on obtient le deuxième point du lemme ■

Nous pouvons passer à la démonstration de la proposition :

Démonstration :

2. \Rightarrow 1. Par hypothèse, on a déjà que $(\mu_n)_n$ converge faiblement vers μ . Le théorème de représentation de Skorohod nous dit qu'il existe $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et des variables aléatoires sur cet espace : $(X_n)_n, X$ telles que :

$$\begin{aligned} \forall n \geq 1 \quad \mathcal{L}(X_n) &= \mu_n, & \mathcal{L}(X) &= \mu \\ (X_n)_n &\longrightarrow X & \text{p.s} \end{aligned}$$

En utilisant l'autre hypothèse, on a :

$$\lim_{R \rightarrow +\infty} \sup_n \int_{|x| > R} |x|^2 d\mu_n(x) = \lim_{R \rightarrow +\infty} \sup_n \int_{|X_n| > R} |X_n|^2 d\mathbb{P} = 0.$$

Autrement dit la suite $(|X_n|^2)_n$ est équi-intégrable. Et la convergence presque sûre entraîne la convergence en probabilité, donc $(X_n)_n$ converge en probabilité vers X . Par le Lemme (A.1) (voir Annexe), cela est équivalent à dire que :

$$\begin{aligned} (X_n)_n &\longrightarrow X & \text{dans } L^2(\mathbb{P}) \\ \text{i.e } \mathbb{E}(|X_n - X|^2) &\longrightarrow 0. \end{aligned}$$

De plus, d'après la définition de W_2 on a l'inégalité :

$$W_2^2(\mu_n, \mu) = \inf_{(Y_n, Y) \in \mathcal{M}(\mu_n, \mu)} \mathbb{E}(|Y_n - Y|^2) \leq \mathbb{E}(|X_n - X|^2) \longrightarrow 0.$$

On a donc montré que $(\mu_n)_n$ converge vers μ au sens de W_2 .

1. \Rightarrow 2. Par hypothèse on a : $\lim_{n \rightarrow +\infty} W_2(\mu_n, \mu) = 0$.

* Le lemme 4.1 nous donne l'inégalité :

$$\limsup_{n \rightarrow +\infty} \int |x|^2 d\mu_n(x) \leq \int |x|^2 d\mu(x).$$

Pour $\epsilon > 0$, il existe donc $n_0 \in \mathbb{N}^*$ tel que, pour tout $n \geq n_0$:

$$\int |x|^2 d\mu_n(x) \leq \int |x|^2 d\mu(x) + \epsilon.$$

Soit $R > 0$. La fonction sur $\mathbb{R}^d \rightarrow \mathbb{R}_+$, $x \mapsto 1_{\{|x| < R\}} |x|^2$ est semi-continue inférieurement et donc :

$$\liminf_{n \rightarrow +\infty} \int_{|x| < R} |x|^2 d\mu_n(x) \geq \int_{|x| < R} |x|^2 d\mu(x).$$

Il existe alors $n_1 \geq n_0$ tel que, pour tout $n \geq n_1$:

$$\int_{|x| < R} |x|^2 d\mu_n \geq \int_{|x| < R} |x|^2 d\mu(x) - \epsilon.$$

Ainsi pour $n \geq n_1$, on obtient :

$$\int_{|x| \geq R} |x|^2 d\mu_n \leq \int_{|x| \geq R} |x|^2 d\mu + 2\epsilon.$$

En prenant le *sup* sur $n \geq 1$ et la limite quand R tend vers $+\infty$, il vient que :

$$\lim_{R \rightarrow +\infty} \sup_{n \geq 1} \int_{|x| \geq R} |x|^2 d\mu_n = 0. \quad (4.7)$$

* Comme $(\mu_n)_n$ converge au sens de W_2 vers μ , le lemme 4.1 indique en particulier que la suite est tendue. Il existe donc une sous-suite $(\mu_{n_k})_k$ qui converge faiblement vers une mesure μ_∞ . De plus cette sous-suite vérifie aussi (4.7). On peut donc appliquer le sens 2. \Rightarrow 1. à cette sous-suite : $(\mu_{n_k})_k$ converge au sens de W_2 vers μ_∞ . Mais par hypothèse $(\mu_n)_n$ converge vers μ au sens de W_2 donc par unicité de la limite, il en découle que $\mu_\infty = \mu$. Ainsi, toute sous-suite de $(\mu_n)_n$ converge faiblement vers μ , et donc $(\mu_n)_n$ converge faiblement vers μ ■

Enfin terminons cette partie, en donnant la propriété fondamentale de la distance de Wasserstein :

PROPOSITION 4.6 *L'espace métrique $(P_2(\mathbb{R}^d), W_2)$ est complet.*

Démonstration :

Prenons une suite de Cauchy $(\mu_n)_n \subset P_2(\mathbb{R}^d)$ pour la distance W_2 . Soit $\epsilon > 0$. Il existe un entier $n_0 > 0$ tel que pour tout $n, m \geq n_0$, on ait $W_2(\mu_n, \mu_m) < \epsilon$.

* Montrons dans un premier temps que $(\mu_n)_n$ est tendue. Soit $m \geq n_0$. On a par inégalité triangulaire :

$$W_2(\mu_1, \mu_n) \leq W_2(\mu_1, \mu_{n_0}) + W_2(\mu_{n_0}, \mu_m) < W_2(\mu_1, \mu_{n_0}) + \epsilon \leq C_0 < +\infty.$$

Ainsi :

$$\begin{aligned} \sup_{m \geq n_0} W_2(\mu_1, \mu_m) &\leq C_0 < +\infty & \text{et} & \quad \sup_{1 \leq m \leq n_0} W_2(\mu_1, \mu_m) = C_1 < +\infty \\ \text{donc } \sup_{m \geq 1} W_2(\mu_1, \mu_m) &\leq \max(C_0, C_1) < +\infty. \end{aligned}$$

Donc $(\mu_n)_n$ est bornée pour W_2 et par le lemme 4.1, cela implique que la suite $(\mu_n)_n$ est bien tendue.

* On en déduit qu'il existe une sous-suite $(\mu_{n_k})_k$ qui converge faiblement vers une mesure μ . On considère pour chaque mesure de cette sous-suite, un couplage $\gamma_{n, n_k} \in \mathcal{M}_0^2(\mu_n, \mu_{n_k})$. Comme $(\mu_n)_n$ est tendue, il s'ensuit que $(\gamma_{n, n_k})_k$. Il existe donc une sous-suite $(\gamma_{n, n_{k_p}})_p$ qui converge faiblement vers $\gamma_{n, \infty}$. Pour ne pas alourdir les notations, on va supposer que c'est $(\gamma_{n, n_k})_k$ qui converge faiblement vers $\gamma_{n, \infty}$. Soit $\varphi \in \mathcal{C}_b(\mathbb{R}^d \times \mathbb{R}^d)$. La convergence faible se traduit par :

$$\int \varphi d\gamma_{n, n_k} \longrightarrow \int \varphi d\gamma_{n, \infty}.$$

Soit alors $\psi \in \mathcal{C}_b(\mathbb{R}^d)$. Posons $\varphi = \psi \circ \pi_1 \in \mathcal{C}_b(\mathbb{R}^d \times \mathbb{R}^d)$, alors on a :

$$\int \varphi d\gamma_{n,n_k} = \int \psi d\mu_{n_k} \longrightarrow \int \psi d\mu.$$

Donc par unicité de la limite, on a que $\int \psi d\mu = \int \varphi d\gamma_{n,\infty}$. Cela signifie que $(\pi_2)_*\gamma_{n,\infty} = \mu$ ■

Remarque : Peut-être que cette démonstration aurait été plus succincte en utilisant la caractérisation de la convergence W_2 donnée par la proposition 4.5. Faute de temps, nous n'avons pas pu y réfléchir de manière approfondie.

Deuxième partie

Couplages avec une loi de Poisson

Cette partie est consacrée à l'étude de cas discrets.

La loi de Poisson de paramètre $\lambda > 0$ est la loi portée par \mathbb{N} , de moyenne λ et variance λ :

$$\mathcal{P}(\lambda) = \sum_{n \in \mathbb{N}} \frac{\lambda^n}{n!} e^{-\lambda} \cdot \delta_n, \text{ de fonction génératrice : } s \mapsto e^{-\lambda(1-s)}.$$

5 Quelques Couplages classiques

5.1 Couplages entre deux lois de Poisson $\mathcal{P}(\lambda)$ et $\mathcal{P}(\lambda')$ avec $\lambda > \lambda'$

Si nous calculons la variation totale entre ces deux lois, et puisque $e^{-\lambda} < e^{-\lambda'}$, alors que $\left\{ n \in \mathbb{N} / \frac{\lambda^n}{n!} e^{-\lambda} < \frac{\lambda'^n}{n!} e^{-\lambda'} \right\}$ est une demi-droite d'entiers : $[N_{\lambda, \lambda'}, +\infty[$, avec $N_{\lambda, \lambda'} \geq 1$, on obtient :

$$\begin{aligned} \|\mathcal{P}(\lambda) - \mathcal{P}(\lambda')\|_{VT} &= \sum_{n \in \mathbb{N}} \left(\frac{\lambda^n}{n!} e^{-\lambda} - \frac{\lambda'^n}{n!} e^{-\lambda'} \right)^+ \\ &= \sum_{n \geq N_{\lambda, \lambda'}} \left(\frac{\lambda^n}{n!} e^{-\lambda} - \frac{\lambda'^n}{n!} e^{-\lambda'} \right) \\ &= (e^{-\lambda'} - e^{-\lambda}) + \sum_{1 \leq n < N_{\lambda, \lambda'}} \left(\frac{\lambda'^n}{n!} e^{-\lambda'} - \frac{\lambda^n}{n!} e^{-\lambda} \right) \\ &= e^{-\lambda'} (1 - e^{-(\lambda - \lambda')}) + \sum_{1 \leq n < N_{\lambda, \lambda'}} \left(\frac{\lambda'^n}{n!} e^{-\lambda'} - \frac{\lambda^n}{n!} e^{-\lambda} \right) \end{aligned}$$

Notons qu'il n'est en général pas facile d'avoir un estimé effectif de ces quantités.

Cependant dans le cas où : $0 < \lambda' < \lambda < 1$, on constate que : $\lambda' e^{-\lambda'} < \lambda e^{-\lambda}$ (il suffit d'étudier les variations de la fonction qui : $x \mapsto x e^{-x}$), et que de ce fait $N_{\lambda, \lambda'} = 1$, avec ainsi :

$$\|\mathcal{P}(\lambda) - \mathcal{P}(\lambda')\|_{VT} = e^{-\lambda'} (1 - e^{-(\lambda - \lambda')}) \leq 1 - e^{-(\lambda - \lambda')} \leq \lambda - \lambda'$$

En fait nous pouvons retrouver cette majoration dans le cas général de façon assez simple en utilisant un couplage particulier (suffisamment bon, bien que non optimal).

Nous utiliserons le fait (dont la vérification est immédiate en termes de fonctions génératrices), que la somme de deux variables de Poisson indépendantes est une variable de Poisson (de paramètre la somme des paramètres). Si X' et X'' sont deux variables aléatoires de Poisson de paramètres respectifs λ' et $\lambda'' = \lambda - \lambda'$, toutes deux définies sur le même espace de probabilité et indépendantes entre elles, leur somme $X := X' + X''$ est une variable de Poisson de paramètre λ . Le couple (X, X') réalise un couplage γ entre les lois de Poisson $\mathcal{P}(\lambda)$ et $\mathcal{P}(\lambda')$. Vu que :

$$\gamma(\Delta_{\mathbb{N}}^c) = \mathbb{P}[X \neq X'] = \mathbb{P}[X'' \neq 0] = 1 - e^{-(\lambda - \lambda')}$$

on en déduit que :

$$\|\mathcal{P}(\lambda) - \mathcal{P}(\lambda')\|_{VT} \leq 1 - e^{-(\lambda - \lambda')} \leq \lambda - \lambda'.$$

5.2 Couplages d'une loi de Bernoulli avec une loi de Poisson

La loi de Bernoulli de paramètre $p \in]0, 1[$ est la loi portée par $\{0, 1\}$, de moyenne p et variance $p(1 - p)$:

$$\mathcal{B}(p) = (1 - p)\delta_0 + p\delta_1, \text{ de fonction génératrice : } s \mapsto 1 - p(1 - s).$$

La distance en variation totale entre μ , la loi de Poisson de paramètre $\lambda > 0$ et ν , la loi de Bernoulli de paramètre p est :

$$\begin{aligned} \|\mathcal{P}(\lambda) - \mathcal{B}(p)\|_{VT} &= (e^{-\lambda} - (1-p))^+ + (\lambda e^{-\lambda} - p)^+ + \sum_{i \geq 2} \frac{\lambda^i}{i!} e^{-\lambda} \\ &= (e^{-\lambda} - (1-p))^+ + (\lambda e^{-\lambda} - p)^+ + 1 - e^{-\lambda}(1+\lambda) \end{aligned}$$

Remarque : Ce résultat se simplifie dans les deux cas particuliers suivants :

Dans le cas où $p = \lambda \in]0, 1[$, vu que : $1 - \lambda \leq e^{-\lambda} \leq 1$, on a :

$$\|\mathcal{P}(\lambda) - \mathcal{B}(\lambda)\|_{VT} = (e^{-\lambda} - (1-\lambda)) + 0 + 1 - e^{-\lambda}(1+\lambda) = \lambda(1 - e^{-\lambda}) \leq \lambda^2 = p^2$$

Dans le cas où $p = 1 - e^{-\lambda}$, vu que : $(1+\lambda)e^{-\lambda} \leq 1$, on a :

$$\|\mathcal{P}(\lambda) - \mathcal{B}(1 - e^{-\lambda})\|_{VT} = \begin{cases} 1 - e^{-\lambda}(1+\lambda) = p + (1-p) \ln(1-p) \leq p^2 \leq \lambda^2 \\ \lambda^2 e^{-\lambda} \sum_{\ell \geq 0} \frac{\lambda^\ell}{(\ell+2)!} \leq \frac{\lambda^2}{2} e^{-\lambda} \sum_{\ell \geq 0} \frac{\lambda^\ell}{\ell!} = \frac{1}{2} \lambda^2 \end{cases}$$

Dans cette partie, X et Y seront des variables aléatoires définies sur le même espace probabilisé, ayant pour loi respective $\mu = \mathcal{P}(\lambda)$ et $\nu = \mathcal{B}(p)$ où $0 < \lambda < 1$ et $0 < p < 1$, réalisant donc un couplage γ entre μ et ν .

* Dans le cas où $p = \lambda \in]0, 1[$, considérons deux couplages possibles γ et $\tilde{\gamma}$ entre X et Y .

Un premier couplage : Ce couplage est valide sous la condition $\frac{\lambda}{1+\lambda} \leq e^{-\lambda}$ donc en particulier si $\frac{\lambda}{1+\lambda} \leq 1 - \lambda$ à savoir si $0 < \lambda \leq \frac{\sqrt{5}-1}{2}$. Il est décrit dans le tableau ci dessous :

$$\gamma_{k,\ell} = \mathbb{P}[X = k, Y = \ell]$$

$Y \backslash X$	0	1	2	k	$\mathcal{B}(p)$
0	$e^{-\lambda} - \lambda(1 - e^{-\lambda})$	0	$\frac{\lambda^2}{2} e^{-\lambda}$	$\frac{\lambda^k}{k!} e^{-\lambda}$	$1 - \lambda$
1	$\lambda(1 - e^{-\lambda})$	$\lambda e^{-\lambda}$	0	0	λ
$\mathcal{P}(\lambda)$	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\frac{\lambda^2}{2} e^{-\lambda}$	$\frac{\lambda^k}{k!} e^{-\lambda}$	

Il est tel que :

$$\begin{aligned} \mathbb{P}[X \neq Y] &= 1 - \mathbb{P}[X = 0, Y = 0] - \mathbb{P}[X = 1, Y = 1] \\ &= 1 - \left(e^{-\lambda} - \lambda(1 - e^{-\lambda}) + \lambda e^{-\lambda} \right) = 1 + \lambda - e^{-\lambda}(1 + 2\lambda) \leq 1 + \lambda - (1 - \lambda)(1 + 2\lambda) = 2\lambda^2 \end{aligned}$$

Vu que :

$$\|\mathcal{P}(\lambda) - \mathcal{B}(\lambda)\|_{VT} = \lambda(1 - e^{-\lambda}) < 1 + \lambda - e^{-\lambda}(1 + 2\lambda) = \mathbb{P}[X \neq Y]$$

ce couplage n'est pas optimal. Pour λ assez petit il donne cependant une estimation raisonnable (avec un coefficient multiplicatif de l'ordre de 2) de la variation totale entre la loi de Poisson et la loi de Bernoulli.

Un second couplage : Essayons d'obtenir cette fois un couplage optimal. Pour cela, la Proposition 4.2 nous indique de placer sur la diagonale la valeur $\min(\mathbb{P}(X = i), \mathbb{P}(Y = i))$. Ce qui donne :

$$\tilde{\gamma}_{k,\ell} = \mathbb{P}[X = k, Y = \ell]$$

$Y \backslash X$	0	1	2	k	$\mathcal{B}(p)$
0	$1 - \lambda$	0	0	0	$1 - \lambda$
1	$e^{-\lambda} - (1 - \lambda)$	$\lambda e^{-\lambda}$	$\frac{\lambda^2}{2} e^{-\lambda}$	$\frac{\lambda^k}{k!} e^{-\lambda}$	λ
$\mathcal{P}(\lambda)$	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\frac{\lambda^2}{2} e^{-\lambda}$	$\frac{\lambda^k}{k!} e^{-\lambda}$	

On a :

$$\mathbb{P}[X \neq Y] = 1 - \mathbb{P}[X = 0, Y = 0] - \mathbb{P}[X = 1, Y = 1] = 1 - (1 - \lambda) - \lambda e^{-\lambda} = \lambda(1 - e^{-\lambda}) \leq \lambda^2$$

On constate que :

$$\mathbb{P}(X \neq Y) = \lambda(1 - e^{-\lambda}) = \|\mu - \nu\|_{VT},$$

ce qui confirme bien le fait que ce couplage est optimal (et c'est en fait le seul).

* Dans le cas où $p = 1 - e^{-\lambda}$, on a à nouveau un couplage optimal et un seul $\hat{\gamma}$, qui est donné sous la forme :

$$\hat{\gamma}_{k,\ell} = \mathbb{P}[X = k, Y = \ell]$$

$Y \backslash X$	0	1	2	k	$\mathcal{B}(p)$
0	$1 - p$	0	0	0	$1 - p$
1	0	$\lambda e^{-\lambda}$	$\frac{\lambda^2}{2} e^{-\lambda}$	$\frac{\lambda^k}{k!} e^{-\lambda}$	p
$\mathcal{P}(\lambda)$	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\frac{\lambda^2}{2} e^{-\lambda}$	$\frac{\lambda^k}{k!} e^{-\lambda}$	

Un tel couplage est réalisé algébriquement en posant : $Y = \min(X, 1)$

$$\mathbb{P}[X \neq Y] = \mathbb{P}[X \geq 2] = e^{-\lambda} \sum_{k \geq 2} \frac{\lambda^k}{k!} = 1 - e^{-\lambda}(1 + \lambda) = \|\mathcal{P}(\lambda) - \mathcal{B}(p)\|_{VT}$$

5.3 Couplages avec une somme de lois de Bernoulli

Soit X suit une loi de Poisson de paramètre λ , on peut toujours écrire $X = \sum_{i=1}^n X_i$ où les X_i suivent respectivement une loi de Poisson de paramètre $p_i \in]0, 1[$, indépendantes et telles que $\sum_{i=1}^n p_i = \lambda$. On considère alors pour $i = 1, \dots, n$, des variables Y_i qui suivent les lois de Bernoulli de paramètre respectif p_i . On note $Y = \sum_{i=1}^n Y_i$.

5.3.1 Cas où tout est indépendant

Ici on suppose tous les Y_i sont indépendants.

Pour chaque $i = 1, \dots, n$, on sait coupler de façon optimale chaque couple (X_i, Y_i) , et les couples étant indépendants entre eux, on obtient en procédant de la sorte :

$$\mathbb{P}(X \neq Y) \leq \sum_{i=1}^n \mathbb{P}(X_i \neq Y_i) \leq \sum_{i=1}^n p_i^2.$$

Remarque : Dans le cas particulier où $p_i = p$ pour tout i , c'est-à-dire si $\mathcal{L}(Y) = \mathcal{B}(n, \frac{\lambda}{n})$, on obtient la majoration :

$$\mathbb{P}(X \neq Y) \leq \frac{\lambda^2}{n} \longrightarrow 0 \quad \text{quand } n \rightarrow \infty.$$

Ce qui fournit un ordre de vitesse de convergence de la Loi Binomiale vers la Loi de Poisson.

