Distance en variation totale et couplages

v1 février 2017 / relue très rapidement

1 Distance en variation totale et couplages

Cadre. On considère un ensemble E fini ou dénombrable. On munit E de la tribu discrète. Si μ est une mesure de probabilité sur E et si $x \in E$ on écrit $\mu(x)$ pour $\mu(\{x\})$.

Définition. Soient μ et ν deux mesures de probabilité sur E. On appelle distance en variation totale entre μ et ν la quantité

$$d(\mu, \nu) = \sup\{|\mu(A) - \nu(A)|, A \subset E\}.$$

Remarques.
- Cela définit une distance sur l'ensemble des mesures de probabilité sur E.
- Cette distance est à valeurs dans $[0, 1]$.

Définition. Si μ et ν sont deux mesures de probabilité sur E, on appelle couplage de μ et ν tout couple (X, Y) de variables aléatoires à valeurs dans E tel que X suit la loi μ et Y suit la loi ν.

Théorème 1 *Soient μ et ν deux mesures de probabilité sur E. Alors*

$$\begin{aligned} d(\mu, \nu) &= \max\{|\mu(A) - \nu(A)|, A \subset E\} \\ &= \frac{1}{2} \sum_{x \in E} |\mu(x) - \nu(x)| \\ &= 1 - \sum_{x \in E} \mu(x) \wedge \nu(x) \\ &= \min\{P(X \neq Y),\ (X, Y) \text{ couplage de } \mu \text{ et } \nu\} \end{aligned}$$

Remarque. Le théorème reste vrai (avec quelques aménagements dans les énoncés) dans un cadre général : E ensemble quelconque muni d'une tribu quelconque. La preuve reste essentiellement la même. Décrivons maintenant les aménagements nécessaires.
- Commençons par un autre cas particulier, celui de mesures sur les boréliens de $\mathbb{R}$ admettant une densité par rapport à la mesure de Lebesgue. Soient μ et ν deux mesures admettant pour densité f et g par rapport à la mesure de Lebesgue. Alors la deuxième égalité (par exemple) devient

$$d(\mu, \nu) = \frac{1}{2} \int_{\mathbb{R}} |f(x) - g(x)| dx.$$

— Dans le cadre général, μ et ν sont absolument continues par rapport à la mesure $m = \mu + \nu$. Par le théorème de Radon-Nikodym, μ et ν admettent donc des densités f et g par rapport à m. La deuxième égalité (par exemple) devient

$$d(\mu, \nu) = \frac{1}{2} \int_E |f(x) - g(x)| m(dx).$$

Preuve. Posons

$$A_* = \{x \in E : \mu(x) > \nu(x)\}.$$

Pour tout $A \subset E$ on a On a

$$m := \mu(A_*^c) - \nu(A_*^c) \leq \mu(A) - \nu(A) \leq \mu(A_*) - \nu(A_*) =: M.$$

Mais

$$m + M = \mu(E) - \nu(E) = 0.$$

Ainsi $m = -M$. On en déduit

$$|\mu(A) - \nu(A)| \leq M$$

puis

$$d(\mu, \nu) = M = |\mu(A_*) - \nu(A_*)|.$$

Cela prouve le premier point : le supremum est atteint.

De ce qui précède, on déduit également

$$d(\mu, \nu) = \frac{1}{2}(M + M) = \frac{1}{2}(M - m) = \frac{1}{2} \sum_{x \in E} |\mu(x) - \nu(x)|.$$

La deuxième égalité est établie.

La troisième égalité se déduit simplement de ce qui précède et de l'égalité suivante, valable pour tout $a, b \in \mathbb{R}$,

$$a \wedge b = \frac{a + b}{2} - \frac{|a - b|}{2}.$$

La dernière égalité demande plus de travail. Soit (X, Y) un couplage de μ et ν. Pour tout $A \subset E$ on a

$$\begin{aligned}
\mu(A) - \nu(A) &= P(X \in A) - P(Y \in A) \\
&\leq P(X \in A) - P(Y \in A \text{ et } X = Y) \\
&= P(\{X \in A\} \setminus \{Y \in A \text{ et } X = Y\}) \\
&\leq P(X \neq Y).
\end{aligned}$$

On obtient similairement $\nu(A) - \mu(A) \leq P(X \neq Y)$. Ainsi

$$|\mu(A) - \nu(A)| \leq P(X \neq Y)$$

puis

$$d(\mu, \nu) \leq \inf_{(X,Y)} P(X \neq Y) \tag{1}$$

où l'infimum porte sur tous les couplages de μ et ν.

