



HAL
open science

Sur la distance de deux lois de probabilité

Maurice Fréchet

► **To cite this version:**

Maurice Fréchet. Sur la distance de deux lois de probabilité. Annales de l'ISUP, 1957, VI (3), pp.183-198. hal-04093677

HAL Id: hal-04093677

<https://hal.science/hal-04093677>

Submitted on 10 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

SUR LA DISTANCE DE DEUX LOIS DE PROBABILITÉ

par

Maurice FRÉCHET

RESUME : L'auteur indique une expression explicite de la distance de deux lois de probabilités, selon la première définition de Paul Lévy. Il indique aussi une modification commode de cette définition.

INTRODUCTION

M. Paul Lévy a résolu tant de problèmes qu'il n'a pas toujours eu le temps d'exploiter toutes ses idées.

Par exemple, il a donné (1) trois intéressantes définitions de la distance "de deux lois de probabilité" en laissant aux lecteurs le soin de les utiliser.

J'ai pensé qu'une bonne manière de rendre hommage à l'oeuvre si riche de M. Paul Lévy, consisterait à prolonger son étude.

Après une brève comparaison de ses trois définitions, j'étudierai plus particulièrement la première.

Mon étude a été facilitée par les résultats énoncés dans ma note aux C.R. (2) "Sur les tableaux de corrélation dont les marges sont données" et par l'application qui en a été donnée par Bass (3) et par Dall'Aglio (4).

Rappel :

Rappelons que :

Une distance est définie sur un ensemble E si l'on associe à tout couple a, b d'éléments de E , un nombre réel appelé distance de a et de b et représenté par la notation (a, b) , de façon à vérifier les conditions suivantes :

$$1^\circ (a, b) = (b, a) \geq 0 ;$$

2° $(a, b) \neq 0$ si a et b sont considérés comme distincts dans E et inversement ;

(1) Voir p. 329 de la 2e édition de notre ouvrage : "Recherches théoriques modernes sur le Calcul des Probabilités (Premier livre), chez Gauthier-Villars, 1950.

(2) Tome 242, 1956, p. 2426-2428.

(3) C.R. 21 février 1955, tome 240, n° 8.

(4) Sugli estremi dei momenti delle funzione di ripartizione doppia (Ann. Scuole Norm. Sup. di Pisa, Sc. Fis. a Mat., série III, Vol. X, 1956, p. 35-74).

3° Pour 3 éléments quelconques a, b, c de E , on a "l'inégalité triangulaire" :

$$(a, b) \leq (a, c) + (c, b).$$

En outre, dire que a_n de E converge vers a de E , c'est dire que la distance (a_n, a) tend vers zéro avec $\frac{1}{n}$.

Distinction entre deux "distances" :

Appliquons cette définition au Calcul des Probabilités. Dans cette théorie, il y a deux distances à distinguer. Soient X, Y deux éléments aléatoires, déterminés simultanément dans chaque épreuve d'une catégorie déterminée.

On peut définir une distance (X, Y) de X, Y dans une épreuve déterminée. Mais il y a aussi intérêt à définir une "distance globale" de X et Y dans l'ensemble des épreuves. Nous représenterons cette distance globale par la notation :

$$([X], [Y]).$$

Il faudra s'assurer à chaque fois que les conditions 1°, 2°, 3° sont vérifiées. Mais en général, on conviendra, en ce qui concerne la condition 2°, de considérer comme non distincts deux éléments aléatoires qui ne sont que "presque certainement" identiques. Par exemple, nous avons prouvé que :

$$\text{Prob}(X \neq Y)$$

est une distance globale.

Une distance globale plus souvent considérée est rappelée en note ci-après. Pour nous en tenir au cas où X, Y sont deux nombres aléatoires, observons que pour déterminer une distance globale, il ne suffit pas de connaître les fonctions de répartition respectives :

$$F(x) = \text{Prob.}(X < x) \quad ; \quad G(y) = \text{Prob.}(Y < y)$$

qui déterminent les lois de probabilité L, L' de X et de Y , mais aussi la fonction de répartition :

$$H(x, y) = \text{Prob.}(X < x \text{ et } Y < y)$$

du couple ordonné X, Y (1).

Au contraire, la distance (L, L') des deux lois L, L' , doit être déterminée indépendamment de $H(x, y)$ par $F(x)$ et $G(y)$.

(1) Par exemple, on prend souvent comme "distance globale" de X et de Y , la quantité

$$\sqrt{\mathcal{N}_H(X-Y)^2} = \sqrt{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x-y)^2 d_x d_y H(x, y)}$$

Les deux dernières définitions de (L, L') par P. Lévy sont purement analytiques. Elles se présentent comme modifications de définitions traditionnelles en Analyse de la distance de deux fonctions F et G , modifications motivées par les discontinuités possibles des deux fonctions F et G .

La première définition proposée par M. Paul Lévy est, au contraire, une définition qui pourrait légitimement paraître arbitraire et artificielle (si on l'énonçait sous forme purement analytique) à un mathématicien ignorant du Calcul des Probabilités. Elle fait, en effet, appel à la notion de nombre aléatoire et c'est quand on fait effectivement appel à cette notion qu'elle prend, au contraire, un sens intuitif et naturel.

Il va être curieux de constater que cette première définition conduit à une expression explicite assez simple de la distance, contrairement à ce qu'on aurait pu attendre.

Rappel :

La première définition proposée par M. Paul Lévy est la suivante :

La distance de deux lois de probabilité L, L' , dont les fonctions de répartition sont $F(x), G(y)$, est la borne inférieure des "distances globales" de deux nombres aléatoires X, Y dont la fonction de répartition $H(x, y)$ est quelconque pourvu qu'elle donne à X, Y , séparément, les fonctions de répartition $F(x), G(y)$.