5.3.2 Cas où il existe de la dépendance

Ici on ne suppose plus l'indépendance des Y_i . On note μ la loi de Y et ν celle de X . Le but de cette section est de coupler μ et ν , et de majorer l'erreur.

Soient pour $i = 1, \dots, n$, $\lambda'_i = -\log(1 - p_i)$. Par hypothèse sur les p_i , on a $\lambda_i > 0$, et on considère alors des variables aléatoires indépendantes Y'_i de loi de Poisson de paramètre λ'_i . Ces paramètres ont été choisis de manière à ce que Y_i et $\min(Y'_i, 1)$ suivent la même loi. En effet pour $i = 1, \dots, n$:

$$\begin{aligned} \mathbb{P}(Y'_i = 0) &= e^{-\lambda'_i} = 1 - p_i = \mathbb{P}(Y_i = 0), \\ \mathbb{P}(Y'_i \geq 1) &= 1 - \mathbb{P}(Y'_i = 0) = p_i = \mathbb{P}(Y_i = 1). \end{aligned}$$

On note $Y' = \sum_{i=1}^n Y'_i$. Il découle de l'indépendance des Y'_i , que $\mathcal{L}(Y') = \mathcal{P}(\lambda')$ où $\lambda' = \sum_{i=1}^n \lambda'_i$.

Remarque : Il est clair que pour tout $i = 1, \dots, n$, Y_i et Y'_i ne sont pas indépendantes, et on ne peut donc pas appliquer la méthode précédente de couplage.

Nous cherchons à approximer Y par Y' . Ainsi :

$$\mathbb{P}(Y \neq Y') \leq \sum_{i=1}^n \mathbb{P}(Y_i \neq Y'_i) = \sum_{i=1}^n \mathbb{P}(Y'_i \geq 2).$$

Mais pour $\alpha > 0$ on a :

$$\sum_{i=2}^{\infty} e^{-\alpha} \frac{\alpha^i}{i!} = \alpha^2 \sum_{i=0}^{\infty} e^{-\alpha} \frac{\alpha^i}{(i+2)!} \leq \frac{\alpha^2}{2}.$$

Cela implique que $\mathbb{P}(Y'_i \geq 2) \leq \frac{(\lambda'_i)^2}{2}$. Finalement on obtient :

$$\mathbb{P}(Y \neq Y') \leq \frac{1}{2} \sum_{i=1}^n (\lambda'_i)^2.$$

De plus, si tous les $p_i \in [0, \frac{1}{2}]$, alors un calcul élémentaire donne :

$$-\log(1 - p_i) \leq \sqrt{2}p_i,$$

et dans ce cas :

$$\boxed{\|\mu - \mathcal{P}(\lambda')\|_{VT} \leq \mathbb{P}(Y \neq Y') \leq \sum_{i=1}^n p_i^2} \tag{5.8}$$

Cette approximation n'est pas celle que l'on recherche, puisque les paramètres $\lambda' = \sum_{i=1}^n \lambda'_i$ et $\lambda = \sum_{i=1}^n p_i$ sont différents. Il nous faut donc nous ramener à une troisième loi $\mathcal{P}(\lambda)$. Soient Z une variable qui suit la loi de Poisson de paramètre λ , et Z' une variable qui suit la loi de Poisson de paramètre λ' .

On a pour tout i entier ≥ 1 ,

$$\lambda'_i = -\log(1 - p_i) \geq p_i \implies \lambda' \geq \lambda.$$

Ainsi on peut écrire $\lambda' = \lambda + (\lambda' - \lambda)$ et considérer Z'' une variable aléatoire qui suit la loi de Poisson de paramètre $(\lambda' - \lambda)$. Le couplage naturel entre Z et Z' est alors le suivant :

$$\mathbb{P}(Z \neq Z') = \mathbb{P}(Z'' \neq 0) = 1 - e^{-(\lambda' - \lambda)} \leq \lambda' - \lambda.$$

De plus $\lambda' - \lambda = \sum_{i=1}^n (\lambda'_i - p_i) = -\sum_{i=1}^n (\log(1 - p_i) + p_i)$. Et on vérifie facilement que si tous les p_i sont dans $[0, \frac{1}{2}]$, alors :

$$-\log(1 - p_i) - p_i \leq p_i^2.$$

Cela signifie finalement que :

$$\boxed{\|\nu - \mathcal{P}(\lambda')\|_{VT} \leq \mathbb{P}(Z \neq Z') \leq \sum_{i=1}^n p_i^2} \quad (5.9)$$

En combinant (5.8) et (5.9), il vient sous conditions que tous les $p_i \in [0, \frac{1}{2}]$, l'inégalité suivante :

$$\|\mu - \nu\|_{VT} \leq \|\mu - \mathcal{P}(\lambda')\|_{VT} + \|\nu - \mathcal{P}(\lambda')\|_{VT} \leq 2 \sum_{i=1}^n p_i^2.$$

6 Méthode de Stein

6.1 Présentation

Présentons de manière générale la *méthode de Stein*, c'est-à-dire dans le cadre le plus abstrait possible. Les données sont les suivantes. On regarde (S, \mathcal{S}, μ) un espace de probabilité. Soit $\mathcal{X}_0 = \{h : S \rightarrow \mathbb{R} \text{ } \mu\text{-intégrable}\}$. Dans ce qui suit, j'explique pourquoi et comment nous utiliserons la *méthode de Stein*.

- But : On veut savoir calculer $\int h d\mu$ pour tout $h \in \mathcal{X}_0$.
- Problème : Le calcul de ces intégrales peut être très fastidieux, notamment si μ est compliquée.
- Idée : On remplace alors μ par μ_0 une probabilité sur (S, \mathcal{S}) proche de μ et simple de manière à ce que $\int h d\mu_0$ soit facilement calculable pour tout $h \in \mathcal{X}_0$.
- Méthode de Stein : Méthode pour estimer la distance $\|\mu - \mu_0\|_{VT}$, et ainsi l'erreur commise en remplaçant μ par μ_0 .

La méthode de Stein consiste à considérer une équation, qui nous donnera des informations sur $\|\mu - \mu_0\|_{VT}$. Pour cela, on introduit \mathcal{F}_0 un espace de fonctions quelconque et $T_0 : \mathcal{F}_0 \rightarrow \mathcal{X}_0$ nommé *opérateur de Stein* qui satisfait à l'assertion suivante : pour tout $h \in \mathcal{X}_0$ l'équation de Stein

$$T_0(fh) = h - \int_S h d\mu_0 \quad (6.10)$$

a une unique solution $f_h \in \mathcal{F}_0$, que l'on appelle la *transformée de Stein associée à h*.

Remarque : Si $h \in \mathcal{X}_0$ et f est la transformée de Stein associée, alors on voit immédiatement que :

$$\int T_0(f) d\mu_0 = 0,$$

et

$$\int T_0(f) d\mu \leq \|\mu - \mu_0\|_{VT}.$$

D'après ce que nous venons de voir, tout repose sur l'opérateur T_0 . Certains mathématiciens se sont donc penchés sur la construction d'un tel opérateur. En voici une méthode générale, c'est celle que nous utiliserons pour trouver un opérateur de Stein dans le cas de la loi de Poisson.

Un moyen de construction de T_0 :

1. Choisir un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et un couple de variables (X, Y) qui a pour marginales μ_0 et tel que sa loi soit la même que la loi de (Y, X) .
2. Choisir une application $\alpha : \mathcal{F}_0 \rightarrow \mathcal{F} = \{F : S^2 \rightarrow \mathbb{R}, \mathbb{E}(|F(X, Y)|) < \infty \text{ et } F(x, y) = -F(y, x)\}$.
3. Soit maintenant $T : \mathcal{F} \rightarrow \mathcal{X}_0$ défini par $T(F)(x) = \mathbb{E}(F(X, Y)|X = x)$ pour tout $x \in S$. On prend alors $\underline{T_0} = T \circ \alpha$ comme opérateur de Stein.

Sans plus attendre, appliquons tout cela à un cas concret : lorsque μ_0 est une loi de Poisson.

6.2 Cas de la loi de Poisson

En gardant les notations précédentes, nous regardons l'espace mesurable $(S, \mathcal{S}) = (\mathbb{N}, \mathcal{P}(\mathbb{N}))$ muni d'une quelconque mesure μ , et $\mathcal{F}_0 = \{h : \mathbb{N} \rightarrow \mathbb{R} \text{ } \mu\text{-mesurable}\}$. L'objet de cette partie est de comparer μ à une loi de Poisson qui vit aussi sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$, via la méthode de Stein. Prenons alors $\mu_0 = \mathcal{P}(\lambda)$.

Essayons de construire un opérateur de Stein pour μ_0 . Pour cela nous adoptons la méthode énoncée ci-dessus. Soit $\{Z_t, t \in \mathbb{R}_+\}$ un processus stationnaire de naissance (avec intensité λ) et de mort (avec intensité i au rang i) sur \mathbb{N} . On sait que la loi stationnaire d'un tel processus est $\mathcal{P}(\lambda) = \mu_0$. Pour $u > 0$ on a alors (Z_0, Z_u) est un couple de variables de marginale μ_0 qui a la même loi que (Z_u, Z_0) .

Maintenant considérons l'application $\alpha : \mathcal{F}_0 \rightarrow \mathcal{F}$ qui à $g \in \mathcal{F}_0$ associe $\alpha(g) : (k, l) \mapsto g(k) - g(l)$ pour tout $k, l \in \mathbb{N}$. Enfin on a $T : \mathcal{F} \rightarrow \mathcal{F}_0$ qui à $F \in \mathcal{F}$ associe $T(F) : k \mapsto \mathbb{E}(F(Z_0, Z_u)|Z_0 = k)$ pour tout $k \in \mathbb{N}$.

Ainsi pour $g \in \mathcal{F}_0$ et $k \in \mathbb{N}$ on a :

$$\begin{aligned} \lim_{u \rightarrow 0^+} \frac{1}{u} (T \circ \alpha)(g)(k) &= \lim_{u \rightarrow 0^+} \frac{1}{u} \mathbb{E}(g(Z_u) - g(Z_0)|Z_0 = k) \\ &= \lim_{u \rightarrow 0^+} \frac{1}{u} \mathbb{E}(g(Z_u) - g(k)|Z_0 = k). \end{aligned}$$

Or grâce aux propriétés d'un processus de naissance et de mort, on sait que :

$$\frac{1}{u} (\mathbb{P}(Z_u = k) - \mathbb{P}(Z_0 = k)) = k\mathbb{P}(Z_0 = k - 1) + \lambda\mathbb{P}(Z_0 = k + 1) - (k + \lambda)\mathbb{P}(Z_0 = k),$$

et donc :

$$\begin{aligned} \lim_{u \rightarrow 0^+} \frac{1}{u} (T \circ \alpha)(g)(k) &= \lambda g(k + 1) + k g(k - 1) - (\lambda + k) g(k) \\ &= T_0(f)(k), \end{aligned}$$

avec $f(k) = g(k) - g(k - 1)$ et $T_0(f)(k) = \lambda f(k + 1) - k f(k)$. Finalement on a trouvé une forme de l'opérateur de Stein, qui est :

$\begin{aligned} T_0 : \mathcal{F}_0 &\rightarrow \mathcal{F}_0 \\ f &\mapsto T_0(f) : k \mapsto \lambda f(k + 1) - k f(k) \end{aligned}$

Maintenant que nous avons notre opérateur de Stein, énonçons une première proposition.

PROPOSITION 6.1 *L'équation de Stein (6.10) avec le T_0 que nous venons de construire, admet pour tout $h \in \mathcal{X}_0$ une solution f , unique à $f(0)$ près. De plus f peut être explicitée de la sorte :*

$$f(k) = \frac{(k-1)!}{\lambda^k} \sum_{i=0}^{k-1} \left(h(i) - \int h d\mu_0 \right) \frac{\lambda^i}{i!} \quad (6.11)$$

$$= -\frac{(k-1)!}{\lambda^k} \sum_{i=k}^{\infty} \left(h(i) - \int h d\mu_0 \right) \frac{\lambda^i}{i!} \quad (6.12)$$

Démonstration :

Le premier point résulte de la définition de l'équation de Stein. De plus pour $h \in \mathcal{X}_0$ et f la transformée de Stein associée, on choisit $f(0) = 0$ et on a en remplaçant dans (6.10) : $\lambda f(k+1) - kf(k) = h(k) - \int h d\mu_0$. Donc :

$$\begin{aligned} f(k+1) &= \frac{k}{\lambda} \left(f(k) + \frac{1}{k} \left(h(k) - \int h d\mu_0 \right) \right) \\ &= \frac{k(k-1)}{\lambda^2} \left(f(k-1) + \frac{1}{k-1} \left(h(k-1) - \int h d\mu_0 \right) + \frac{\lambda}{k(k-1)} \left(h(k) - \int h d\mu_0 \right) \right) \\ &= \frac{k!}{\lambda^{k+1}} \left(h(0) - \int h d\mu_0 + \dots + \frac{\lambda^{k-1}}{(k-2)!} \left(h(k-1) - \int h d\mu_0 \right) + \frac{\lambda^k}{k!} \left(h(k) - \int h d\mu_0 \right) \right) \\ &= \frac{k!}{\lambda^{k+1}} \sum_{i=0}^k \left(h(i) - \int h d\mu_0 \right) \frac{\lambda^i}{i!}. \end{aligned}$$

Ce qui démontre (6.11). Pour (6.12), il suffit de remarquer que :

$$\sum_{i=0}^{\infty} \left(h(i) - \int h d\mu_0 \right) \frac{\lambda^i}{i!} = e^\lambda \int h d\mu_0 - \int h d\mu_0 e^\lambda = 0.$$

■

Les expressions (6.11) et (6.12) vont nous être utiles pour les prochains résultats que nous allons établir. Mais avant, énonçons le Lemme de Stein qui caractérise la loi de Poisson sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$.

LEMME 6.1 (de Stein) *Soit μ une probabilité sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$. On a :*

$$\mu = \mathcal{P}(\lambda) \iff \int_{\mathbb{N}} T_0 f d\mu = 0 \quad \forall f : \mathbb{N} \longrightarrow \mathbb{R} \text{ bornée.}$$

Démonstration :

Condition nécessaire : $\mu = \mathcal{P}(\lambda)$, alors par définition de l'opérateur de Stein on a pour $f : \mathbb{N} \longrightarrow \mathbb{R}$ bornée :

$$\begin{aligned} \int_{\mathbb{N}} T_0 f d\mu &= e^{-\lambda} \sum_{i=0}^{\infty} T_0 f(i) \frac{\lambda^i}{i!} \\ &= e^{-\lambda} \sum_{i=0}^{\infty} (\lambda f(i+1) - i f(i)) \frac{\lambda^i}{i!} \\ &= e^{-\lambda} \left(\sum_{i=0}^{\infty} \frac{\lambda^{i+1}}{i!} f(i+1) - \sum_{i=1}^{\infty} \frac{\lambda^i}{i!} f(i) \right) = 0. \end{aligned}$$

Condition suffisante : $\int_{\mathbb{N}} T_0 f d\mu = 0$ pour toute $f : \mathbb{N} \longrightarrow \mathbb{R}$ bornée. Soient $A \subset \mathbb{N}$, $h_A = 1_A$ et f_A la transformée de Stein associée à h_A . f_A est donc bornée et par hypothèse on a donc :

$$\begin{aligned} \int_{\mathbb{N}} T_0 f_A d\mu = 0 &= \int_{\mathbb{N}} h_A d\mu - \int_{\mathbb{N}} h_A d\mu_0 \\ &= \mu(A) - \mu_0(A). \end{aligned}$$

Ceci étant vrai pour tout $A \subset \mathbb{N}$, on en déduit que $\mu = \mu_0 = \mathcal{P}(\lambda)$ ■

Introduisons pour le reste du paragraphe les constantes suivantes, appelées parfois *facteurs de Stein* :

$$\begin{aligned} k_1(\lambda) &= 1 \wedge \sqrt{\frac{2}{\lambda e}}, \\ k_2(\lambda) &= \frac{1 - e^{-\lambda}}{\lambda}. \end{aligned}$$

Voici une première proposition. Notons $\|f\| = \sup_{i \in \mathbb{N}} |f(i)|$ la norme uniforme, et $\Delta f(k) = f(k+1) - f(k)$.

PROPOSITION 6.2 *Soient $h \in \mathcal{X}_0$ et f la solution de l'équation de Stein (6.10). On a :*

1. $\|f\| \leq k_1(\lambda)(\sup_i h(i) - \inf_i h(i))$
2. $\|\Delta f\| \leq k_2(\lambda)(\sup_i h(i) - \inf_i h(i))$

Démonstration :

* Considérons déjà un cas particulier, celui où $h = 1_{\{k\}}$ avec $k \in \mathbb{N}$. On note alors $f_{\{k\}}$ la transformée de Stein associée. Soit $i \in \mathbb{N}$. Distinguons les cas selon la position de i par rapport à k .

Si $i < k$ alors on développe $f_{\{k\}}$ grâce à (6.11), ce qui donne :

$$f_{\{k\}}(i) = -e^{-\lambda} \frac{\lambda^k}{k!} \frac{(i-1)!}{\lambda^i} \sum_{j=0}^{i-1} \frac{\lambda^j}{j!}.$$

Or $-e^{-\lambda} \frac{\lambda^k}{k!}$ est toujours négatif. De plus :

$$\begin{aligned} \frac{(i-1)!}{\lambda^i} \sum_{j=0}^{i-1} \frac{\lambda^j}{j!} &= \sum_{j=0}^{i-1} \frac{(i-1) \cdots (i-j+1)}{\lambda^{i-j}} \\ &\leq \sum_{j=0}^i \frac{i(i-1) \cdots (i-j+1)}{\lambda^{i+1-j}}. \end{aligned}$$

Cela signifie que pour $i < k$, $f_{\{k\}}$ est décroissante en i et donc $\Delta f_{\{k\}}(i) \leq 0$.

Si $i > k$ alors on développe $f_{\{k\}}$ grâce à (6.12), ce qui donne :

$$f_{\{k\}}(i) = e^{-\lambda} \frac{\lambda^k}{k!} \frac{(i-1)!}{\lambda^i} \sum_{j=k}^{\infty} \frac{\lambda^j}{j!}.$$

Or $e^{-\lambda} \frac{\lambda^k}{k!}$ est toujours positif. De plus :

$$\begin{aligned} \frac{(i-1)!}{\lambda^i} \sum_{j=i}^{\infty} \frac{\lambda^j}{j!} &= \sum_{j=i}^{\infty} \frac{\lambda^{j-i}}{i(i+1) \cdots j} \\ &\geq \sum_{j=i+1}^{\infty} \frac{\lambda^{j-i-1}}{(i+1) \cdots j}. \end{aligned}$$

Ce qui signifie encore que pour $i > k$, $f_{\{k\}}$ est décroissante en i et donc $\Delta f_{\{k\}}(i) \leq 0$.

Si $i = k$ alors on a par calcul explicite, en utilisant (6.12) pour $f_{\{k\}}(k+1)$ et (6.11) pour $f_{\{k\}}(k)$:

$$\begin{aligned}
\Delta f_{\{k\}}(k) &= e^{-\lambda} \left(\frac{k!}{\lambda^{k+1}} \frac{\lambda^k}{k!} \sum_{j=k+1}^{\infty} \frac{\lambda^j}{j!} + \frac{(k-1)!}{\lambda^k} \frac{\lambda^k}{k!} \sum_{j=0}^{k-1} \frac{\lambda^j}{j!} \right) \\
&= \frac{e^{-\lambda}}{\lambda} \left(\sum_{j=k+1}^{\infty} \frac{\lambda^j}{j!} + \frac{\lambda}{k} \sum_{j=0}^{k-1} \frac{\lambda^j}{j!} \right) \\
&= \frac{e^{-\lambda}}{\lambda} \left(\sum_{j=k+1}^{\infty} \frac{\lambda^j}{j!} + \frac{1}{k} \sum_{j=1}^k \frac{\lambda^j}{(j-1)!} \right) \\
&\leq \frac{e^{-\lambda}}{\lambda} \left(\sum_{j=1}^{\infty} \frac{\lambda^j}{j!} \right) = \frac{1 - e^{-\lambda}}{\lambda}.
\end{aligned}$$

Il s'agit de l'inégalité voulue pour $f_{\{k\}}$.