Pour conclure, il reste à montrer qu'il existe un couplage (X, Y) tel que

$$d(\mu, \nu) = P(X \neq Y).$$

Les cas $d(\mu,\nu)=0$ et $d(\mu,\nu)=1$ sont élémentaires. Supposons dans la suite que ce n'est pas le cas. Autrement dit, supposons $0<d(\mu,\nu)<1$. Posons

$$\alpha = 1-d(\mu,\nu)=\sum_{x\in E}\mu(x)\wedge\nu(x).$$

Soient U,V,W trois v.a. indépendantes à valeurs dans E. On suppose, pour tout $x\in E$,

$$\begin{aligned}
P(U=x) &= \frac{\mu(x)\wedge\nu(x)}{\alpha},\\
P(V=x) &= \frac{\mu(x)-\mu(x)\wedge\nu(x)}{1-\alpha},\\
P(W=x) &= \frac{\nu(x)-\mu(x)\wedge\nu(x)}{1-\alpha}.
\end{aligned}$$

Soit B une v.a. de Bernoulli de paramètre α indépendante de U,V et W. On définit enfin X et Y de la manière suivante. Si $B=1$, on pose $X=U$ et $Y=U$. Sinon, on pose $X=V$ et $Y=W$. On vérifie que X est de loi μ et que Y est de loi ν. Par ailleurs, on a

$$P(X\neq Y)\leq P(B=0)=1-\alpha=d(\mu,\nu).$$

On peut vérifier directement que l'inégalité précédente est une égalité, mais on peut également se contenter de remarquer que l'inégalité précédente et (1) permettent de conclure. □

2 Distance en variation totale et convergence en loi

Cadre. On considère un ensemble E fini ou dénombrable. On munit E de la tribu discrète et de la topologie discrète. Si μ est une mesure de probabilité sur E et si $x\in E$ on écrit $\mu(x)$ pour $\mu(\{x\})$.

Théorème 2 *Soit $(\mu_n)_n$ une suite de mesures de probabilité sur E. Soit μ une mesure de probabilité sur E. Les propriétés suivantes sont équivalentes.*

1. *Convergence étroite : pour toute application $\phi : E\to\mathbb{R}$ continue et bornée, $\int\phi d\mu_n\to\int\phi d\mu$.*
2. *Convergence simple : pour tout $x\in E$, $\mu_n(x)\to\mu(x)$.*
3. *Convergence en variation totale : $d(\mu_n,\mu)\to 0$.*

Rappel. Soit $(X_n)_n$ une suite de v.a. à valeurs dans E. Soit X une v.a. à valeurs dans E. On dit que $(X_n)_n$ converge en loi vers X lorsque la suite des lois des X_n converge étroitement vers la loi de X.

Remarques.

- Comme E est muni de la topologie discrète, toute application de E dans $\mathbb{R}$ est continue. La topologie induite par celle de $\mathbb{R}$ sur l'ensemble $\{0\}\cup\{1/n,n\geq 1\}$ est-elle discrète ?
- La convergence en variation totale peut s'énoncer comme suit :

  $$\sum_{x\in E}|\mu_n(x)-\mu(x)|\to 0.$$

  C'est donc une convergence dans L^1.
- Les implications « 3 implique 2 » et « 3 implique 1 » restent vraies dans un cadre général.
- Les autres implications sont fausses en toute généralité. Considérer par exemple la suite des $\delta_{1/n}$ ou des exemples avec des mesures de probabilité sans atomes.

Preuve. Montrons « 3 implique 1 ». Soit $\phi : E \to \mathbb{R}$ continue et bornée. Pour tout entier n on a

$$\begin{aligned}
\left|\int \phi d\mu_n - \int \phi d\mu\right| &= \left|\sum_{x\in E} \phi(x)\mu_n(x) - \sum_{x\in E} \phi(x)\mu(x)\right| \\
&\leq \|\phi\|_\infty \sum_{x\in E} |\mu_n(x) - \mu(x)| \\
&= 2\|\phi\|_\infty d(\mu_n, \mu).
\end{aligned}$$

On en déduit l'implication recherchée. Une preuve plus robuste repose sur l'existence d'un couplage optimal.

L'implication « 1 implique 2 » est élémentaire : pour tout $x \in E$, $1_{\{x\}}$ est continue.