* Considérons maintenant $h \in \mathcal{X}_0$ quelconque. On ne change pas l'équation de Stein (6.10) en remplaçant h par $h_+ = h - \inf_i h(i) \geq 0$. On peut donc considérer h positive. Pour f la transformée de Stein associée, remarquons que l'on a pour $i \in \mathbb{N}$:

$$\begin{aligned}
\sum_{k=0}^{\infty} h(k) f_{\{k\}}(i) &= \sum_{k=0}^{\infty} h(k) \frac{(i-1)!}{\lambda^i} \sum_{j=0}^{i-1} \left(\delta_{kj} - \frac{\lambda^k e^{-\lambda}}{k!} \right) \frac{\lambda^j}{j!} \\
&= \frac{(i-1)!}{\lambda^i} \sum_{j=0}^{i-1} \frac{\lambda^j}{j!} \left(\sum_{k=0}^{\infty} h(k) \delta_{kj} - \sum_{k=0}^{\infty} h(k) \frac{\lambda^k e^{-\lambda}}{k!} \right) \\
&= \frac{(i-1)!}{\lambda^i} \sum_{j=0}^{i-1} \frac{\lambda^j}{j!} \left(h(j) - \int h d\mu_0 \right) \\
&= f(i).
\end{aligned}$$

L'échange des deux signes somme est justifié par le fait que si on note $M = \|h\|$ alors :

$$\begin{aligned}
\frac{(i-1)!}{\lambda^i} \sum_{j=0}^{i-1} \sum_{k=0}^{\infty} \left(h(k) \left(\delta_{kj} - \frac{\lambda^k e^{-\lambda}}{k!} \right) \frac{\lambda^j}{j!} \right) &\leq \frac{(i-1)!}{\lambda^i} 2M \sum_{j=0}^{i-1} \frac{\lambda^j}{j!} \\
&\leq \frac{(i-1)!}{\lambda^i} e^{\lambda} 2M < +\infty.
\end{aligned}$$

Finalement on majore chaque $\Delta f_{\{k\}}(i)$ par 0 sauf si $i = k$, on majore par $k_2(\lambda)$. On obtient ainsi :

$$\begin{aligned}
\Delta f(i) &= \sum_{k=0}^{\infty} h(k) (f_{\{k\}}(i+1) - f_{\{k\}}(i)) \\
&\leq k_2(\lambda) (\sup_i h(i)),
\end{aligned}$$

et comme on avait pris h positive, pour h quelconque on obtient :

$$\sup_i \Delta f(i) \leq k_2(\lambda) (\sup_i h(i) - \inf_i h(i)).$$

En remplaçant h par $-h$, la solution de l'équation de Stein devient $-f$ grâce à la linéarité de T_0 et on obtient de la même façon :

$$\sup_i \Delta f(i) \geq -k_2(\lambda) (\sup_i h(i) - \inf_i h(i)).$$

En conclusion on a bien l'inégalité voulue à savoir :

$$\|\Delta f\| \leq k_2(\lambda)(\sup_i h(i) - \inf_i h(i)).$$

■

Passons maintenant au vif du sujet. On considère un ensemble discret fini, disons $\Gamma = \{1, \dots, n\}$. Soit $\{X_i\}_{i \in \Gamma}$ un ensemble fini de variables aléatoires qui suivent chacune une loi de Bernoulli de paramètre p_i . Notons leur somme $W = \sum_{i \in \Gamma} X_i$ et μ la loi de W . On note aussi $\lambda = \sum_{i \in \Gamma} p_i = \mathbb{E}(W)$.

Rappelons que $\|\mu - \mu_0\|_{VT} = \sup_{A \subset \mathbb{N}} |\mu(A) - \mu_0(A)|$. Ainsi si $A \subset \mathbb{N}$ et $h = 1_A$, f_A la transformée de Stein associée, alors on a par le théorème du transfert :

$$\mu(A) - \mu_0(A) = \int_{\mathbb{N}} T_0 f_A d\mu = \int_{\mathbb{N}} \lambda f_A(i+1) - i f_A(i) d\mu(i) = \mathbb{E}(\lambda f_A(W+1) - W f_A(W)).$$

Et donc :

$$\boxed{\|\mu - \mu_0\|_{VT} = \sup_{A \subset \mathbb{N}} |\mathbb{E}(\lambda f_A(W+1) - W f_A(W))|}$$

On va alors pouvoir par la suite, se concentrer sur $|\mathbb{E}(\lambda f_A(W+1) - W f_A(W))|$. Énonçons maintenant le théorème suivant, établi par Chen (1975).

THÉORÈME 6.1 *Pour chaque $i \in \Gamma$, on se donne une partition de $\Gamma \setminus \{i\} = \Gamma_i^s \sqcup \Gamma_i^w$. On pose alors $Z_i = \sum_{j \in \Gamma_i^s} X_j$ et $W_i = \sum_{j \in \Gamma_i^w} X_j$. Alors :*

$$\|\mu - \mu_0\|_{VT} \leq k_2(\lambda) \sum_{i \in \Gamma} (p_i \mathbb{E}(X_i + Z_i) + \mathbb{E}(X_i Z_i)) + k_1(\lambda) \sum_{i \in \Gamma} \mathbb{E}|p_i - \mathbb{E}(X_i | W_i)|.$$

Remarque : Dans l'énoncé de ce Théorème 6.1, la partition de $\Gamma \setminus \{i\}$ est quelconque, mais dans la pratique on s'arrangera en général à partitionner de la façon suivante : $\Gamma_i^s = \{j \in \Gamma \setminus \{i\}, X_j \text{ dépend fortement de } X_i\}$ et $\Gamma_i^w = (\Gamma \setminus \{i\}) \setminus \Gamma_i^s$. En particulier quand les X_j sont toutes indépendantes, on a $\Gamma_i^s = \emptyset$, $\mathbb{E}(X_j | W_j) = \mathbb{E}(X_j)$ et le résultat du théorème se réduit à :

$$\|\mu - \mu_0\|_{VT} \leq k_2(\lambda) \sum_{i \in \Gamma} p_i^2.$$

Un exemple d'application de ce théorème est donné par le *Problème de la date d'anniversaire*.

Démonstration :

Soit $A \subset \mathbb{N}$. On a en gardant toujours les mêmes notations :

$$\begin{aligned} \mathbb{E}(\lambda f_A(W+1) - W f_A(W)) &= \sum_{i \in \Gamma} \mathbb{E}(p_i f_A(W+1) - X_i f_A(W)) \\ &= \sum_{i \in \Gamma} \mathbb{E}(p_i f_A(W+1) - p_i f_A(W_i+1) + p_i f_A(W_i+1) \\ &\quad - X_i f_A(W_i+1) + X_i f_A(W_i+1) - X_i f_A(W)). \end{aligned}$$

Et pour $i \in \Gamma$, comme f_A prend des valeurs entières, on a :

$$\begin{aligned} |f_A(W+1) - f_A(W_i+1)| &\leq \|\Delta f_A\| (W - W_i) = \|\Delta f_A\| (X_i + Z_i), \\ |X_i f_A(W_i+1) - X_i f_A(W)| &\leq \|\Delta f_A\| |X_i| |W_i+1 - W| = \|\Delta f_A\| X_i Z_i, \end{aligned}$$

puis pour $p_i f_A(W_i + 1) - X_i f_A(W_i + 1)$, on regarde l'espérance de l'espérance conditionnelle par rapport à W_i :

$$\begin{aligned} |\mathbb{E}(p_i f_A(W_i + 1) - X_i f_A(W_i + 1))| &= |\mathbb{E}(\mathbb{E}(p_i f_A(W_i + 1) - X_i f_A(W_i + 1)|W_i))| \\ &= |\mathbb{E}(f_A(W_i + 1)(p_i - \mathbb{E}(X_i|W_i)))| \\ &\leq \|f_A\| \mathbb{E}|p_i - \mathbb{E}(X_i|W_i)|. \end{aligned}$$

On obtient ainsi :

$$|\mathbb{E}(\lambda f_A(W + 1) - W f_A(W))| \leq \|\Delta f_A\| \sum_{i \in \Gamma} (p_i \mathbb{E}(X_i + Z_i) + \mathbb{E}(X_i Z_i)) + \|f_A\| \sum_{i \in \Gamma} |p_i - \mathbb{E}(X_i|W_i)|.$$

Le résultat découle de cette inégalité combinée à la Proposition 6.2, en remarquant qu'ici h est une indicatrice, on a $\sup_i h(i) - \inf_i h(i) = 1$ ■

Après cette approche locale, regardons celle qui motive notre étude de la méthode de Stein : l'*approche par couplage*, qui permet comme le précédent Théorème de majorer la variation totale entre une mesure μ et une loi de Poisson. On garde les mêmes notations que ci-dessus.

THÉORÈME 6.2 *On se donne $W = \sum_{i \in \Gamma} X_i$ et pour $i \in \Gamma$, les ensembles Γ_i^s , Γ_i^w et les variables aléatoires $Z_i = \sum_{j \in \Gamma_i^s} X_j$, $W_i = \sum_{j \in \Gamma_i^w} X_j$ comme considérés plus haut. Soient alors \widetilde{W}_i^1 et W_i^1 de lois respectives $\mathcal{L}(W_i|X_i = 1)$ et $\mathcal{L}(W_i)$ définies sur le même espace de probabilité. Alors on a :*

$$\|\mathcal{L}(W) - \mathcal{P}(\lambda)\| \leq k_2(\lambda) \sum_{i \in \Gamma} (p_i \mathbb{E}(X_i + Z_i) \mathbb{E}(X_i Z_i)) + k_2(\lambda) \sum_{i \in \Gamma} p_i \mathbb{E}|W_i^1 - \widetilde{W}_i^1|.$$

Remarque : Nous verrons une application directe de ce résultat dans le deuxième exemple de la sous-partie suivante.

Démonstration :

On procède de façon similaire à la démonstration du Théorème précédent. Soient $A \subset \mathbb{N}$, $h = 1_A$ et f_A la transformée de Stein associée. On écrit alors :

$$\begin{aligned} \mathbb{E}(\lambda f_A(W + 1) - W f_A(W)) &\leq \|\Delta f_A\| \sum_{i \in \Gamma} (p_i \mathbb{E}(X_i + Z_i) + \mathbb{E}(X_i Z_i)) + \\ &\quad \sum_{i \in \Gamma} \mathbb{E}(p_i f_A(W_i + 1) - X_i f_A(W_i + 1)). \end{aligned}$$

Mais cette fois le second terme du membre de droite, nous allons le majorer de la façon suivante :

$$\begin{aligned} |\mathbb{E}(p_i f_A(W_i + 1) - X_i f_A(W_i + 1))| &= |p_i \mathbb{E}(f_A(W_i + 1)) - \mathbb{E}(X_i) \mathbb{E}(f_A(W_i + 1)|X_i = 1)| \\ &= |p_i (\mathbb{E}(f_A(W_i + 1)) - \mathbb{E}(f_A(W_i + 1)|X_i = 1))| \\ &= |p_i \mathbb{E}(f_A(W_i^1) - f_A(\widetilde{W}_i^1 + 1))| \\ &\leq p_i \|\Delta f_A\| \mathbb{E}|W_i^1 - \widetilde{W}_i^1| \end{aligned}$$

Le résultat en découle en appliquant la Proposition 6.2 ■

Remarque : Dans l'exemple *Un problème classique d'occupation* qui applique directement ce résultat, nous avons à faire à des variables de Bernoulli qui vérifient $\widetilde{X}_{i,j} \leq X_j$ pour tout j (voir les notations ci-après), ce qui nous verrons, simplifiera considérablement nos calculs. Il semble donc important d'introduire cette notion plus formellement.

Soient toujours pareil X_1, \dots, X_n variables de Bernoulli et $W = \sum_{i \in \Gamma} X_i$. Pour $i \in \Gamma$, on considère $\{\widetilde{X}_{i,j}, j \neq i\}$ et $\{X_{i,j}, j \neq i\}$ des variables aléatoires sur le même espace de probabilité telles que

$$\mathcal{L}(\widetilde{X}_{i,j}, j \neq i) = \mathcal{L}(X_j, j \neq i | X_i = 1), \quad (6.13)$$

$$\mathcal{L}(X_{i,j}, j \neq i) = \mathcal{L}(X_j, j \neq i). \quad (6.14)$$

On décompose $\Gamma \setminus \{i\}$ en sous ensembles $\Gamma_i^+ = \{j \in \Gamma; \tilde{X}_{i,j} \geq X_{i,j}\}$, $\Gamma_i^- = \{j \in \Gamma; \tilde{X}_{i,j} \leq X_{i,j}\}$ et Γ_i^0 .

Définition 6.1 Les variables X_1, \dots, X_n sont dites :

- positivement liées si $\Gamma_i^- = \Gamma_i^0 = \emptyset$
- négativement liées si $\Gamma_i^+ = \Gamma_i^0 = \emptyset$

Voici un Lemme qui caractérise les variables *positivement liées* et celles *négativement liées*.

LEMME 6.2 Les variables de Bernoulli X_1, \dots, X_n sont positivement (resp. négativement) liées si et seulement si, pour tout $i \in \Gamma$ et toute fonction croissante $\varphi : \{0, 1\}^{n-1} \rightarrow \{0, 1\}$ on a :

$$\mathbb{E}(\varphi(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n) | X_i = 1) \geq (\text{resp. } \leq) \mathbb{E}(\varphi(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)).$$

Enonçons à présent le dernier Théorème de cette section.

THÉORÈME 6.3 Dans ces mêmes conditions, on a la majoration suivante avec $\lambda = \mathbb{E}(W)$:

$$\begin{aligned} \|\mathcal{L}(W) - \mathcal{P}(\lambda)\|_{VT} &\leq k_2(\lambda) \left(\sum_{i \in \Gamma} p_i^2 + \sum_{i \in \Gamma} \sum_{j \in \Gamma_i^+} \text{Cov}(X_i, X_j) \right) \\ &\quad + k_2(\lambda) \left(\sum_{i \in \Gamma} \sum_{j \in \Gamma_i^-} |\text{Cov}(X_i, X_j)| + \sum_{i \in \Gamma} \sum_{j \in \Gamma_i^0} (\mathbb{E}(X_i X_j) + p_i p_j) \right). \end{aligned}$$

En particulier, si X_1, \dots, X_n sont positivement liées alors :

$$\|\mathcal{L}(W) - \mathcal{P}(\lambda)\|_{VT} \leq k_2(\lambda) \left(\text{Var}(W) - \lambda + 2 \sum_{i \in \Gamma} p_i^2 \right).$$

Et si X_1, \dots, X_n sont négativement liées alors :

$$\|\mathcal{L}(W) - \mathcal{P}(\lambda)\|_{VT} \leq k_2(\lambda) \left(\lambda - \text{Var}(W) \right).$$

Remarque : Une application directe de ce Théorème se place dans le troisième exemple ci-dessous.

Démonstration :

On se ramène aux hypothèses du Théorème 6.2. Ainsi pour $i \in \Gamma$, on choisit $\Gamma_i^s = \emptyset$, $\Gamma_i^w = \Gamma \setminus \{i\}$ et $\tilde{W}_i = \sum_{j \neq i} \tilde{X}_{i,j}$. Dans ce cas, le Théorème 6.2 nous donne la majoration :

$$\|\mathcal{L}(W) - \mathcal{P}(\lambda)\|_{VT} \leq k_2(\lambda) \sum_{i \in \Gamma} \left(p_i \left(\mathbb{E}(X_i) + \mathbb{E}|W_i - \tilde{W}_i| \right) \right). \quad (6.15)$$

De plus on écrit :

$$\begin{aligned} p_i \mathbb{E}|W_i^1 - \tilde{W}_i^1| &= p_i \mathbb{E} \left| \sum_{j \in \Gamma_i^w} (X_j - \tilde{X}_{i,j}^1) \right| \\ &\leq p_i \mathbb{E} \left(\sum_{j \in \Gamma_i^+} (\tilde{X}_{i,j} - X_j) \right) + p_i \mathbb{E} \left(\sum_{j \in \Gamma_i^-} (X_j - \tilde{X}_{i,j}) \right) + p_i \mathbb{E} \left(\sum_{j \in \Gamma_i^0} (\tilde{X}_{i,j} + X_j) \right). \end{aligned}$$

Mais on a : $p_i \mathbb{E}(\tilde{X}_{i,j}) = p_i \mathbb{E}(X_j | X_i = 1) = \mathbb{E}(1_{X_i=1} X_j) = \mathbb{E}(1_{X_i=1} X_i X_j) = \mathbb{E}(X_i X_j)$ et donc : $p_i \mathbb{E}(\tilde{X}_{i,j} - X_j) = \mathbb{E}(X_i X_j) - \mathbb{E}(X_i)(X_j) = \text{Cov}(X_i, X_j)$. On en déduit que :

$$p_i \mathbb{E}|W_i^1 - \tilde{W}_i^1| \leq \sum_{j \in \Gamma_i^+} \text{Cov}(X_i, X_j) + \sum_{j \in \Gamma_i^-} |\text{Cov}(X_i, X_j)| + \sum_{j \in \Gamma_i^0} (\mathbb{E}(X_i X_j) + p_i p_j).$$

Et le résultat en découle en appliquant cette majoration dans (6.15) ■

Nous pouvons maintenant nous concentrer sur trois exemples qui reflètent ce que l'on vient d'établir.

6.3 Quelques applications

6.3.1 Un problème de la date d'anniversaire

Considérons n personnes prises au hasard de sorte que leur date d'anniversaire soit aléatoire. Soit W le nombre de paires de personnes qui ont la même date d'anniversaire. Si μ est la loi de W , et $\lambda = \mathbb{E}(W) = \frac{1}{365} \binom{n}{2}$ alors on a :

$$\|\mu - \mathcal{P}(\lambda)\|_{VT} \leq \frac{8\lambda(1 - e^{-\lambda})}{n - 1}.$$

Démonstration :

Prenons $\Gamma = \{(i_1, i_2) \subset \{1, \dots, n\}\}$ l'ensemble des sous-ensembles de $\{1, \dots, n\}$ à 2 éléments. On regarde les variables X_i où $i = \{i_1, i_2\}$ qui associent 1 si les personnes i_1 et i_2 ont la même date d'anniversaire, et 0 sinon. Il découle que $\mathbb{E}(X_i) = \frac{1}{365}$ pour tout $i \in \Gamma$. Ainsi avec le W introduit en haut, on voit que $W = \sum_{i \in \Gamma} X_i$. Pour $i \in \Gamma$ fixé, on choisit $\Gamma_i^s = \{j \in \Gamma, i \cap j \neq \emptyset\}$. De plus $\Gamma_i^w = \{j \in \Gamma, i \cap j = \emptyset\}$ et on remarque que si $j \in \Gamma_i^w$ alors X_j est indépendante à X_i , c'est-à-dire que le deuxième terme du membre de droite de l'inégalité du Théorème 6.1 est nul. En appliquant ce Théorème et en respectant les notations, cela nous donne la majoration :

$$\begin{aligned} \|\mu - \mathcal{P}(\lambda)\|_{VT} &\leq k_2(\lambda) \sum_{i \in \Gamma} (p_i \mathbb{E}(X_i + Z_i) + \mathbb{E}(X_i Z_i)) \\ &= k_2(\lambda) \binom{n}{2} \left(\frac{2(n-1) + 1}{365^2} + \frac{2(n-1)}{365^2} \right) \\ &= k_2(\lambda) \binom{n}{2} \frac{4n-3}{365^2} \leq \frac{8\lambda(1 - e^{-\lambda})}{n-1}. \end{aligned}$$

■

Remarque : Notons que le problème peut être rendu général, en considérant n boules tirés indépendamment de d boîtes ayant la même probabilité d'apparition. On a dans ce cas $\mathbb{E}(W) = \frac{1}{d} \binom{n}{2}$ et le même résultat.

6.3.2 Un problème classique d'occupation

On distribue r boules de manière indépendante dans n boîtes équiprobables. Si W est le nombre de boîtes vides après cette distribution, alors on a pour $\lambda = \mathbb{E}(W)$:

$$\|\mathcal{L}(W) - \mathcal{P}(\lambda)\|_{VT} \leq \left(1 - \exp \left\{ -n \left(\frac{n-1}{n} \right)^r \right\} \right) \left(n \left(\frac{n-1}{n} \right)^r - (n-1) \left(\frac{n-2}{n-1} \right)^r \right).$$

De plus si $r = na_n$ avec $\lim_{n \rightarrow \infty} a_n = \infty$ alors :

$$\|\mathcal{L}(W) - \mathcal{P}(\lambda)\|_{VT} = O_{n \rightarrow \infty}(a_n e^{-a_n}).$$

Et si $a_n = \log(n) - \log(c)$ alors $\lim_{n \rightarrow \infty} \lambda = c$.

Démonstration :

Modélisons le problème énoncé. Numérotons les boules de 1 à m et notons pour $i = 1, \dots, r$ la variable T_i qui représente le numéro de la boîte dans laquelle on envoie la boule i . Ainsi les T_1, \dots, T_r sont i.i.d et de loi $\mathcal{U}(\{1, \dots, n\})$ par hypothèse. Pour $i \in \Gamma$, on considère $S_i = \sum_{j=1}^n 1_{\{T_j=i\}}$ le nombre de boules dans la boîte i . On considère X_i la variable aléatoire qui vaut 1 si la boîte i est vide et 0 sinon. Ainsi :

$$X_i = 1_{\{S_i=0\}} = \prod_{j=1}^n 1_{\{T_j \neq i\}}.$$

Par indépendance des T_j , la probabilité d'envoyer les r boules ailleurs que dans la boîte i est $p_i := \left(\frac{n-1}{n}\right)^r$. Ainsi X_i suit la loi $\mathcal{B}(p_i)$ et p_i ne dépend en fait pas de i . On a par définition $W = \sum_{i \in \Gamma} X_i$ et $\lambda = \mathbb{E}(W) = \sum_{i \in \Gamma} p_i = n \left(\frac{n-1}{n}\right)^r$. Choisissons maintenant dans le Théorème 6.2, $\Gamma_i^s = \emptyset$ et alors $\Gamma_i^w = \Gamma_i \setminus \{i\}$. Fixons $i \in \Gamma$, et considérons l'action suivante :

(*) *Prendre les boules qui sont dans la boîte i , et de les disperser uniformément dans les autres boîtes.* Sur $S_i = l$ avec $0 \leq l \leq r$ (c'est-à-dire qu'après la première distribution, la boîte i contient l boules) on considère \tilde{T}_j la variable qui représente le numéro de la boîte où on envoie la boule j . Ainsi pour $j \in \Gamma_i^w$ on a :

$$\mathcal{L}(\tilde{T}_j | S_i = l) = \mathcal{U}(\{1, \dots, n\} \setminus \{i\}).$$

De plus les \tilde{T}_j sachant $S_i = l$ sont indépendantes entre elles, mais aussi indépendantes des T_1, \dots, T_r . On considère maintenant pour tout $j \in \Gamma_i^w$, la variable aléatoire $\tilde{X}_{i,j}$ qui vaut 1 si la boîte j est vide, après avoir réalisé (*), et 0 sinon. Ainsi si $\tilde{S}_j^i = \sum_{k=0}^l 1_{\{\tilde{T}_k=j\}}$ est le nombre de boules dans j après (*), alors :

$$\tilde{X}_{i,j} = 1_{\tilde{S}_j^i=0}.$$

On peut donc écrire pour tout $\Gamma' \subset \Gamma_i^w$:

$$P(\tilde{X}_{i,j}^1 = 1, \forall j \in \Gamma') = \left(\frac{n - |\Gamma'| - 1}{n - 1}\right)^r = P(X_j = 1 | X_i = 1, \forall j \in \Gamma').$$

Cela signifie :

$$\mathcal{L}(\tilde{X}_{i,j}^1, j \in \Gamma_i^w) = \mathcal{L}(X_j, j \in \Gamma_i^w | X_i = 1). \quad (6.16)$$

Notons enfin $\tilde{W}_i^1 = \sum_{j \in \Gamma_i^w} \tilde{X}_{i,j}$ et $W_i^1 = W_i = \sum_{j \in \Gamma_i^w} X_j$. Et (6.16) nous implique :

$$\mathcal{L}(\tilde{W}_i^1) = \mathcal{L}(W_i | X_i = 1).$$

Comme $\tilde{X}_{i,j}^1 = 1 \Rightarrow X_j = 1$ pour tout $j \in \Gamma_i^w$, il est clair que $\tilde{X}_{i,j}^1 \leq X_j$ pour tout $j \in \Gamma_i^w$. Ainsi par le Théorème 6.2 :

$$\begin{aligned} \|\mathcal{L}(W) - \mathcal{P}(\lambda)\|_{VT} &\leq k_2(\lambda) \left(\sum_{i \in \Gamma} p_i \mathbb{E}(X_i) + \sum_{i \in \Gamma} p_i \mathbb{E}|W_i^1 - \tilde{W}_i^1| \right) \\ &= k_2(\lambda) \sum_{i \in \Gamma} p_i \left(p_i + \mathbb{E}|W_i^1 - \tilde{W}_i^1| \right) \\ &= k_2(\lambda) \sum_{i \in \Gamma} p_i \mathbb{E} \left(X_i + \sum_{j \in \Gamma_i^w} (X_j - \tilde{X}_{i,j}^1) \right) \\ &= k_2(\lambda) \sum_{i \in \Gamma} p_i \mathbb{E} \left(W - \sum_{j \in \Gamma_i^w} \tilde{X}_{i,j}^1 \right) \\ &= k_2(\lambda) \left(\mathbb{E}(W)^2 - \sum_{i \in \Gamma} \sum_{j \in \Gamma_i^w} \mathbb{E}(X_i) \mathbb{E}(X_i X_j) / P(1_{X_i}) \right) \\ &= k_2(\lambda) \left(\mathbb{E}(W)^2 - \sum_{i \in \Gamma} \sum_{j \in \Gamma_i^w} \mathbb{E}(X_i X_j) \right). \end{aligned}$$

De plus on sait que :

$$\begin{aligned} \mathbb{E}(X_i X_j) &= \int_{\{0,1\}} \left(\int_{\{0,1\}} xy dP(y | X_i = x) \right) dP(x) \\ &= \int_{\{0,1\}} x P(X_j = 1 | X_i = x) dP(x) \\ &= P(X_j = 1 | X_i = 1) P(X_i = 1) = \left(\frac{n-2}{n-1}\right)^r \left(\frac{n-1}{n}\right)^r. \end{aligned}$$

Et finalement :

$$\begin{aligned} \|\mathcal{L}(W) - \mathcal{P}(\lambda)\|_{VT} &\leq k_2(\lambda) \left(\mathbb{E}(W)^2 - n(n-1) \left(\frac{n-2}{n-1} \right)^r \left(\frac{n-1}{n} \right)^r \right) \\ &= \left(1 - \exp \left\{ -n \left(\frac{n-1}{n} \right)^r \right\} \right) \left(n \left(\frac{n-1}{n} \right)^r - (n-1) \left(\frac{n-2}{n-1} \right)^r \right). \end{aligned}$$

Dans le cas où $r = na_n$ avec $a_n = \log(n) - \log(c)$, on a :

$$\begin{aligned} \lambda &= n \left(\frac{n-1}{n} \right)^r \\ &= n \left(\frac{n-1}{n} \right)^{n(\log(\frac{n}{c}))} \\ &= n \exp \left\{ \log\left(\frac{n}{c}\right) \log \left(1 - \frac{1}{n} \right) n \right\} \\ &\sim_{n \rightarrow \infty} n \exp \left\{ \log\left(\frac{c}{n}\right) \right\} = c. \end{aligned}$$

■

Remarque : Dans cet exemple, les variables X_1, \dots, X_n sont négativement liées.

6.3.3 Approximation de la loi hypergéométrique

On dispose de N boîtes dont n rouges ($n \leq N$), les autres noires, et de m boules ($m \leq N$). On suppose que chaque boîte peut contenir au plus une boule. On distribue les m boules uniformément à l'intérieur des N boîtes. On désigne par W le nombre de boîtes rouges remplies (de 1 boule) après la distribution. Alors W suit une loi hypergéométrique et avec $\lambda = \mathbb{E}(W)$, on a :

$$\|\mathcal{L}(W) - \mathcal{P}(\lambda)\|_{VT} \leq (1 - \exp(-\lambda)) \frac{n+m}{N-1}.$$

Démonstration :

Comme dans l'exemple précédent, modélisons le problème donné, de telle sorte qu'on se retrouve dans les hypothèses du Théorème 6.3. Soit $\Gamma = \{1, \dots, N\}$. Numérotons les m boules de 1 à m et les N boîtes de 1 à N de telle sorte que les boîtes $1, \dots, n$ sont exactement les boîtes rouges. On parlera de configuration $(\epsilon_1, \dots, \epsilon_N) \in \{0, 1\}^N$ où ϵ_i vaut 1 si la boîte i est remplie et 0 sinon. L'ensemble des configurations possibles est $\{(\epsilon_1, \dots, \epsilon_N) \in \{0, 1\}^N, \sum_{i=1}^N \epsilon_i = m\}$. On peut l'identifier avec l'ensemble des boîtes qui sont pleines i.e. $\mathcal{P}_m(\{1, \dots, N\})$. Dans toute la suite, on notera donc indifféremment (et avec abus) la configuration $C = (\epsilon_1, \dots, \epsilon_N) = \{i \in \Gamma, \exists k \in \{1, \dots, m\}, T_k = i\}$. On notera aussi pour alléger les notations $\mathcal{I}_k = \{(i_1, \dots, i_k) \in \Gamma^k, 2 \text{ à } 2 \text{ distincts}\}$ où $k = 1, \dots, m$, et pour $A \in \mathcal{P}_k(\{1, \dots, N\})$, $\mathcal{I}_k(A) = \{(i_1, \dots, i_k) \in \Gamma^k, 2 \text{ à } 2 \text{ distincts}, \{i_1, \dots, i_k\} = A\}$. On remarque immédiatement que $\text{Card}(\mathcal{I}_k(A)) = k!$.

Plaçons les boules l'une après l'autre et désignons par T_k où $k \in \{1, \dots, m\}$, la variable aléatoire qui prend pour valeur le numéro de la boîte où la boule k est envoyée. Ainsi on a pour tout $k = 1, \dots, m-1$:

$$\begin{aligned} \mathcal{L}(T_1) &= \mathcal{U}(\{1, \dots, N\}) \\ \forall i_1 \in \mathcal{I}_1 \quad \mathcal{L}(T_2|T_1 = i_1) &= \mathcal{U}(\{1, \dots, N\} \setminus \{i_1\}) \\ &\dots \\ \forall (i_1, \dots, i_k) \in \mathcal{I}_k \quad \mathcal{L}(T_{k+1}|T_1 = i_1, \dots, T_k = i_k) &= \mathcal{U}(\{1, \dots, N\} \setminus \{i_1, \dots, i_k\}) \end{aligned}$$

Le vecteur aléatoire (T_1, \dots, T_m) suit une loi uniforme sur \mathcal{I}_m . En effet pour $(i_1, \dots, i_m) \in \mathcal{I}_m$:

$$\begin{aligned} \mathbb{P}(T_1 = i_1, \dots, T_m = i_m) &= \mathbb{P}(T_1 = i_1)\mathbb{P}(T_2 = i_2|T_1 = i_1) \cdots \mathbb{P}(T_m = i_m|T_1 = i_1, \dots, T_{m-1} = i_{m-1}) \\ &= \frac{1}{N} \frac{1}{N-1} \cdots \frac{1}{N-(m-1)} = \frac{(N-m)!}{N!}. \end{aligned}$$

On voit en effectuant cette distribution que chaque configuration est prise de manière uniforme sur $\mathcal{P}_m\{1, \dots, N\}$. En effet pour $A \in \mathcal{P}_m(\{1, \dots, N\})$:

$$\begin{aligned} \mathbb{P}(\{T_1, \dots, T_m\} = A) &= \mathbb{P}((T_1, \dots, T_m) = (i_1, \dots, i_m); \underline{i} := (i_1, \dots, i_m) \in \mathcal{I}_m(A)) \\ &= \sum_{\underline{i} \in \mathcal{I}_m(A)} \mathbb{P}(T_1 = i_1, \dots, T_m = i_m) \\ &= \sum_{\underline{i} \in \mathcal{I}_m(A)} \frac{(N-m)!}{N!} = \frac{1}{\binom{N}{m}}. \end{aligned}$$

En particulier cela implique que chaque T_k suit une loi uniforme sur $\{1, \dots, N\}$. On considère pour $i \in \Gamma$ la variable aléatoire X_j qui vaut 1 s'il y a une boule dans la boîte j (i.e si $j \in C$) et 0 sinon. Alors $X_j = \sum_{k=1}^m 1_{T_k=j}$ suit une loi de Bernoulli de paramètre $p = \mathbb{E}(X_j) = \sum_{k=1}^m \mathbb{P}(T_k = j) = \frac{m}{N}$. Par l'énoncé, W est le nombre de boîtes rouges remplies après la distribution. Ainsi $W = \sum_{i=1}^n X_i$. Et on a immédiatement $\lambda = \mathbb{E}(W) = np$. Montrons que W suit une loi hypergéométrique. Pour $k = 1, \dots, n$ on a par uniformité des configurations :

$$\begin{aligned} \mathbb{P}(W = k) &= \mathbb{P}(\text{Card}(C \cap \{1, \dots, n\}) = k) \\ &= \mathbb{P}(C = A \sqcup B, A \in \mathcal{P}_k(\{1, \dots, n\}), B \in \mathcal{P}_{m-k}(\{1, \dots, N-n\})) \\ &= \frac{\binom{n}{k} \binom{N-n}{m-k}}{\binom{N}{m}}. \end{aligned}$$

Cela signifie que W suit la loi hypergéométrique $\mathcal{H}(m, \frac{n}{N}, N) = \mathcal{H}(n, p, N)$, dont on sait que la variance est $\text{Var}(W) = np(1-p)\frac{N-n}{N-1}$. Montrons maintenant que les X_1, \dots, X_n sont négativement liées, afin de pouvoir appliquer une version du Théorème 6.3. Pour cela, considérons une deuxième manière de distribuer les boules, et fixons $i \in \Gamma$ dans toute la suite.

* Construction des variables $\tilde{X}_{i,j}^1$.

On note dans la même idée que plus haut, pour $k = 1, \dots, m$, $\mathcal{I}_k^i = \{(i_1, \dots, i_k) \in (\Gamma \setminus \{i\})^k, 2 \text{ à } 2 \text{ distincts}\}$. Considérons les actions suivantes :

(*) *On choisit de manière uniforme une boule $l \in \{1, \dots, m\}$.*

(**) *On place la boule l dans la boîte i , et on distribue les $m-1$ restantes dans les $N-1$ autres boîtes de manière uniforme.*

Soit alors B la variable aléatoire qui a pour valeur le numéro de la boule tirée en (*). Ainsi $\mathcal{L}(B) = \mathcal{U}(\{1, \dots, m\})$. Pour $k = 1, \dots, m$, on désigne par T'_k la variable aléatoire qui a pour valeur le numéro de la boîte où on envoie la boule k . Remarquons tout de suite que

$$\{B = l\} \iff \{T'_l = i\}.$$

Ou autrement dit, $\mathcal{L}(T'_l|B = l) = \delta_i$. Ainsi la variable aléatoire T'_B est constante égale à i . On a alors pour tout $l \in \{1, \dots, m\}$, si $1 < k < l$:

$$\begin{aligned} \mathcal{L}(T'_1|B = l) &= \mathcal{U}(\{1, \dots, N\} \setminus \{i\}) \\ &\dots \end{aligned}$$

$$\forall (i_1, \dots, i_{k-1}) \in \mathcal{I}_{k-1}^i \quad \mathcal{L}(T'_k|B = l, T'_1 = i_1, \dots, T'_{k-1} = i_{k-1}) = \mathcal{U}(\{1, \dots, N\} \setminus \{i, i_1, \dots, i_{k-1}\})$$

et en notant \wedge_l le fait d'omettre dans un vecteur la composante l , on a pour tout $k = l + 1, \dots, m$,

$$\begin{aligned} \forall (i_1, \dots, \wedge_l, \dots, i_{k-1}) \in \mathcal{I}_{k-1}^i : \\ \mathcal{L}(T'_k | B = l, T'_1 = i_1, \dots, \wedge_l, \dots, T'_{k-1} = i_{k-1}) = \mathcal{U}(\{1, \dots, N\} \setminus \{i, i_1, \dots, \wedge_l, \dots, i_{k-1}\}). \end{aligned}$$

Par le même raisonnement qu'on a fait pour la distribution uniforme, on a pour $(i_1, \dots, \wedge_l, \dots, i_m) \in \mathcal{I}_{m-1}^i$:

$$\mathbb{P}(T'_1 = i_1, \dots, \wedge_l, \dots, T'_m = i_m | B = l) = \frac{(N - m)!}{(N - 1)!}.$$

Pour $j \in \{1, \dots, n\}$, on note X'_j la variable aléatoire qui vaut 1 s'il y a une boule dans la boîte j après avoir réalisé (*) et (**), et 0 sinon. Ainsi $X'_j = \sum_{k=1}^m 1_{T'_k=j}$. Ces nouvelles variables, comme nous allons le voir, feront office des $\tilde{X}_{i,j}^1$. On obtient alors une nouvelle configuration C' .

* Construction des variables $X_{i,j}^1$.

On considère l'action suivante :

(***) On effectue un tirage à pile ou face indépendamment de ce qui précède, avec une probabilité p que l'on tire pile.

On note cette variable aléatoire S . Ainsi $\mathcal{L}(S) = \mathcal{B}(p)$ et S est indépendante des variables C, B, T'_1, \dots, T'_m .

On obtient alors une nouvelle configuration C'' définie comme suit :

→ si $S = \text{pile}$, alors on ne touche à rien. En particulier, on a $C'' = C$.

→ si $S = \text{face}$, alors on choisit une boîte de manière uniforme parmi les boîtes vides restantes, de manière indépendante de tout ce qui précède, et on y place la boule de la boîte i . Ainsi $C'' \neq C'$.