Prouvons maintenant « 2 implique 3 ». On suppose qu'il y a convergence simple. Soit $\varepsilon > 0$. On fixe un sous-ensemble fini $A \subset E$ tel que

$$\mu(A) \geq 1 - \varepsilon.$$

On fixe N tel que, pour tout $n \geq N$,

$$\sum_{x\in A} |\mu_n(x) - \mu(x)| \ \leq \varepsilon.$$

On a

$$\begin{aligned}
\sum_{x\in A^c} |\mu_n(x) - \mu(x)| &\leq \sum_{x\in A^c} \mu_n(x) + \sum_{x\in A^c} \mu(x) \\
&= 1 - \sum_{x\in A} \mu_n(x) + \sum_{x\in A^c} \mu(x) \\
&= 1 - \sum_{x\in A} \mu(x) + \sum_{x\in A} \big(\mu(x) - \mu_n(x)\big) + \sum_{x\in A^c} \mu(x) \\
&\leq 3\varepsilon.
\end{aligned}$$

Cela conclut la preuve. □

Remarque. L'implication « 2 implique 3 » est un cas particulier du lemme de Scheffé dont voici une version. Soient (f_n) une suite d'applications mesurables et positives. Soit f une application mesurable et positive. Toutes ces applications sont définies sur un même espace mesuré. On suppose que les intégrales de toutes ces applications valent 1. On suppose que $(f_n)_n$ converge presque partout vers f. Alors $(f_n)_n$ converge dans L^1 vers f. Indications pour la preuve :

$$(f - f_n)_+ \leq f \text{ et } |f - f_n| = 2(f - f_n)_+ + (f - f_n).$$

3 Inégalité de Le Cam

Lemme 3 *Soit $p \in [0, 1]$. Il existe un couple de v.a. (X, Y) tel que X suit une loi de Bernoulli de paramètre p, Y suit une loi de Poisson de paramètre p et $P(X \neq Y) \leq p^2$.*

Preuve 1. Voici une première preuve reposant sur les outils développés précédemment. En utilisant la troisième égalité du premier théorème, on obtient

$$d(\mathcal{B}(p), \mathcal{P}(p)) = 1 - (1-p) \wedge e^{-p} - p \wedge pe^{-p} = p - pe^{-p} = p(1 - e^{-p}) \leq p^2.$$

La quatrième égalité du théorème 1 permet alors de conclure. □

Preuve 2. Voici une preuve plus pédestre mais reposant en partie sur les mêmes idées. Soit U une v.a. de loi uniforme sur $[0,1]$. On pose

$$X = 1_{]1-p,1]} \text{ et } Y = \sum_{n\geq 1} n 1_{]q_{n-1},q_n]}$$

où, pour tout entier $n \geq 0$, q_n est la probabilité qu'une variable aléatoire de loi de Poisson de paramètre p soit inférieure ou égale à n. Alors X suit une loi de Bernoulli de paramètre p tandis que Y suit une loi de Poisson de paramètre p. Notons que c'est simplement la construction de Skorokhod (mais prouver que X et Y suivent les lois indiquées se fait facilement à la main). On constate

$$P(X = Y) = P(X = Y = 0) + P(X = Y = 1) = 1 - p + pe^{-p}$$

d'où

$$P(X \neq Y) = p(1 - e^{-p}) \leq p^2.$$

Cela conclut la preuve. □

Théorème 4 (Inégalité de Le Cam) *Soit $(p_1, \dots, p_n)$ une suite de réels de $[0,1]$. Soit S la somme de n v.a. indépendantes de lois de Bernoulli de paramètres $p_1, \dots, p_n$. Notons λ la somme de p_i. Alors*

$$d(loi(S), \mathcal{P}(\lambda)) \leq \sum_i p_i^2.$$

Preuve. On se donne, grâce au lemme précédent, $(X_1, Y_1), \dots, (X_n, Y_n)$ des couples indépendants des v.a. tels que, pour tout $i \in \{1, \dots, n\}$:
- X_i est de loi $\mathcal{B}(p_i)$.
- Y_i de loi $\mathcal{P}(p_i)$.
- $P(X_i \neq Y_i) \leq p_i^2$.

Alors :
- $X_1 + \cdots + X_n$ a la même loi que S.
- $Y_1 + \cdots + Y_n$ suit la loi $\mathcal{P}(\lambda)$.
- $P(X_1 + \cdots + X_n \neq Y_1 + \cdots + Y_n) \leq \sum_i P(X_i \neq Y_i) \leq \sum_i p_i^2$.

On conclut alors avec la dernière inégalité du premier théorème. On peut noter que l'on a en fait uniquement besoin de (1), inégalité qui se prouve de simplement en quelques lignes. □

Remarque. L'inégalité de Le Cam permet en particulier d'établir des convergences en loi de somme de Bernoulli vers une loi de Poisson. Elle permet notamment de retrouver le résultat élémentaire suivant : pour tout $\lambda > 0$, la loi binomiale de paramètres n et λ/n converge en loi vers la loi de Poisson de paramètre λ lorsque n tend vers l'infini.