On désigne maintenant par T''_k où $k = 1, \dots, m$, la variable aléatoire qui prend le numéro de la boîte où la boule j est envoyée après (*), (**), et (***) . Ainsi sur $\{B = l\} \cap \{S = \text{face}\}$ on a :

$$\begin{aligned} \forall k \neq l \quad T''_k &= T'_k, \\ T''_l &\in \Gamma \setminus C', \\ T'' \text{ est indépendante de } T'_1, \dots, \wedge_l, \dots, T'_m, \\ C'' &= C' \setminus \{i\} \cup T''_l. \end{aligned}$$

On note de plus pour $j \in \{1, \dots, n\}$, la variable aléatoire X''_j qui vaut 1 si la boîte j est pleine après (*), (**), et (***) (i.e si $j \in C''$), et 0 sinon. Comme $C' \setminus \{i\} \subset C''$, on a pour $j \in \{1, \dots, n\} \setminus \{i\}$:

$$X'_j \leq X''_j.$$

Il est clair aussi que l'ensemble des configurations possibles, après cette distribution, est égal à celui des configurations possibles après distribution uniforme. Il faut montrer que les configurations après telle ou telle distribution ont la même probabilité d'apparaître. D'après ce qui précède, sur $\{B = l\}$ on a $(T'_1, \dots, \wedge_l, \dots, T'_m) = (T''_1, \dots, \wedge_l, \dots, T''_m)$ et en particulier :

$$\mathcal{L}(T''_1, \dots, \wedge_l, \dots, T''_m | B = l) = \mathcal{L}(T'_1, \dots, \wedge_l, \dots, T'_m | B = l).$$

De plus

$$\begin{cases} \mathcal{L}(T''_l | S = \text{pile}, B = l, T''_1 = i_1, \dots, \wedge_l, \dots, T''_m = i_m) &= \delta_i \\ \mathcal{L}(T''_l | S = \text{face}, B = l, T''_1 = i_1, \dots, \wedge_l, \dots, T''_m = i_m) &= \mathcal{U}(\{1, \dots, N\} \setminus \{i, i_1, \dots, i_m\}) \end{cases}$$

Montrons que pour tout k , T''_k suit une loi uniforme sur Γ . Pour tout $k = 1, \dots, m$ on a :

$$\{T''_l = i\} = \{T'_l = i\} \cap \{S = \text{pile}\} = \{B = l\} \cap \{S = \text{pile}\}$$

et par indépendance de S par rapport à B , il en découle que :

$$\mathbb{P}(T''_l = i) = \mathbb{P}(B = l)\mathbb{P}(S = \text{pile}) = \frac{1}{N}.$$

Et pour $j \neq i$,

$$\begin{aligned}
\mathbb{P}(T_l'' = j) &= \sum_{k=1}^m \sum_{(i_1, \dots, i_m) \in \mathcal{I}_{m-1}^i} \mathbb{P}(T_l'' = i | S = \text{face}, B = k, T_1'' = i_1, \dots, \wedge_l, \dots, T_m'' = i_m) \\
&\quad \times \mathbb{P}(S = \text{face}) \mathbb{P}(B = k) \mathbb{P}(T_1'' = i_1, \dots, \wedge_l, \dots, T_m'' = i_m | B = k) \\
&= \sum_{k=1}^m \frac{1}{N-m} \frac{N-m}{N} \frac{1}{m} \sum_{(i_1, \dots, i_m) \in \mathcal{I}_{m-1}^i} \mathbb{P}(T_1' = i_1, \dots, \wedge_l, \dots, T_m' = i_m | B = k) \\
&= \sum_{k=1}^m \frac{1}{Nm} \sum_{(i_1, \dots, i_m) \in \mathcal{I}_{m-1}^i} \frac{1}{N-2} \cdots \frac{1}{N-2-(m-2)} \\
&= \frac{1}{N} (m-1)! \binom{N-2}{m-1} \frac{(N-m-1)!}{(N-2)!} = \frac{1}{N}.
\end{aligned}$$

Prenons maintenant $(i_1, \dots, i_m) \in \mathcal{I}_m$. S'il existe $l \in \{1, \dots, m\}$ tel que $i_l = i$, alors grâce à l'indépendance de S par rapport à T_1', \dots, T_m' et B on a :

$$\begin{aligned}
\mathbb{P}(T_1'' = i_1, \dots, T_m'' = i_m) &= \mathbb{P}(T_l'' = i) \mathbb{P}(T_1'' = i_1, \dots, \wedge_l, \dots, T_m'' = i_m | T_l'' = i) \\
&= \frac{1}{N} \mathbb{P}(T_1'' = i_1, \dots, \wedge_l, \dots, T_m'' = i_m | B = l, S = \text{pile}) \\
&= \frac{1}{N} \mathbb{P}(T_1' = i_1, \dots, \wedge_l, \dots, T_m' = i_m | B = l, S = \text{pile}) \\
&= \frac{1}{N} \mathbb{P}(T_1' = i_1, \dots, \wedge_l, \dots, T_m' = i_m | B = l) \\
&= \frac{1}{N} \frac{(N-m)!}{(N-1)!} = \frac{(N-m)!}{N!} \\
&= \mathbb{P}(T_1 = i_1, \dots, T_m = i_m).
\end{aligned}$$

Maintenant si $\forall l \in \{1, \dots, m\}$, on a $T_l \neq i$, cela signifie qu'on a obtenu $S = \text{face}$ et alors :

$$\begin{aligned}
\mathbb{P}(T_1'' = i_1, \dots, T_m'' = i_m) &= \sum_{l=1}^m \mathbb{P}(T_1'' = i_1, \dots, T_m'' = i_m, S = \text{face}, B = l) \\
&= \sum_{l=1}^m \mathbb{P}(T_1' = i_1, \dots, \wedge_l, \dots, T_m' = i_m, T_l'' = i_l, S = \text{face}, B = l) \\
&= \sum_{l=1}^m \mathbb{P}(T_l'' = i_l, S = \text{face}, B = l) \mathbb{P}(T_1' = i_1, \dots, \wedge_l, \dots, T_m' = i_m | B = l) \\
&= \sum_{l=1}^m \mathbb{P}(T_l'' = i_l) \mathbb{P}(B = l) \mathbb{P}(T_1' = i_1, \dots, \wedge_l, \dots, T_m' = i_m | B = l) \\
&= \sum_{l=1}^m \frac{1}{N} \frac{1}{m} \frac{(N-m)!}{(N-1)!} = \frac{(N-m)!}{N!} \\
&= \mathbb{P}(T_1 = i_1, \dots, T_m = i_m).
\end{aligned}$$

Cela signifie finalement que :

$$\mathcal{L}(X_j'', j) = \mathcal{L}(X_j, j).$$

On a donc (6.14). Poursuivons : on a pour $(k_1, \dots, k_n) \in \{0, 1\}^n$

$$\mathbb{P}(X_1' = k_1, \dots, X_n' = k_n) = \begin{cases} 0 & \text{si } k_i = 0 \\ \mathbb{P}(X_1' = k_1, \dots, X_i' = 1, \dots, X_n' = k_n) & \text{si } k_i = 1 \end{cases}$$

De plus par indépendance de la variable S par rapport aux X'_1, \dots, X'_n on a :

$$\begin{aligned} \mathbb{P}(X'_1 = k_1, \dots, X'_i = 1, \dots, X'_n = k_n) &= \mathbb{P}(X''_1 = k_1, \dots, X''_i = 1, \dots, X''_n = k_n | S = \text{pile}) \\ &= \frac{1}{\mathbb{P}(S = \text{pile})} \mathbb{P}(X''_1 = k_1, \dots, X''_i = 1, \dots, X''_n = k_n) \\ &= \frac{1}{p} \mathbb{P}(X_1 = k_1, \dots, X_i = 1, \dots, X_n = k_n). \end{aligned}$$

Or on a :

$$\mathbb{P}(X_1 = k_1, \dots, X_n = k_n | X_i = 1) = \begin{cases} 0 & \text{si } k_i = 0 \\ \frac{1}{\mathbb{P}(X_i=1)} \mathbb{P}(X_1 = k_1, \dots, X_i = 1, \dots, X_n = k_n) & \text{si } k_i = 1 \end{cases}$$

Comme $\mathbb{P}(X_i = 1) = p$ par définition, on en déduit l'égalité :

$$\mathbb{P}(X'_1 = k_1, \dots, X'_n = k_n) = \mathbb{P}(X_1 = k_1, \dots, X_n = k_n | X_i = 1),$$

pour tout $(k_1, \dots, k_n) \in \{0, 1\}^n$. Nos variables vérifient donc (6.13).

Tout cela signifie que les X_j sont négativement liées. On peut donc appliquer le Théorème 6.3 qui donne :

$$\begin{aligned} \|\mathcal{L}(W) - \mathcal{P}(\lambda)\|_{VT} &\leq k_2(\lambda) (\lambda - \text{Var}(W)) \\ &= (1 - \exp(-\lambda)) \left(1 - (1-p) \frac{N-n}{N-1}\right) \\ &\leq (1 - \exp(-\lambda)) \left(\left(\frac{n}{N} + p\right) \frac{N}{N-1}\right) \\ &= (1 - \exp(-\lambda)) \frac{n+m}{N-1}. \end{aligned}$$

■

Remarque : La majoration est efficace lorsque n et m sont petits par rapport à N .

Troisième partie

Entropie conditionnelle

Dans cette partie, nous établissons des résultats sur la notion d'*entropie*. L'entropie conditionnelle mesure d'une certaine manière, l'écart entre deux probabilités, tout comme la distance en variation totale (introduite en première partie). L'entropie intervient en particulier dans la théorie de l'information. Nous introduisons dans un premier temps un autre écart d_2 entre deux probabilités, avec lequel on fera le lien avec l'entropie. Nous appliquerons ces résultats, dans un deuxième temps, pour aboutir à une inégalité de Poincaré. L'inégalité de Poincaré obtenue fait partie des outils pour obtenir des résultats isopérimétriques.

7 Liaison entropie et écart d_2

On reprendra les notations de [2] tout au long de cette partie.

Soient ν et μ deux probabilités sur \mathbb{R}^n . Rappelons que $\mathcal{M}(\nu, \mu)$ est l'ensemble des couplages entre ν et μ . Considérons :

– Une pseudo-distance sur l'espace des probabilités sur un même espace (ici \mathbb{R}^n) :

$$d_2(\nu, \mu) = \inf_{\gamma \in \mathcal{M}(\nu, \mu)} \sup_{\alpha} \iint \sum_{i=1}^n \alpha_i(y) 1_{x_i \neq y_i} d\gamma(x, y), \quad (7.17)$$

où le sup est pris sur les $\alpha = (\alpha_1, \dots, \alpha_n)$ tels que $\alpha_i \geq 0$ et $\int \sum_{i=1}^n \alpha_i^2(y) d\mu(y) \leq 1$.

– Pour $f : \mathbb{R}^n \rightarrow \mathbb{R}_+$ on définit l'*entropie* de f selon μ est donnée par :

$$Ent_{\mu}(f) = \int_{\mathbb{R}^n} f \log(f) d\mu - \int_{\mathbb{R}^n} f d\mu \cdot \log \left(\int_{\mathbb{R}^n} f d\mu \right). \quad (7.18)$$

Remarque : Enonçons tout de suite deux cas particuliers :

– Si f est constante positive, disons $a \geq 0$, alors :

$$Ent_{\mu}(a) = a \log(a) - a(\log(a) + \log(1)) = 0,$$

et ceci même si $a = 0$ grâce au prolongement par continuité en 0 de la fonction $x \mapsto x \log(x)$.

– Si f est une densité de probabilité par rapport à μ , alors l'entropie relative à μ se réduit à :

$$Ent_{\mu}(f) = \int_{\mathbb{R}^d} f \log(f) d\mu.$$

Les méthodes de couplage vont nous permettre de lier les quantités (7.17) et (7.18). J'ai repris le même énoncé du théorème que P-M. Samson dans [2]. Cependant celui-ci ne propose une démonstration que dans le cas où μ est à densité *strictement positive*. Nous allons donner dans un premier temps, une démonstration plus générale en dimension 1. Nous donnerons dans un second temps le théorème en dimension n quelconque, et une esquisse de la démonstration.

7.1 En dimension un

THÉORÈME 7.1 *Si μ est à densité par rapport à une mesure de référence ρ sur \mathbb{R} et si ν est absolument continue par rapport à μ , de dérivée de Radon-Nikodym $\frac{d\nu}{d\mu}$ alors on a les inégalités suivantes :*

$$d_2(\nu, \mu) \leq \sqrt{2Ent_{\mu} \left(\frac{d\nu}{d\mu} \right)}, \quad (7.19)$$

$$d_2(\mu, \nu) \leq \sqrt{2Ent_{\mu} \left(\frac{d\nu}{d\mu} \right)}. \quad (7.20)$$

Remarque : Nous supposons toujours $\nu \neq \mu$. Dans le cas contraire on a $d_2(\nu, \nu) = 0$ et le Théorème 7.1 n'a pas grand intérêt.

Pour démontrer ce Théorème 7.1, nous avons besoin du lemme calculatoire suivant, qui est cependant vrai en dimension n quelconque :

LEMME 7.1 *Soient μ et ν deux probabilités de densité respective f strictement positive et g par rapport à une mesure ρ sur \mathbb{R}^n . Alors nous avons les deux inégalités suivantes :*

$$d_\rho(f|g) := \left(\int \left[1 - \frac{f}{g}\right]_+^2 g d\rho \right)^{\frac{1}{2}} \leq \sqrt{2Ent_\mu\left(\frac{g}{f}\right)} \quad (7.21)$$

$$d_\rho(g|f) := \left(\int \left[1 - \frac{g}{f}\right]_+^2 f d\rho \right)^{\frac{1}{2}} \leq \sqrt{2Ent_\mu\left(\frac{g}{f}\right)} \quad (7.22)$$

Démonstration :

Posons $u = \frac{g}{f}$. On a :

$$Ent_P(u) = \int (u \log(u) - u + 1)g d\nu.$$

Soient alors pour $u > 0$:

$$\begin{cases} \psi(u) &= u \log(u) - u + 1 \\ \Phi(u) &= \frac{\psi(u)}{u} \end{cases}$$

Sur $]0, 1]$ l'application $u \mapsto \psi(u) - \frac{1}{2}(1-u)^2$ est dérivable de dérivée :

$$\log(u) + 1 - u \leq 0$$

Donc elle est décroissante sur $]0, 1]$ et vaut 0 en 1, donc est positive sur tout $]0, 1]$ i.e :

$$\psi(u) \geq \frac{1}{2}(1-u)^2. \quad (7.23)$$

Remarquons que (7.23) est aussi vraie pour $u = 0$.

Supposons maintenant que g ne s'annule pas.

On peut alors considérer $\frac{f}{g} = \frac{1}{u}$. Sur $[1, +\infty[$ l'application $u \mapsto \Phi(u) - \frac{1}{2}(1 - \frac{1}{u})^2$ est dérivable de dérivée :

$$\frac{1}{u} - \frac{1}{u^3} \geq 0.$$

Donc elle est croissante sur $[1, +\infty[$ et vaut 0 en 1, donc est positive sur tout $[1, +\infty[$ i.e :

$$\Phi(u) \geq \frac{1}{2}\left(1 - \frac{1}{u}\right)^2.$$

Finalement on a pour tout $u \geq 0$:

$$\psi(u) = \psi(u)1_{u \leq 1} + u\Phi(u)1_{u > 1} \geq \frac{1}{2}[1-u]_+^2 + \frac{1}{2}[u-1]_+^2.$$

Et cette inégalité est aussi vraie quand g (i.e u) s'annule. Ainsi :

$$Ent_\mu(u) \geq \frac{1}{2} \left(\int [1-u]_+^2 f d\rho + \int \left[1 - \frac{1}{u}\right]_+^2 g d\rho \right),$$

d'où :

$$2Ent_\mu(u) \geq d_\rho^2(f|g) + d_\rho^2(g|f).$$

■

→ Nous pouvons maintenant passer au Théorème 7.1. Dans [2], l'auteur écrit une chose fausse : il affirme que γ (voir suite) est une mesure à densité alors que ça n'en n'est pas une. Voici donc une preuve rédigée correctement de ce théorème.

Démonstration :

On considère f la densité de μ par rapport à ρ . Si f n'est pas strictement positive alors on choisit $\rho = \mu$ comme mesure de référence pour palier à cette éventualité. Et comme ν est absolument continue par rapport à μ , ν est à densité, disons g , par rapport à ρ sur \mathbb{R} . Ainsi :

$$g = \frac{d\nu}{d\mu} f.$$

Notons $i_\Delta := i_{\Delta_{\mathbb{R}}} : x \mapsto (x, x)$.

Considérons la mesure $\gamma : \{A \subset \mathbb{R} \times \mathbb{R}; \rho\text{-mesurable}\} \rightarrow \mathbb{R}_+$ définie par :

$$(x, y) \mapsto \left(i_{\Delta_*} \left(\min(f, g) \cdot \rho \right) \right) (x, y) + \frac{[g(y) - f(y)]_+ \cdot [f(x) - g(x)]_+}{\|\nu - \mu\|_{VT}} \cdot d\rho \otimes \rho(x, y)$$

D'après (4.2), on a bien que $\gamma \in \mathcal{M}(\mu, \nu)$.

Soit maintenant $\alpha : \mathbb{R} \rightarrow \mathbb{R}_+$ une application μ -mesurable telle que $\int_{\mathbb{R}} \alpha(y)^2 d\nu(y) \leq 1$. On a alors en utilisant le théorème de Fubini :

$$\begin{aligned} \iint_{\mathbb{R}^2} \alpha(y) 1_{y \neq x} d\gamma(x, y) &= \iint_{\mathbb{R}^2} \alpha(y) \mathbb{1}_{y \neq x} \frac{[g(y) - f(y)]_+ \cdot [f(x) - g(x)]_+}{\|\nu - \mu\|_{VT}} d\rho \otimes \rho(x, y) \\ &= \int_{\mathbb{R}} \alpha(y) [g(y) - f(y)]_+ \mathbb{1}_B d\rho(y) \end{aligned}$$

où $B = \{y \in \mathbb{R} | g(y) > 0\}$. Et en appliquant l'inégalité de Cauchy-Schwarz à $\alpha\sqrt{g}$ et $\left[1 - \frac{f}{g}\right]_+ \sqrt{g} 1_B$, on obtient :

$$\begin{aligned} \iint_{\mathbb{R}^2} \alpha(y) 1_{y \neq x} d\gamma(x, y) &\leq \left(\int_{\mathbb{R}} \alpha(y)^2 g(y) d\rho(y) \right)^{\frac{1}{2}} \cdot \left(\int_B \left[1 - \frac{f(y)}{g(y)}\right]_+ g(y) d\rho(y) \right)^{\frac{1}{2}} \\ &\leq \left(\int_{\mathbb{R}} \left[1 - \frac{f(y)}{g(y)}\right]_+ g(y) d\rho(y) \right)^{\frac{1}{2}} \\ &= d_\rho(f|g). \end{aligned}$$

En prenant $\beta : \mathbb{R} \rightarrow \mathbb{R}_+$ ρ -mesurable telle que $\int_{\mathbb{R}} \beta(x)^2 d\mu(x) \leq 1$, on obtient de la même façon :

$$\iint \beta(x) 1_{y \neq x} d\gamma(x, y) \leq d_\rho(g|f).$$

Appliquons maintenant le Lemme 7.1. En effet, les inégalités (7.21) et (7.22) nous fournissent immédiatement :

$$\begin{aligned} \iint \alpha(y) 1_{y \neq x} d\gamma(x, y) &\leq \sqrt{2Ent_\mu \left(\frac{g}{f} \right)}, \\ \iint \beta(x) 1_{y \neq x} d\gamma(x, y) &\leq \sqrt{2Ent_\mu \left(\frac{g}{f} \right)}. \end{aligned}$$

Ceci étant vrai pour tout α et β , on peut prendre le sup sur ces fonctions ρ -mesurables qui vérifient $\|\alpha\|_{L^2(\nu)} \leq 1$ et $\|\beta\|_{L^2(\mu)} \leq 1$ et ainsi :

$$\begin{aligned} \sup_{\alpha} \iint \alpha(y) 1_{y \neq x} d\gamma(x, y) &\leq \sqrt{2Ent_{\mu} \left(\frac{g}{f} \right)}, \\ \sup_{\beta} \iint \beta(x) 1_{y \neq x} d\gamma(x, y) &\leq \sqrt{2Ent_{\mu} \left(\frac{g}{f} \right)}. \end{aligned}$$

Et comme on a trouvé une mesure $\gamma \in \mathcal{M}(\nu, \mu)$ pour laquelle ces inégalités sont vraies, on a finalement le résultat :

$$\begin{aligned} \inf_{\gamma \in \mathcal{M}(\nu, \mu)} \sup_{\alpha} \iint \alpha(y) 1_{y \neq x} d\gamma(x, y) &\leq \sqrt{2Ent_{\mu} \left(\frac{g}{f} \right)}, \\ \inf_{\gamma \in \mathcal{M}(\nu, \mu)} \sup_{\beta} \iint \beta(x) 1_{y \neq x} d\gamma(x, y) &\leq \sqrt{2Ent_{\mu} \left(\frac{g}{f} \right)}. \end{aligned}$$

■

7.2 En dimension finie quelconque

→ Nous allons énoncer l'équivalent du théorème 7.1 en dimension finie quelconque ($n \geq 1$). Pour cela, nous avons besoin d'autres notations, reprises là encore dans [2].

Considérons un échantillon X de taille n de variables aléatoires sur \mathbb{R}^n et de loi μ . Soit $\Gamma = (\gamma_i^j)_{1 \leq i, j \leq n}$ la matrice de dépendance des X_1, \dots, X_n définie de la manière suivante. Pour $i \geq j$:

$$\gamma_i^j = \begin{cases} 0 & \text{si } i > j \\ 1 & \text{si } i = j \end{cases}$$

Et pour $i < j$:

$$\gamma_i^j = \left(\sup_{(x_i, y_i) \in \mathbb{R}^2} \sup_{y_1^{i-1} \in \mathbb{R}^{i-1}} a_j(y_1^{i-1}, x_i, y_i) \right)^{\frac{1}{2}},$$

où $a_j(y_1^{i-1}, x_i, y_i) = \left\| \mathcal{L}(X_j^n | X_1^{i-1} = y_1^{i-1}, X_i = x_i) - \mathcal{L}(X_j^n | X_1^{i-1} = y_1^{i-1}, X_i = y_i) \right\|_{VT}$

et $X_i^j := (X_i, \dots, X_j)$. On note $\|\Gamma\| = \left(\sum_{i, j=1}^n (\gamma_i^j)^2 \right)^{\frac{1}{2}}$ la norme de la topologie euclidienne. Remarquons que si $n = 1$, l'échantillon est de taille 1 et la matrice Γ est réduite à (1) donc $\|\Gamma\| = 1$. La qualité des majorations obtenues dépend donc fortement de la dimension de l'espace considéré.

Nous sommes en mesure d'énoncer le théorème dans le cadre général.

THÉORÈME 7.2 *Si μ est à densité strictement positive par rapport à $\rho_1 \otimes \dots \otimes \rho_n$ sur \mathbb{R}^n et ν une probabilité sur \mathbb{R}^n admettant une dérivée de Radon-Nikodym $\frac{d\nu}{d\mu}$ par rapport à μ , alors on a les inégalités :*

$$d_2(\nu, \mu) \leq \|\Gamma\| \sqrt{2Ent_{\mu} \left(\frac{d\nu}{d\mu} \right)}, \quad (7.24)$$

$$d_2(\mu, \nu) \leq \|\Gamma\| \sqrt{2Ent_{\mu} \left(\frac{d\nu}{d\mu} \right)}. \quad (7.25)$$

Démonstration :

Pour $n = 1$: est traité par le théorème 7.1.

Pour $n \geq 1$: on raisonne par récurrence et on suppose le résultat vrai pour $n - 1$. Soient f est la densité de μ et g celle de ν par rapport $\rho_1 \otimes \cdots \otimes \rho_n$. La relation

$$g = \frac{d\nu}{d\mu} f,$$

impose également g à être strictement positive. Considérons $\alpha = (\alpha_1, \dots, \alpha_n)$ et $\beta = (\beta_1, \dots, \beta_n)$ des vecteurs de fonctions positives telles que :

$$\int \sum_{i=1}^n \alpha_i(y)^2 d\nu(y) \leq 1$$

$$\int \sum_{i=1}^n \beta_i(x)^2 d\mu(x) \leq 1$$

Comme $f > 0$ on peut écrire :

$$f(x_1, \dots, x_n) = f_n(x_n | x_1, \dots, x_{n-1}) \cdots f_2(x_2 | x_1) f_1(x_1),$$

avec :

$$f_j(x_j | x_1, \dots, x_{j-1}) = \frac{\int f(x_1, \dots, x_j, z_{j+1}, \dots, z_n) d\rho_{j+1}(z_{j+1}) \cdots d\rho_n(z_n)}{\int f(x_1, \dots, x_{j-1}, z_j, \dots, z_n) d\rho_j(z_j) \cdots d\rho_n(z_n)}.$$

On a $\int f_j(x_j | x_1, \dots, x_{j-1}) d\rho_j(x_j) = 1$. Soient alors $F_j(\cdot | x_1, \dots, x_{j-1})$ les probabilités qui ont pour densité respective $f_j(\cdot | x_1, \dots, x_{j-1})$ par rapport à ρ_j pour $j = 1, \dots, n$. Introduisons maintenant pour $1 \leq i < j \leq k \leq n$:

$$f_j^k(x_j, \dots, x_k | x_1, \dots, x_i) = \int f(x_1, \dots, x_i, z_{i+1}, \dots, z_{j-1}, x_j, \dots, x_k, z_{k+1}, \dots, z_n) d\rho_{i+1}(z_{i+1}) \cdots d\rho_{j-1}(z_{j-1}) d\rho_{k+1}(z_{k+1}) \cdots d\rho_n(z_n).$$

Soit alors $F_j^k(\cdot, \dots, \cdot | x_1, \dots, x_i)$ la probabilité ayant $f_j^k(\cdot, \dots, \cdot | x_1, \dots, x_i)$ pour densité par rapport à $d\rho_j \otimes \cdots \otimes d\rho_k$. On définit de façon similaire pour g les g_n, \dots, g_1 et G_j pour $j = 1, \dots, n$. Et pour $1 \leq i < j \leq k \leq n$ les $g_j^k(\cdot, \dots, \cdot | y_1, \dots, y_i)$ et $G_j^k(\cdot, \dots, \cdot | y_1, \dots, y_i)$. Soit enfin pour $i = 1, \dots, n$:

$$E_i = \int Ent_{F_i(\cdot | y_1, \dots, y_{i-1})} \left(\frac{g_i(\cdot | y_1, \dots, y_{i-1})}{f_i(\cdot | y_1, \dots, y_{i-1})} \right) dG_1^{i-1}(y_1, \dots, y_{i-1}).$$

LEMME 7.2 On a :

$$\sum_{i=1}^n E_i = Ent_{\mu} \left(\frac{d\nu}{d\mu} \right).$$

Démonstration :

Rappelons que :

$$Ent_{\mu} \left(\frac{d\nu}{d\mu} \right) = Ent_{\mu} \left(\frac{g}{f} \right) = \int \frac{g}{f} \log \left(\frac{g}{f} \right) d\mu.$$

On réécrit f et g , pour $(y_1, \dots, y_n) \in \mathbb{R}^n$:

$$\frac{g(y_1, \dots, y_n)}{f(y_1, \dots, y_n)} = \prod_{i=1}^n \frac{g_i(y_i | y_1, \dots, y_{i-1})}{f_i(y_i | y_1, \dots, y_{i-1})}.$$

Donc :

$$\begin{aligned} Ent_\mu \left(\frac{g}{f} \right) &= \sum_{i=1}^n \int \cdots \int \log \left(\frac{g_i(y_i|y_1, \dots, y_{i-1})}{f_i(y_i|y_1, \dots, y_{i-1})} \right) \frac{\prod_{i=1}^n g_i(y_i|y_1, \dots, y_{i-1})}{f(y_1, \dots, y_n)} \\ &\quad \times f(y_1, \dots, y_n) d\rho_1 \otimes \cdots \otimes \rho_n \\ &= \sum_{i=1}^n \int \log \left(\frac{g_i(y_i|y_1, \dots, y_{i-1})}{f_i(y_i|y_1, \dots, y_{i-1})} \right) dG_1^n(y_1, \dots, y_n). \end{aligned}$$

Mais d'après les notations ci-dessus :

$$\begin{aligned} dG_1^n(y_1, \dots, y_n) &= G_n(dy_n|y_1, \dots, y_{n-1}) \times \cdots \times G_1(dy_1) \\ &= g_n(y_n|y_1, \dots, y_{n-1}) \times \cdots \times g_1(y_1) \cdot d(\rho_n \otimes \cdots \otimes \rho_1)(y_n, \dots, y_1). \end{aligned}$$

Donc :

$$\begin{aligned} &\int \log \left(\frac{g_i(y_i|y_1, \dots, y_{i-1})}{f_i(y_i|y_1, \dots, y_{i-1})} \right) dG_1^n(y_1, \dots, y_n) \\ &= \int \log \left(\frac{g_i(y_i|y_1, \dots, y_{i-1})}{f_i(y_i|y_1, \dots, y_{i-1})} \right) g_i(y_i|y_1, \dots, y_{i-1}) \cdots g_1(y_1) d(\rho_i \otimes \cdots \otimes \rho_1)(y_i, \dots, y_1) \\ &\quad \times \int g_n(y_n|y_1, \dots, y_{n-1}) \cdots g_{i+1}(y_{i+1}|y-1, \dots, y_i) d(\rho_n \otimes \cdots \otimes \rho_{i+1})(y_n, \dots, y_{i+1}) \\ &= \int \log \left(\frac{g_i(y_i|y_1, \dots, y_{i-1})}{f_i(y_i|y_1, \dots, y_{i-1})} \right) G_i(dy_1|y_1, \dots, y_{i-1}) G_1^{i-1}(dy_1, \dots, dy_{i-1}). \end{aligned}$$

Et on écrit :

$$Ent_{F_i(\cdot|y_1, \dots, y_{i-1})} \left(\frac{g_i(y_i|y_1, \dots, y_{i-1})}{f_i(y_i|y_1, \dots, y_{i-1})} \right) = \int \log \left(\frac{g_i(y_i|y_1, \dots, y_{i-1})}{f_i(y_i|y_1, \dots, y_{i-1})} \right) G_i(dy_1|y_1, \dots, y_{i-1}).$$

Finalement on obtient le résultat :

$$\begin{aligned} Ent_\mu \left(\frac{g}{f} \right) &= \sum_{i=1}^n \int Ent_{F_i(\cdot|y_1, \dots, y_{i-1})} \left(\frac{g_i(y_i|y_1, \dots, y_{i-1})}{f_i(y_i|y_1, \dots, y_{i-1})} \right) G_1^{i-1}(dy_1, \dots, dy_{i-1}) \\ &= \sum_{i=1}^n E_i. \end{aligned}$$

■

Nous allons voir que ce Lemme est une clef importante de la démonstration. Pour alléger les notations, notons maintenant pour $1 \leq j \leq n$:

$$\begin{cases} \Delta_j &= \int \alpha_j(y)^2 d\nu(y) \\ \tilde{\Delta}_j &= \int \beta_j(x)^2 d\mu(x) \end{cases}$$

On a par hypothèse sur α et β : $\sum_{j=1}^n \Delta_j \leq 1$ et $\sum_{j=1}^n \tilde{\Delta}_j \leq 1$.

→ Dans [2], pour obtenir le résultat du Théorème 7.2, l'auteur construit une probabilité $\gamma \in \mathcal{M}(\nu, \mu)$ qui vérifie les inégalités suivantes :

$$\iint \alpha_j(y) \mathbb{1}_{y_j \neq x_j} d\gamma(x, y) \leq \sum_{i=1}^j \gamma_i^j (2E_i)^{\frac{1}{2}} (\Delta_j)^{\frac{1}{2}}, \quad (7.26)$$

$$\iint \beta_j(x) \mathbb{1}_{y_j \neq x_j} d\gamma(x, y) \leq \sum_{i=1}^j \gamma_i^j (2E_i)^{\frac{1}{2}} (\tilde{\Delta}_j)^{\frac{1}{2}}. \quad (7.27)$$

Avec ces inégalités (7.26) et (7.27), le résultat du Théorème 7.2 s'en suit facilement. En effet, partant de (7.26), montrons (7.24) ((7.27) impliquera (7.25) de la même façon). En sommant sur j , en appliquant plusieurs fois l'inégalité de Cauchy-Schwarz et en se rappelant que $\gamma_i^j = 0$ pour $i > j$, on a :

$$\begin{aligned}
 \iint \sum_{j=1}^n \alpha_j(y) \mathbb{1}_{x_j \neq y_j} d\gamma(x, y) &\leq \sum_{j=1}^n \left((\Delta_j)^{\frac{1}{2}} \sum_{i=1}^j (2E_i)^{\frac{1}{2}} \gamma_i^j \right) \\
 &\leq \left(\sum_{j=1}^n \Delta_j \right)^{\frac{1}{2}} \left(\sum_{j=1}^n \left(\sum_{i=1}^j (2E_i)^{\frac{1}{2}} \gamma_i^j \right)^2 \right)^{\frac{1}{2}} \\
 &= \left(\sum_{j=1}^n \Delta_j \right)^{\frac{1}{2}} \left(\sum_{j=1}^n \left(\sum_{i=1}^n (2E_i)^{\frac{1}{2}} \gamma_i^j \right)^2 \right)^{\frac{1}{2}} \\
 &\leq \left(\sum_{j=1}^n \Delta_j \right)^{\frac{1}{2}} \left(\sum_{j=1}^n \left(\sum_{i=1}^n (2E_i) \right) \left(\sum_{i=1}^n (\gamma_i^j)^2 \right) \right)^{\frac{1}{2}} \\
 &= \left(\sum_{j=1}^n \Delta_j \right)^{\frac{1}{2}} \left(\left(\sum_{i=1}^n (2E_i) \right) \left(\sum_{i,j=1}^n (\gamma_i^j)^2 \right) \right)^{\frac{1}{2}} \\
 &= \|\Gamma\| \left(2 \sum_{i=1}^n E_i \right)^{\frac{1}{2}} \left(\sum_{j=1}^n \Delta_j \right)^{\frac{1}{2}}.
 \end{aligned}$$

Grâce au Lemme et par hypothèse sur les Δ_j , on obtient :

$$\iint \sum_{j=1}^n \alpha_j(y) \mathbb{1}_{x_j \neq y_j} d\gamma(x, y) \leq \|\Gamma\| \sqrt{2Ent_\mu \left(\frac{g}{f} \right)}.$$

Ceci étant vrai pour tout α vérifiant les conditions de départ, et le membre de droite de l'inégalité ne dépendant pas de α on peut prendre le *sup* sur celui de gauche, puis enfin l'*inf* sur les couplages $\gamma \in \mathcal{M}(Q, P)$ pour obtenir ce que l'on veut, à savoir :

$$d_2(\mu, \nu) \leq \|\Gamma\| \sqrt{2Ent_\mu \left(\frac{g}{f} \right)}.$$

■

8 Application : une inégalité de Poincaré

Gardons les mêmes notations que la partie précédente. On se donne pour toute la suite : une mesure de probabilité μ sur $[0, 1]^n$ et $f : [0, 1]^n \rightarrow \mathbb{R}$ une fonction μ -mesurable, continue sur $[0, 1]$ et \mathcal{C}^1 sur $]0, 1[^n$. On notera pour $j \in \{1, \dots, n\}$, l'opérateur $\partial_j = \frac{\partial}{\partial_j}$.

LEMME 8.1 *Si f est convexe alors pour tout $x, y \in [0, 1]^n$:*

$$f(y_1, \dots, y_n) - f(x_1, \dots, x_n) \leq \sum_{j=1}^n |\partial_j(y)| \mathbb{1}_{y_j \neq x_j}. \quad (8.28)$$

Si f est concave alors pour tout $x, y \in [0, 1]^n$:

$$f(y_1, \dots, y_n) - f(x_1, \dots, x_n) \leq \sum_{j=1}^n |\partial_j(x)| \mathbb{1}_{y_j \neq x_j}. \quad (8.29)$$

Démonstration :

Supposons f convexe.

* Prouvons déjà le lemme pour $n = 1$. Soient $x, y \in [0, 1]^n$. f' est continue sur $]0, 1[$ et alors on écrit $f(y) - f(x) = \int_x^y f'(t) dt$. De plus la convexité de f entraîne la croissance de f'

→ Si $y = x$, l'inégalité est triviale.

→ Si $y > x$ alors :

$$f(y) - f(x) \leq f'(y) \int_x^y dt = f'(y)(y - x).$$

→ Si $y < x$ alors :

$$f(y) - f(x) \geq f'(y)(y - x).$$

Finalement on a donc $f(y) - f(x) \leq |f'(y)| \mathbb{1}_{y \neq x}$.

* Supposons maintenant $n > 1$. Pour $x, y \in [0, 1]^n$, il est facile de voir que l'application

$$g : t \in [0, 1] \longmapsto f((1 - t)x + ty)$$

est convexe. Elle est de plus continue sur $[0, 1]$ et \mathcal{C}^1 sur $]0, 1[$. On peut donc lui appliquer l'inégalité précédente pour en particulier 0 et 1 :

$$g(1) - g(0) = f(y) - f(x) \leq |g'(1)|.$$

On en déduit que :

$$f(y) - f(x) \leq \sum_{j=1}^n |\partial_j f(y)| \mathbb{1}_{y_j \neq x_j}.$$

Cela étant vrai pour tout $x, y \in [0, 1]^n$, on a l'inégalité (8.28).

Dans le cas où f est concave, on applique l'inégalité précédente à $-f$ qui alors convexe, et on retrouve bien (8.29) ■

Passons maintenant à la proposition qui nous permet d'obtenir facilement l'inégalité de Poincaré recherchée. On dénote par ∇f le gradient de f et $|\nabla f|$ sa norme euclidienne dans \mathbb{R}^n .

PROPOSITION 8.1 *Si f est convexe alors :*

$$Ent_\mu(e^f) \leq 2 \|\Gamma\|^2 \int |\nabla f|^2 e^f d\mu. \quad (8.30)$$

Si f est concave alors :

$$Ent_\mu(e^f) \leq 2 \|\Gamma\|^2 \int |\nabla f|^2 d\mu \int e^f d\mu. \quad (8.31)$$

Démonstration :

Souvenons nous de la définition de l'entropie conditionnelle en (7.18) :

$$Ent_\mu(e^f) = \int_{[0,1]^n} f(y) e^{f(y)} d\mu(y) - \left(\int_{[0,1]^n} e^{f(x)} d\mu(x) \right) \left(\log \int_{[0,1]^n} e^{f(x)} d\mu(x) \right).$$

Comme $-\log$ est convexe, on peut appliquer l'inégalité de Jensen au deuxième facteur du terme de gauche, et avec $\mathbb{E}_\mu(e^f) = \int e^f d\mu$ il vient que :

$$\frac{Ent_\mu(e^f)}{\mathbb{E}_\mu(e^f)} \leq \int f(y) \frac{e^{f(y)}}{\mathbb{E}_\mu(e^f)} d\mu(y) - \int f(x) d\mu(x).$$

Soit maintenant μ^f la mesure de probabilité sur $[0, 1]^n$ qui admet la densité $e^f/\mathbb{E}_\mu(e^f)$ par rapport à μ . On a alors la relation suivante $\frac{d\mu^f}{d\mu} = \frac{e^f}{\mathbb{E}_\mu(e^f)}$. Soit $\gamma \in \mathcal{M}(\mu, \mu^f)$. Alors l'inégalité précédente nous dit exactement :

$$\frac{Ent_\mu(e^f)}{\mathbb{E}_\mu(e^f)} \leq \iint f(y) - f(x) d\gamma(x, y). \quad (8.32)$$

Si f est convexe, alors en réinjectant (8.28) du Lemme précédent dans (8.32) on obtient :

$$\frac{Ent_\mu(e^f)}{\mathbb{E}_\mu(e^f)} \leq \iint \sum_{j=1}^n |\partial_j f(y)| \mathbb{1}_{y_j \neq x_j} d\gamma(x, y).$$

Si on note $I_{\mu^f}(f) = \int |\nabla f(y)|^2 d\mu^f(y)$ alors on a deux cas :

→ Si $I_{\mu^f}(f) = 0$ alors cela signifie que toutes les dérivées partielles de f sont nulles, i.e f est constante.

Et si f est constante, l'entropie de e^f selon μ est nulle, donc l'inégalité (8.30) est immédiate.

→ Sinon $I_{\mu^f}(f) > 0$ et on peut écrire :

$$\frac{Ent_\mu(e^f)}{\mathbb{E}_\mu(e^f)} \leq \iint \left(\sum_{j=1}^n \frac{|\partial_j f(y)|}{I_{\mu^f}(f)^{1/2}} \mathbb{1}_{y_j \neq x_j} d\gamma(x, y) \right) (I_{\mu^f}(f))^{1/2}.$$

De plus on a que :

$$\int \sum_{j=1}^n \frac{|\partial_j f(y)|^2}{I_{\mu^f}(f)} d\mu^f(y) = \frac{I_{\mu^f}(f)}{I_{\mu^f}(f)} = 1.$$

Cela signifie que le sup est atteint dans la définition de $d_2(\mu, \mu^f)$ donnée en (7.17). De plus comme ici $\gamma \in \mathcal{M}(\mu, \mu^f)$ est un couplage particulier, on en déduit que :

$$\frac{Ent_\mu(e^f)}{\mathbb{E}_\mu(e^f)} \leq d_2(\mu, \mu^f) \left(\int \sum_{j=1}^n |\partial_j f(y)|^2 d\mu^f(y) \right)^{1/2}.$$

Puis en appliquant le Théorème (7.2) à $d_2(\mu, \mu^f)$, on a :

$$\frac{Ent_\mu(e^f)}{\mathbb{E}_\mu(e^f)} \leq \|\Gamma\| \left(2 \frac{Ent_\mu(e^f)}{\mathbb{E}_\mu(e^f)} \right)^{1/2} \left(\int |\nabla f|^2 \frac{e^f}{\mathbb{E}_\mu(e^f)} d\mu \right)^{1/2}.$$

Finalement, on obtient (8.30) en divisant cette inégalité par $\mathbb{E}_\mu(e^f)$.

De même si f est concave on obtient successivement les inégalités suivantes grâce à (8.29), (7.17) avec $d_2(\mu^f, \mu)$, puis au Théorème (7.2) :

$$\begin{aligned} \frac{Ent_\mu(e^f)}{\mathbb{E}_\mu(e^f)} &\leq \sum_{j=1}^n \iint |\partial_j f(x)| \mathbb{1}_{y_j \neq x_j} d\gamma(x, y) \\ &\leq d_2(\mu^f, \mu) \left(\int \sum_{j=1}^n |\partial_j f(x)|^2 d\mu(x) \right)^{1/2} \\ &\leq \|\Gamma\| \left(2 \frac{Ent_\mu(e^f)}{\mathbb{E}_\mu(e^f)} \right)^{1/2} \left(\int |\nabla f|^2 d\mu \right)^{1/2}. \end{aligned}$$

On obtient (8.31) en divisant la dernière inégalité par $\mathbb{E}_\mu(e^f)$ ■

Enonçons maintenant l'inégalité de Poincaré désirée.

PROPOSITION 8.2 *Si f est convexe ou concave, alors :*

$$\int f^2 d\mu - \left(\int f d\mu \right)^2 \leq 4 \|\Gamma\|^2 \int |\nabla f|^2 d\mu.$$

Démonstration :

Remarquons d'abord que l'on peut supposer que $\mathbb{E}_\mu(f) = 0$. Dans ce cas si on démontre l'inégalité de Poincaré, alors en remplaçant f par $f - \mathbb{E}_\mu(f)$:

$$\int (f - \mathbb{E}_\mu(f))^2 d\mu = \int f^2 d\mu - (\mathbb{E}_\mu(f))^2.$$

* Traitons le cas où f est convexe. Fixons $a > \epsilon > 0$. On peut appliquer l'inégalité (8.30) à ϵf , ce qui donne :

$$Ent_\mu(e^{\epsilon f}) \leq 2\epsilon^2 \|\Gamma\|^2 \int |\nabla f|^2 e^{\epsilon f} d\mu,$$

c'est-à-dire :

$$\epsilon \int f(y) e^{\epsilon f(y)} d\mu(y) \leq 2\epsilon^2 \|\Gamma\|^2 \int |\nabla f|^2 e^{\epsilon f} d\mu + \left(\int e^{\epsilon f(x)} d\mu(x) \right) \left(\log \int e^{\epsilon f(x)} d\mu(x) \right). \quad (8.33)$$

D'autre part :

- pour tout $x \in [0, 1]^n$, $\epsilon \mapsto e^{\epsilon f(x)}$ est continue en 0,
- $e^{\epsilon f(x)} \leq e^{af(x)}$ pour tout $\epsilon \in [0, a]$, pour tout $x \in [0, 1]^n$,
- $x \mapsto e^{af(x)}$ est μ -intégrable sur $[0, 1]^n$.

On peut donc appliquer le théorème de convergence dominée qui nous dit que :

$$\lim_{\epsilon \rightarrow 0} \int_{[0,1]^n} e^{\epsilon f(x)} d\mu(x) = \int_{[0,1]^n} d\mu = 1.$$

Par continuité de la fonction log, on en déduit que :

$$\lim_{\epsilon \rightarrow 0} \log \int e^{\epsilon f(x)} d\mu(x) = 0.$$

De la même façon, avec le théorème de convergence dominée, on vérifie facilement que :

$$\lim_{\epsilon \rightarrow 0} \int |\nabla f|^2 e^{\epsilon f} d\mu = \int |\nabla f|^2 d\mu.$$

De plus, comme on a $\mathbb{E}_\mu(f) = 0$, on peut écrire :

$$\int f e^{\epsilon f} d\mu = \epsilon \int f \frac{e^{\epsilon f} - 1}{\epsilon} d\mu.$$

On vérifie que :

- pour tout $x \in [0, 1]^n$, $\epsilon \mapsto \frac{e^{\epsilon f(x)} - 1}{\epsilon}$ est continue en 0 de limite $f(x)$,
- $\left| \frac{e^{\epsilon f(x)} - 1}{\epsilon} \right| \leq |f(x)| e^{|f(x)|}$ pour tout $\epsilon \in [0, a]$ et tout $x \in [0, 1]^n$,
- $x \mapsto |f(x)| e^{|f(x)|}$ est μ -intégrable sur $[0, 1]^n$.

Ce qui nous permet une fois encore d'appliquer le théorème de convergence dominée :

$$\lim_{\epsilon \rightarrow 0} \int f(x) \frac{e^{\epsilon f(x)} - 1}{\epsilon} d\mu(x) = \int f(x)^2 d\mu(x).$$

Enfin on écrit :

$$\begin{aligned} \log \int e^{\epsilon f(x)} d\mu(x) &= \log \left(1 + \int (e^{\epsilon f(x)} - 1) d\mu(x) \right) \\ &= \log \left(1 + \epsilon^2 \int \frac{e^{\epsilon f(x)} - 1 - \epsilon f(x)}{\epsilon^2} d\mu(x) \right), \end{aligned}$$

et on vérifie que :

- pour tout $x \in [0, 1]^n$, $\epsilon \mapsto \frac{e^{\epsilon f(x)} - 1 - \epsilon f(x)}{\epsilon^2}$ est continue en 0 de limite $\frac{f(x)^2}{2}$,
- $\left| \frac{e^{\epsilon f(x)} - 1 - \epsilon f(x)}{\epsilon^2} \right| \leq \frac{|f(x)|^2}{2} e^{f(x)}$ pour tout $\epsilon \in [0, a]$, et tout $x \in [0, 1]^n$,
- $x \mapsto \frac{|f(x)|^2}{2} e^{f(x)}$ est μ -intégrable sur $[0, 1]^n$.

En appliquant une fois de plus le théorème de convergence dominée, on a :

$$\lim_{\epsilon \rightarrow 0} \log \int e^{\epsilon f(x)} d\mu(x) = \frac{\epsilon^2}{2} \int f^2(x) d\mu(x).$$

Finalemnt en divisant tout par ϵ^2 et en passant à la limite quand $\epsilon \rightarrow 0$ dans (8.33), il vient que :

$$\int f(x)^2 d\mu(x) \leq 2 \|\Gamma\| \int |\nabla f(x)|^2 d\mu(x) + \frac{1}{2} \int f(x)^2 d\mu(x).$$

Cela nous donne bien le résultat.

* Dans le cas où f est concave, on obtient le résultat en appliquant (8.31) à ϵf , puis en utilisant de la même façon le théorème de convergence dominée ■

Quatrième partie

Introduction à la théorie du transport

L'objet de cette partie est d'établir et d'étudier certaines équations différentielles sur l'espace des mesures. Pour cela, nous avons besoin de mettre sur l'espace des probabilités, une structure pseudo-différentiable. C'est la raison pour laquelle on introduit la notion de mesure dérivée.

9 Mesure dérivée et application à W_2

Nous souhaitons généraliser la notion de dérivée de fonction à celle de *dérivée de mesure*.

Définition 9.1 Soient (M, d) est un espace métrique et I un intervalle de \mathbb{R} . On dit que $u : I \rightarrow M$ est absolument continue s'il existe une fonction $m \in L^1(I)$ telle que :

$$d(u(s), u(t)) \leq \int_s^t m(\tau) d\tau, \quad \forall s \leq t \in I. \quad (9.34)$$

On admettra le résultat suivant :

THÉORÈME 9.1 Si $u : I \rightarrow E$ est absolument continue, la quantité

$$|u'| (t) := \lim_{\epsilon \rightarrow 0} \frac{d(u(t+\epsilon), u(t))}{|\epsilon|}$$

existe pour presque tout $t \in I$, et est appelée mesure dérivée de u . De plus la fonction $t \mapsto |u'| (t)$ est intégrable sur I et est minimale dans la définition de l'absolue continuité de u au sens suivant : si $m \in L^1(I)$ vérifie (9.34) alors :

$$|u'| (t) \leq m(t) \quad \text{pour presque tout } t \in I.$$

Regardons des résultats de cette notion sur l'espace $(P_2(\mathbb{R}^d), W_2)$. On se donne deux mesures $\mu, \nu \in P_2(\mathbb{R}^d)$ et $\gamma \in \mathcal{M}_0^2(\mu, \nu)$. et on considère la famille de mesures :

$$\mu_t = ((1-t)\pi_1 + t\pi_2)_* \gamma, \quad \forall t \in [0, 1].$$

Si $\varphi \in \mathcal{C}_b(\mathbb{R}^d)$, cela s'écrit :

$$\int_{\mathbb{R}^d} \varphi d\mu_t = \int_{\mathbb{R}^d \times \mathbb{R}^d} \varphi((1-t)x + ty) d\gamma(x, y).$$

On peut voir $(\mu_t)_{t \in [0,1]}$ comme une géodésique qui relie μ à ν dans $P_2(\mathbb{R}^d)$. Le lemme suivant indique que cette famille relie μ à ν à vitesse constante :

LEMME 9.1 Pour tout $0 \leq 1 < t < 1$, on a l'égalité :

$$W_2(\mu_s, \mu_t) = (t-s)W_2(\mu, \nu).$$

Démonstration :

Soit $\gamma_{s,t} = ((1-s)\pi_1 + s\pi_2, (1-t)\pi_1 + t\pi_2)_* \gamma$. Alors :

$$\begin{cases} (\pi_1)_* \gamma_{s,t} &= ((1-s)\pi_1 + s\pi_2)_* \gamma = \mu_s \\ (\pi_2)_* \gamma_{s,t} &= ((1-t)\pi_1 + t\pi_2)_* \gamma = \mu_t. \end{cases}$$

Cela signifie que $\gamma_{s,t} \in \mathcal{C}(\mu_s, \mu_t)$. On a alors :

$$\begin{aligned} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\gamma_{s,t}(x, y) &= \int_{\mathbb{R}^d \times \mathbb{R}^d} |((1-s)x + sy) - ((1-t)x + ty)|^2 d\gamma(x, y) \\ &= |s - t|^2 \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\gamma(x, y) \\ &= |s - t|^2 W_2(\mu, \nu). \end{aligned}$$

Cela implique que $W_2(\mu_s, \mu_t) \leq |s - t|W_2(\mu, \nu)$. S'il existe $s_0 < t_0 < 1$ tels que l'inégalité est stricte alors :

$$\begin{aligned} W_2(\mu, \nu) &\leq W_2(\mu, \mu_{s_0}) + W_2(\mu_{s_0}, \mu_{t_0}) + W_2(\mu_{t_0}, \mu) \\ &< W_2(\mu, \nu)((s_0 - 0) + (t_0 - s_0) + (1 - t_0)) = W_2(\mu, \nu). \end{aligned}$$

Ce qui est absurde. Ainsi l'inégalité est en fait une égalité et le résultat en découle. ■

Ce lemme nous permet d'établir que, pour tout $\epsilon > 0$:

$$\frac{W_2(\mu_{t+\epsilon}, \mu_t)}{|\epsilon|} = W_2(\mu, \nu)$$

et donc que la dérivée de la mesure μ_t est exactement la distance de Wasserstein $W_2(\mu, \nu)$:

$$|\mu'_t| = W_2(\mu, \nu).$$

Remarquons que l'absolue continuité implique la continuité. Ainsi si $t \mapsto \mu_t \in P_2(\mathbb{R}^d)$ est absolument continue sur I alors $\mu_{t+\epsilon} \rightarrow \mu_t$ (au sens de W_2) quand ϵ tend vers 0.

PROPOSITION 9.1 *Pour I intervalle ouvert de \mathbb{R} , on considère une application $t \in I \mapsto \mu_t \in P_2(\mathbb{R}^d)$. Si cette application est absolument continue sur I alors il existe un champ de vecteurs mesurable $V : (t, x) \in I \times \mathbb{R}^d \mapsto V_t(x) \in \mathbb{R}^d$ tel que :*

1. $V_t \in L^2(\mathbb{R}^d, \mu_t)$, et $\|V_t\|_{L^2(\mu_t)} \leq |\mu'_t|$ pour presque tout $t \in I$.
2. μ_t vérifie l'équation

$$\frac{\partial \mu_t}{\partial t} + \nabla(V_t \cdot \mu_t) = 0$$

au sens où, pour toute fonction $\psi : I \times \mathbb{R}^d \mapsto \mathbb{R}$ \mathcal{C}^∞ à support compact :

$$\int_I \int_{\mathbb{R}^d} \frac{\partial}{\partial t} \psi(t, x) + \langle \nabla_x \psi(t, x), V_t(x) \rangle d\mu_t(x) dt = 0$$

3. Pour presque tout $t \in I$, $V \in \overline{\{\nabla \psi; \psi \in \mathcal{C}_c^\infty(\mathbb{R}^d)\}}^{L^2(\mu)}$.

Démonstration :

On peut raisonner dans $I =]0, 1[$ pour se fixer les idées.

* Fixons $\varphi \in \mathcal{C}_c^\infty(\mathbb{R}^d)$. Pour $s < t \in I$, on considère un couplage optimal $\gamma_{s,t} \in \mathcal{M}_0^2(\mu_s, \mu_t)$. On a grâce à l'inégalité de Cauchy-Schwarz :

$$\begin{aligned} |\mu_t(\varphi) - \mu_s(\varphi)| &= \left| \int_{\mathbb{R}^d \times \mathbb{R}^d} (\varphi(y) - \varphi(x)) d\gamma_{s,t}(x, y) \right| \\ &\leq \|\varphi\|_{Lip} \int_{\mathbb{R}^d \times \mathbb{R}^d} |y - x| d\gamma_{s,t}(x, y) = \|\varphi\|_{Lip} W_2(\mu_s, \mu_t). \end{aligned}$$

Comme $t \mapsto \mu_t$ est absolument continue sur I , il existe $m \in L^1(I)$ telle que pour tout $s < t \in I$:

$$W_2(\mu_s, \mu_t) \leq \int_s^t m(\tau) d\tau.$$

Cela entraîne que $t \mapsto \mu_t(\varphi)$ est absolument continue sur I .

* Considérons maintenant la fonction H définie sur $\mathbb{R}^d \times \mathbb{R}^d$ par :

$$H(x, y) = \begin{cases} |\nabla\varphi(x)| & \text{si } y = x \\ \frac{|\varphi(x) - \varphi(y)|}{|x - y|} & \text{si } y \neq x \end{cases}$$

Remarquons qu'on a toujours $|\varphi(y) - \varphi(x)| = |y - x|H(x, y)$, pour tout $x, y \in \mathbb{R}^d$. On écrit :

$$\begin{aligned} \frac{|\mu_{s+\epsilon}(\varphi) - \mu_s(\varphi)|}{|\epsilon|} &= \frac{1}{|\epsilon|} \left| \int_{\mathbb{R}^d \times \mathbb{R}^d} \varphi(y) - \varphi(x) d\gamma_{s, s+\epsilon}(x, y) \right| \\ &\leq \frac{1}{|\epsilon|} \int_{\mathbb{R}^d \times \mathbb{R}^d} |\varphi(y) - \varphi(x)| d\gamma_{s, s+\epsilon}(x, y) \\ &= \frac{1}{|\epsilon|} \int_{\mathbb{R}^d \times \mathbb{R}^d} |y - x| H(x, y) d\gamma_{s, s+\epsilon} \\ &\leq \frac{W_2(\mu_s, \mu_{s+\epsilon})}{|\epsilon|} \left(\int_{\mathbb{R}^d} H^2(x, y) d\gamma_{s, s+\epsilon}(x, y) \right)^{1/2}, \end{aligned}$$

la dernière inégalité provenant de Cauchy-Schwarz. On sait que $\mu_{s+\epsilon} \rightarrow \mu_s$ au sens de W_2 . Cela implique (par la proposition (4.5)) que $\mu_{s+\epsilon}$ converge faiblement vers μ_s . On en déduit que $\{\gamma_{s, s+\epsilon}\}_\epsilon$ est tendue. Quitte à considérer une sous-suite, on peut supposer que $\gamma_{s, s+\epsilon}$ converge faiblement vers une certaine mesure $\hat{\gamma}_s$. On a donc pour $M > 0$, comme $\gamma_{s, s+\epsilon}$ est un couplage optimal :

$$\begin{aligned} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \wedge M d\hat{\gamma}_s(x, y) &= \lim_{\epsilon \rightarrow 0} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \wedge M d\gamma_{s, s+\epsilon} \\ &\leq \liminf_{\epsilon \rightarrow 0} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\gamma_{s, s+\epsilon} \\ &= \lim_{\epsilon \rightarrow 0} W_2(\mu_s, \mu_t)^2 = 0. \end{aligned}$$

De plus $M \mapsto |x - y|^2 \wedge M$ est croissante, et par le théorème de Beppo Lévy on en déduit que :

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\hat{\gamma}_s(x, y) = 0.$$

Cela signifie que la mesure $\hat{\gamma}_s$ est portée par la diagonale $\{x = y\}$. Mais H est bornée et continue sur la diagonale donc pour la mesure $\hat{\gamma}_s$. Ainsi on a la convergence :

$$\lim_{\epsilon \rightarrow 0} \int_{\mathbb{R}^d \times \mathbb{R}^d} H(x, y) d\gamma_{s, s+\epsilon}(x, y) = \int_{\mathbb{R}^d \times \mathbb{R}^d} |H(x, y)| d\hat{\gamma}_s(x, y) = \int_{\mathbb{R}^d} |\nabla\varphi(x)| d\hat{\gamma}_s(x).$$

On en déduit finalement que :

$$\limsup_{\epsilon \rightarrow 0} \frac{|\mu_{s+\epsilon}(\varphi) - \mu_s(\varphi)|}{|\epsilon|} \leq \limsup_{\epsilon \rightarrow 0} \frac{W_2(\mu_s, \mu_{s+\epsilon})}{|\epsilon|} \left(\int_{\mathbb{R}^d} |\nabla\varphi(x)|^2 d\mu_s \right)^{1/2} = |\mu'|_s \|\nabla\varphi\|_{L^2(\mu_s)}. \quad (9.35)$$

* Nous pouvons achever la démonstration. Soit $\psi \in C_c^\infty(I \times \mathbb{R}^d)$. Par le théorème de dérivation à paramètres, on a :

$$\begin{aligned} \int_{I \times \mathbb{R}^d} \frac{\partial}{\partial s} \psi(s, x) d\mu_s(x) ds &= \lim_{\epsilon \rightarrow 0} \int_{I \times \mathbb{R}^d} \frac{1}{\epsilon} (\psi(s, x) - \psi(s - \epsilon, x)) d\mu_s(x) ds \\ &= \lim_{\epsilon \rightarrow 0} \int_I \frac{1}{\epsilon} \left(\int_{\mathbb{R}^d} \psi(s, x) d\mu_s(x) - \int_{\mathbb{R}^d} \psi(s - \epsilon, x) d\mu_s(x) \right) ds. \end{aligned}$$

Mais ψ est à supporte compact, donc il existe $L \times K$ compact de $I \times \mathbb{R}^d$ tel que $Supp(\psi) \subset L \times K$. Comme I est ouvert, on peut prendre $\epsilon > 0$ de sorte que $L + \epsilon \subset I$. Et donc par le changement de variable $s - \epsilon \mapsto s$, on a :

$$\int_L \frac{1}{\epsilon} \left(\int_{\mathbb{R}^d} \psi(s - \epsilon, x) d\mu_s(x) \right) ds = \int_{L+\epsilon} \frac{1}{\epsilon} \left(\int_{\mathbb{R}^d} \psi(s, x) d\mu_{s+\epsilon}(x) \right) ds.$$

Ainsi grâce à (9.35), on obtient :

$$\begin{aligned} \left| \int_{I \times \mathbb{R}^d} \frac{\partial}{\partial s} \psi(s, x) d\mu_s(x) ds \right| &\leq \left(\int_I |\mu'(s)| \|\nabla_x \psi(s, \cdot)\|_{L^2(\mu_s)} ds \right) \\ &\leq \left(\int_I |\mu'|^2(s) ds \right)^{1/2} \left(\int_I \int_{\mathbb{R}^d} |\nabla_x \psi(s, x)|^2 d\mu_s(x) ds \right)^{1/2}. \end{aligned}$$

Posons alors $\mathcal{V} = \{\nabla \psi; \psi \in \mathcal{C}_c^\infty(I \times \mathbb{R}^d)\}$ et $\bar{\mathcal{V}}$ sont adhérence dans $L^2(\tilde{\mu})$, où $d\tilde{\mu} := d\mu_s ds$. La dernière inégalité signifie que l'application

$$\begin{aligned} L : \mathcal{V} &\longrightarrow \mathbb{R} \\ \nabla \psi &\longmapsto - \int_{I \times \mathbb{R}^d} \frac{\partial}{\partial s} \psi(s, x) d\mu_s(x) ds \end{aligned}$$

est bien définie. L est de plus linéaire et donc s'étend de manière unique à $\bar{\mathcal{V}}$ (on note son prolongement encore L). Par le théorème de Riesz-Fisher, il existe donc un élément $V(.,.) \in \bar{\mathcal{V}}$ telle que :

$$L(w) = \langle V, w \rangle_{L^2(\tilde{\mu})} \quad \text{pour } w \in \bar{\mathcal{V}}.$$

En particulier, pour $\nabla \psi \in \mathcal{V}$, on a :

$$- \int_{I \times \mathbb{R}^d} \frac{\partial}{\partial s} \psi(s, x) d\mu_s(x) ds = \int_{I \times \mathbb{R}^d} \langle V(s, x), \nabla_x \psi(s, x) \rangle d\mu_s(x) ds.$$

A savoir, en combinant la dernière inégalité obtenue et en prenant $\nabla_x \psi(s, x) = V(s, x)$, que :

$$\int_I \|V_t\|_{L^2(\mu_t)} dt \leq \int_I |\mu'(t)| dt.$$

Ceci étant vrai pour tout intervalle J ouvert dans I , on en déduit finalement que :

$$\|V_t\|_{L^2(\mu_t)} \leq |\mu'(t)|,$$

pour presque tout $t \in I$ ■

10 Méthode itérative sur des mesures

Notations : $d\gamma_d(x) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{|x|^2}{2}} dx$ représente la mesure Gaussienne sur \mathbb{R}^d . Il s'agira de notre mesure de référence pour toute la suite. On note aussi :

$$\mathcal{K} = \left\{ \rho : \mathbb{R}^d \rightarrow \mathbb{R}_+; \int_{\mathbb{R}^d} \rho d\gamma_d = 1; m_2(\rho\gamma_d) < +\infty \right\}.$$

Et si $\rho \in \mathcal{K}$:

$$\begin{aligned} S(\rho) &= \int_{\mathbb{R}^d} \rho \log \rho d\gamma_d, \\ E_\psi(\rho) &= \int_{\mathbb{R}^d} \psi \rho d\gamma_d + S(\rho) \quad \text{pour } \psi : \mathbb{R}^d \rightarrow \mathbb{R}_+ \text{ mesurable.} \end{aligned}$$

Remarquons que $S(\rho)$ est positif, et qu'il en est de même pour $E_\psi(\rho)$ (que l'on notera $E(\rho)$).

PROPOSITION 10.1 *Soit $\psi : \mathbb{R}^d \rightarrow \mathbb{R}_+$ mesurable. S'il existe une mesure $\rho \in \mathcal{K}$ telle que $E_\psi(\rho) < +\infty$, alors pour tout $\rho^{(0)} \in \mathcal{K}$, et $h > 0$, il existe une unique mesure $\hat{\rho} \in \mathcal{K}$ qui minimise la fonction :*

$$\rho \longmapsto \frac{1}{2} W_2^2(\rho^{(0)}, \rho) + h E_\psi(\rho).$$

Démonstration :

Notons $m = \inf_{\rho \in \mathcal{K}} \{ \frac{1}{2} W_2^2(\rho^{(0)}, \rho) + hE(\rho) \}$. Comme il existe $\rho \in \mathcal{K}$ tel que $E(\rho) < +\infty$, la quantité m est finie. Pour chaque $n \geq 1$, il existe alors $\rho_n \in \mathcal{K}$ tel que :

$$\frac{1}{2} W_2^2(\rho^{(0)}, \rho_n) + hE(\rho_n) \leq \frac{1}{n} + m. \quad (10.36)$$

Cela implique en particulier que $\sup_{n \geq 1} W_2^2(\rho^{(0)}, \rho_n) < +\infty$. Grâce au lemme 4.1, on a :

$$C := \sup_{n \geq 1} \int_{\mathbb{R}^d} |x|^2 \rho_n d\gamma_d < +\infty.$$

Soit $\epsilon > 0$ fixé. Pour $R > 0$ on a :

$$\int_{\mathbb{R}^d} |x|^2 d\rho_n \geq \int_{|x| > R} |x|^2 d\rho_n \geq R^2 \rho_n(|x| > R).$$

Comme la quantité de gauche est uniformément bornée en n , on en déduit que pour R assez grand :

$$\sup_{n \geq 1} \rho_n(|x| > R) \leq \epsilon.$$

Cela signifie que la suite $\{\rho_n \gamma_d\}_{n \geq 1}$ est tendue : quitte à considérer une sous-suite, on peut supposer que $(\rho_n \gamma_d)_n$ converge faiblement vers une mesure μ .

* Pour $R > 0$, on a par définition de la convergence faible :

$$\int |x|^2 \wedge R d\mu = \lim_{n \rightarrow +\infty} \int (|x|^2 \wedge R) \rho_n d\gamma_d \leq C < +\infty.$$

Donc par Beppo Lévi, on a finalement que :

$$m_2(\mu) < +\infty.$$

* (10.36) nous indique aussi que $C_1 := \sup_{n \geq 1} S(\rho_n) < +\infty$. Pour $R > 1$, on a :

$$\int_{\rho_n \geq R} \rho_n d\gamma_d \leq \frac{1}{\log R} \int_{\rho_n \geq R} \rho_n \log \rho_n d\gamma.$$

Mais $\min_{0 \leq x \leq 1} (x \log x) = -\frac{1}{e}$ et par convexité de cette fonction, on a : $x \log x \geq -\frac{1}{e}$ pour tout $x \in \mathbb{R}$. Donc :

$$\int_{\rho_n \geq R} \rho_n \log \rho_n d\gamma_d = S(\rho_n) - \int_{0 \leq \rho_n \leq 1} \rho_n \log \rho_n d\gamma_d \leq S(\rho_n) + \frac{1}{e}.$$

D'où finalement :

$$\int_{\rho_n \geq R} \rho_n d\gamma_d \leq \frac{1}{\log R} (C_1 + \frac{1}{e}) \rightarrow 0 \quad \text{quand } R \rightarrow +\infty$$

donc $\lim_{R \rightarrow +\infty} \sup_{n \geq 1} \int_{\rho_n \geq R} \rho_n d\gamma_d = 0.$

* On a vu que $(\rho_n \gamma_d)_n$ convergeait faiblement vers μ . On veut étendre cette notion de convergence qui est caractérisée sur les fonctions continues bornées, à des fonctions seulement mesurables bornées. Soit $\psi : \mathbb{R}^d \rightarrow \mathbb{R}$ mesurable bornée par C_ψ . On sait que $\mathcal{C}_c(\mathbb{R}^d)$ est dense dans l'ensemble $\mathcal{B}(\gamma_d + \mu) = \{\psi : \mathbb{R}^d \rightarrow \mathbb{R} \text{ mesurable bornée}\}$. Pour $\delta > 0$, il existe donc $\varphi \in \mathcal{C}_c(\mathbb{R}^d)$ telle que $\|\varphi\|_\infty \leq C_\psi$ et :

$$\int |\psi - \varphi| d\gamma_d < \delta, \quad \int |\psi - \varphi| d\mu < \delta.$$

Alors on a :

$$\left| \int \psi \rho_n d\gamma_d - \int \psi d\mu \right| \leq \int |\psi - \varphi| \rho_n d\gamma_d + \int |\psi - \varphi| d\mu + \left| \int \varphi \rho_n d\gamma_d - \int \varphi d\mu \right|. \quad (10.37)$$

Seul le premier terme du membre de droite peut poser problème. Mais :

$$\begin{aligned} \int |\psi - \varphi| \rho_n d\gamma_d &\leq R \int_{\rho_n \leq R} |\psi - \varphi| d\gamma_d + \int_{\rho_n > R} |\psi - \varphi| \rho_n d\gamma_d \\ &\leq R\delta + 2C_\psi \int_{\rho_n > R} \rho_n d\gamma_d. \end{aligned}$$

En faisant $\delta \rightarrow 0$, puis en prenant le *sup* sur n et $R \rightarrow +\infty$, on obtient :

$$\sup_{n \geq 1} \int |\psi - \varphi| \rho_n d\gamma_d = 0.$$

Donc dans (10.37), on fait $\delta \rightarrow 0$, puis on prend le *sup* en n , et $R \rightarrow +\infty$, pour obtenir :

$$\left| \int \psi \rho_n d\gamma_d - \int \psi d\mu \right| \leq \epsilon.$$

On a donc ce qu'on voulait démontrer, à savoir, pour tout $\psi : \mathbb{R}^d \rightarrow \mathbb{R}$ mesurable bornée :

$$\lim_{n \rightarrow +\infty} \int \psi \rho_n d\gamma_d = \int \psi d\mu.$$

* Prenons en particulier pour $E \in \mathcal{B}(\mathbb{R}^d)$, $\psi = 1_E$. La précédente égalité nous dit :

$$\rho_n \gamma_d(E) \rightarrow \mu(E) \quad \text{quand } n \rightarrow +\infty.$$

A savoir : $\gamma_d(E) = 0 \Rightarrow \mu(E)$. Cela signifie que μ est absolument continue par rapport à γ_d . Et le théorème de Radon-Nykodym nous fournit une densité $\hat{\rho} : d\mu = \hat{\rho} \cdot d\gamma_d$. Montrons que $\hat{\rho}$ vérifie bien ce que l'on recherche.

→ On sait que $m_2(\mu) < +\infty$ et donc $\int |x|^2 \hat{\rho} d\gamma_d(x) < +\infty$. On en déduit que $\hat{\rho} \in \mathcal{K}$.

→ Pour $n \geq 1$, on considère un couplage $\gamma_n \in \mathcal{M}_0^2(\rho^{(0)}, \rho_n)$. Comme $(\rho_n \gamma_d)_n$ converge faiblement vers $\hat{\rho} \gamma_d$, la suite $\{\gamma_n\}_n$ est tendue : quitte à considérer une sous-suite, on peut supposer que $(\gamma_n)_n$ converge faiblement vers une mesure $\eta \in \mathcal{M}(\rho^{(0)}, \rho)$. Et pour $R > 0$, on a :

$$\begin{aligned} \int |x - y|^2 \wedge R d\eta &= \lim_{n \rightarrow +\infty} \int |x - y|^2 \wedge R d\gamma_n \leq \liminf_n \int |x - y|^2 d\gamma_n \\ \text{donc } W_2^2(\rho^{(0)}, \hat{\rho}) &\leq \liminf_n \int |x - y|^2 d\gamma_n = \liminf_n W_2^2(\rho^{(0)}, \rho_n). \end{aligned}$$

Par le même raisonnement, on obtient pour $\psi : \mathbb{R}^d \rightarrow \mathbb{R}^d$ mesurable bornée, que :

$$\int \psi \hat{\rho} d\gamma_d \leq \liminf_n \int \psi \rho_n d\gamma.$$

→ Soit $\epsilon > 0$. Traduisons l'inégalité précédente : il existe $n_0 \geq 1$ tel que, pour tout $n \geq n_0$ on ait :

$$\int \psi \hat{\rho} d\gamma_d \leq \epsilon \int \psi \rho_n d\gamma_d,$$

et donc :

$$\frac{1}{2} W_2^2(\rho^{(0)}, \hat{\rho}) + h \int \psi \hat{\rho} d\gamma_d + hS(\rho_n) \leq \frac{1}{2} W_2^2(\rho^{(0)}, \rho_n) + E(\rho_n) + \epsilon.$$

Par (10.36), on en déduit que :

$$\frac{1}{2}W_2^2(\rho^{(0)}, \hat{\rho}) + h \int \psi \hat{\rho} d\gamma_d - \epsilon \leq m + \frac{1}{n}.$$

Considérons alors le sous-ensemble de \mathcal{K} :

$$C = \left\{ \rho \in \mathcal{K}, hS(\rho) \leq \left(m + \frac{1}{n}\right) - \frac{1}{2}W_2^2(\rho^{(0)}, \hat{\rho}) - h \int \psi \hat{\rho} d\gamma_d + \epsilon \right\}.$$

D'après ce qui précède, $\rho_k \in C$ pour tout $k \geq n$ où $n \geq n_0$. De plus C est clairement convexe, grâce à la convexité de $x \mapsto x \log x$. De plus on a vu que $(\rho_n)_n$ converge faiblement vers $\hat{\rho}$, mais que cette convergence a lieu aussi sur les fonctions mesurables bornées (de L^∞). Mais $L^\infty(\mathbb{R}^d, \gamma_d)$ est le dual de $L^1(\mathbb{R}^d, \gamma_d)$, et cela signifie que $(\rho_n)_n$ converge faiblement vers $\hat{\rho}$ dans L^1 . D'où $\hat{\rho} \in \overline{C}^{faible}$. Par le théorème de Hahn-Banach, on sait que :

$$\overline{C}^{faible} = \overline{C}^{forte}.$$

Cela signifie qu'il existe $(\hat{\rho}_q)_q \subset C$ qui converge vers $\hat{\rho}$ dans L^1 . Quitte à considérer une sous-suite, on peut supposer que $(\hat{\rho}_q)_q$ converge vers $\hat{\rho}$ γ_d -presque sûrement. Et par le lemme de Fatou on a :

$$S(\hat{\rho}) \leq \liminf_{q \rightarrow +\infty} S(\hat{\rho}_q).$$

Mais par définition de C on a :

$$hS(\hat{\rho}) \leq hS(\hat{\rho}_q) \leq m + \frac{1}{n} - \frac{1}{2}W_2^2(\rho^{(0)}, \hat{\rho}) - h \int \psi \hat{\rho} d\gamma_d + \epsilon.$$

En faisant $n \rightarrow +\infty$ et $\epsilon \rightarrow 0$, on obtient finalement :

$$\frac{1}{2}W_2^2(\rho^{(0)}, \hat{\rho}) + E(\hat{\rho}) \leq m.$$

Finalement $\hat{\rho}$ réalise le minimum voulu. Si ρ_0 et ρ_1 sont deux mesures de \mathcal{K} qui réalisent ce minimum, alors on va montrer que ces deux mesures sont égales, grâce à des arguments de convexité. Comme la fonction $x \mapsto x \log x$ est convexe, il est clair que $\rho \mapsto E(\rho)$ est convexe. On considère $\rho_t = (1-t)\rho_0 + t\rho_1$ pour $0 \leq t \leq 1$. On a $\rho_t \in \mathcal{K}$ pour tout t entre 0 et 1. Prenons alors des couplages optimaux $\gamma_1 \in \mathcal{M}_0(\rho^{(0)}, \rho_0)$ et $\gamma_2 \in \mathcal{M}_0(\rho^{(0)}, \rho_1)$. On voit alors que la mesure $(1-t)\gamma_1 + t\gamma_2$ appartient à $\mathcal{M}(\rho^{(0)}, \rho_t)$ pour tout $0 \leq t \leq 1$. On a donc si $0 < t < 1$:

$$\begin{aligned} W_2^2(\rho^{(0)}, \rho_t) &< \int |x-y|^2 d((1-t)\gamma_1 + t\gamma_2)(x, y) \\ &= (1-t) \int |x-y|^2 d\gamma_1(x, y) + t \int |x-y|^2 d\gamma_2(x, y) = (1-t)W_2^2(\rho^{(0)}, \rho_0) + tW_2^2(\rho^{(0)}, \rho_1). \end{aligned}$$

E étant convexe et $\rho \mapsto W_2^2(\rho^{(0)}, \rho)$ étant strictement convexe, on a alors :

$$\frac{1}{2}W_2^2(\rho^{(0)}, \rho_t) + E(\rho_t) < (1-t)m + tm = m,$$

et cela contredit que m est la borne inférieure. On en déduit que le minimum atteint est unique. Cela prouve le théorème ■

Partant de $\rho^{(0)} \in \mathcal{K}$, on a construit une unique mesure $\rho^{(1)} \in \mathcal{K}$ qui réalise le minimum de la fonction

$$\rho \mapsto \frac{1}{2}W_2^2(\rho^{(0)}, \rho) + hE(\rho).$$

Partant maintenant de $\rho^{(1)}$, on peut de même construire un $\rho^{(2)}$... Ainsi de suite, et on finit par obtenir une suite de mesures $(\rho^{(n)})_n$ dans \mathcal{K} . Cette méthode d'itérations est en fait un outil très puissant pour résoudre des fonctions aux dérivées partielles, comme celle de Fokker-Plank :

$$\partial_t u - \nabla(\nabla u + u\nabla V) = 0.$$

Appendices

A Théorème de convergence L^p

LEMME A.1 Soit $p \in \mathbb{N}^*$. On considère une suite de variables aléatoires $(X_n)_n$ admettant des moments d'ordre p , toutes définies sur $(\Omega, \mathcal{F}, \mathbb{P})$ et X définie sur $(\Omega, \mathcal{F}, \mathbb{P})$. On a équivalence entre :

1. La suite $(X_n)_n$ converge dans L^p vers X .
2. La suite $(|X_n|^p)_n$ est uniformément intégrable et $(X_n)_n$ converge en probabilité vers X .

Remarquons tout d'abord, que grâce à la croissance de la fonction $x \mapsto x^p$ sur \mathbb{R} et la convexité sur \mathbb{R}_+ , on a pour tout $a, b, c \in \mathbb{R}$:

$$|a - b|^p \leq 2^{p-1}(|a - c|^p + |c - b|^p). \quad (\text{A.38})$$

Démonstration :

1. \Rightarrow 2. Par définition on a $\lim_{n \rightarrow +\infty} \mathbb{E}(|X_n - X|^p) = 0$. L'inégalité de Minkowski, nous dit que pour $n, m \geq 1$, on a :

$$\|X_n - X_m\|_p \leq \|X_n - X\|_p + \|X - X_m\|_p.$$

Cela signifie que $(X_n)_n$ est de Cauchy dans L^p . Soit $N \geq 1$ tel que, pour $n, m \geq N$: $\mathbb{E}(|X_n - X_m|^p) \leq \epsilon/2^p$. Par l'inégalité (A.38), on a pour $A \in \mathcal{F}$ et $n \geq N$:

$$\int_A |X_n|^p dP \leq 2^{p-1} \left(\int_A |X_N|^p dP + \int_A |X_n - X_N|^p dP \right) \leq 2^{p-1} \int_A |X_N|^p dP + \frac{\epsilon}{2}.$$

Et donc en prenant le *sup* sur n , on a pour tout $A \in \mathcal{F}$:

$$\sup_{n \geq 1} \int_A |X_n|^p dP \leq \sup_{n \leq N} \int_A |X_n|^p dP + 2^{p-1} \int_A |X_N|^p dP + \epsilon/2.$$

On en déduit que $\{|X_n|^p\}_{n \geq 1}$ est bornée dans L^1 et donc uniformément intégrable dans L^1 . Par l'inégalité de Markov, on a pour $\epsilon > 0$, et $n, m \geq 1$:

$$P(|X_n - X_m| > \epsilon) \leq \frac{1}{\epsilon^p} \mathbb{E}(|X_n - X_m|^p).$$

Et donc la suite $(X_n)_n$ est de Cauchy en probabilité et converge donc en probabilité vers X .

2. \Rightarrow 1. Soit $\epsilon > 0$. Il découle de l'inégalité (A.38) que :

$$\begin{aligned} \mathbb{E}|X_n - X|^p &\leq \int_{|X_n - X| \leq \epsilon^{1/p}} |X_n - X|^p dP + 2^{p-1} \int_{|X_n - X| > \epsilon^{1/p}} (|X_n|^p + |X|^p) dP \\ &\leq \epsilon + 2^{p-1} \left(\int_{|X_n - X| > \epsilon^{1/p}} |X_n|^p dP + \int_{|X_n - X| > \epsilon^{1/p}} |X|^p dP \right). \end{aligned}$$

La famille $\{|X_n|^p, n \geq 1; |X|^p\}$ étant équi-continue, il existe $\eta > 0$ tel que si $P(A) \leq \eta$ alors :

$$\sup_{n \geq 1} \int_A |X_n|^p dP + \int_A |X|^p dP \leq \frac{\epsilon}{2^{p-1}}.$$

Et par hypothèse, $(X_n)_n$ converge en probabilité vers X donc il existe $N \geq 1$, tel que pour tout $n \geq N$:

$$P(|X_n - X| > \epsilon^{1/p}) \leq \eta.$$

On en déduit finalement que :

$$\limsup_{n \rightarrow +\infty} \mathbb{E}|X_n - X|^p \leq 2\epsilon.$$

Ceci étant vrai pour tout $\epsilon > 0$, il en résulte que $(X_n)_n$ converge dans L^p vers X . ■

Références

- [1] G. Savare L. Ambrosio, N. Gigli. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. 2008.
- [2] P-M. Samson. Concentration of measure inequalities for markov chains and ϕ -mixing processes. *The Annals of Probability*, 2000.
- [3] A.D. Barbour and Louis H.Y. Chen. *An introduction to Stein's Method*.
- [4] S. Fang. Wasserstein space and fokker planck equation. 2007.
- [5] Franck et Patrick Gabriel. Notes sur les couplages discrets, exemples avec la loi de poisson.
- [6] T. Lindvall. *Lectures on the Coupling Method*. Wiley-Interscience, 1992.