Marges :

On peut dire, si l'on veut, que $F(x), G(y)$ sont les "marges" du "tableau de corrélation" défini par $H(x, y)$.

Il nous sera précieux d'utiliser le résultat énoncé dans notre Note aux C.R. du 14 mai 1956 rappelée ci-dessus, suivant lequel, l'ensemble des fonctions $H(x, y)$ n'est autre que celui des fonctions de répartition comprises (extrémités incluses) entre les fonctions de répartition $H_0(x, y)$ et $H_1(x, y)$, ($H_0 \leq H_1$), définies par

$$H_0(x, y) = \text{Sup.} \left[(F(x) + G(y) - 1), 0 \right]$$

$$H_1(x, y) = \text{Inf.} \left[F(x), G(y) \right].$$

Observons que dans cette première définition intervient une distance globale de X et de Y qui peut s'interpréter de différentes façons. Nous en avons donné un premier exemple p. 2 et un second en Note (1), p. 2. Autre exemple : pour obtenir une distance globale de X et Y , soit :

$$([X], [Y])$$

qui tende vers zéro quand X tend vers Y en probabilité et réciproquement, nous avons proposé de prendre pour cette distance la borne inférieure quand le nombre positif ε varie, de la somme

$$\varepsilon + \text{Prob.} (|X - Y| > \varepsilon)$$

On peut aussi prendre pour distance globale de X et de Y , la valeur moyenne de $f(|X - Y|)$ où f est une fonction continue, croissante bornée, définie pour $x \geq 0$, nulle pour $x = 0$ et telle que $f(a+b) \leq f(a) + f(b)$ (On

peut d'ailleurs se contenter de conditions un peu plus générales; voir 2° et 3° parties de mon article paru en 1957 dans le *Giornale Italiane Attuari*].

J'ai donné comme exemples de $f(x)$ les fonctions $\frac{x}{1+x}$, $|\operatorname{th} x|$, $|1 - e^{-x}|$ (1).

Les conditions d'existence d'une distance :

Pour sa première définition de la distance de deux lois L, L' , Lévy montre que cette distance satisfait à la condition 3° ci-dessus p. 2. Il est évident que la condition 1° est satisfaite. La question est moins simple pour la condition 2°.

Il est clair cependant que si deux lois L, L' sont identiques, l'un des couples X, Y satisfaisant respectivement à ces deux lois sera le couple X, X dont la distance est nulle. La borne inférieure des distances $([X], [Y])$ sera donc nulle et (L, L') sera ainsi nul. Le point le plus délicat est la proposition inverse. Quand $(L, L') = 0$, L, L' sont-elles nécessairement identiques ?

Précisons par la notation $([X], [Y])_H$ le fait qu'il s'agit de la distance globale de X et Y quand la fonction de répartition de leur couple est $H(x, y)$.

Dans le cas où cette distance atteint sa borne inférieure (L, L') pour au moins une fonction H particulière, soit $K(x, y)$, alors, par définition de $([X], [Y])_H$ si cette quantité est nulle, X et Y sont non distincts (ou bien identiques dans toute épreuve ou bien "presque certainement" égaux) et alors $F(x) \equiv G(y)$, c'est-à-dire $L \equiv L'$. Mais si la borne inférieure n'est pas atteinte, le raisonnement est moins simple.

Dire que $(L, L') = 0$, c'est dire qu'il existe pour tout n entier, un couple X_n, Y_n admissible (c'est-à-dire tel que X_n, Y_n aient pour lois respectives L, L') pour lequel la distance :

$$(1) \quad ([X_n], [Y_n]) < \frac{1}{n} .$$

Mais on n'a pas prescrit d'avance le choix adopté pour cette dernière distance. Et nous avons montré qu'on peut en concevoir beaucoup qui vérifient les conditions 1°, 2°, 3° ci-dessus et qui soient conformes à notre intuition.

Par exemple, nous avons montré qu'on peut définir des distances $([X], [Y])$ très différentes ayant la propriété commune suivante :

La condition nécessaire et suffisante pour que :

$$\lim_{n \rightarrow \infty} ([X_n], [X]) = 0$$

est que X_n converge "en probabilité" vers X quand $n \rightarrow \infty$.

(1) Voir pour ces deux définitions, les p. 205 et 226 de la 2^e Edition de mes "Recherches théoriques modernes, sur le Calcul des Probabilités", Gauthier-Villars, (1950).

Adoptons une telle définition dans l'énoncé de la définition de la distance de deux lois L, L' .

Partons de (1), on a, pour tout $r > 0$:

$$(2) \quad G(x-\varepsilon) = \text{Prob.} [Y_n < x-\varepsilon] \leq \text{Prob} [X_n < x] + \text{Prob} [|X_n - Y_n| > \varepsilon] \\ = F(x) + \text{Prob.} (|X_n - Y_n| > \varepsilon).$$

Pour les différentes définitions que nous avons données d'une distance de deux nombres aléatoires, compatible avec la convergence en probabilité, on a aussi :

$$(3) \quad ([X], [Y]) = ([X - Y], 0).$$

Par suite, d'après (1) et (3), on a :

$$([X_n - Y_n], 0) < \frac{1}{n}$$

donc $X_n - Y_n$ converge vers zéro en probabilité c'est-à-dire que :

$$(4) \quad 0 = \lim_{n \rightarrow \infty} \text{Prob.} (|X_n - Y_n| > \varepsilon)$$

pour tout $\varepsilon > 0$. Alors d'après (2), en faisant croître n vers l'infini, on a :

$$G(x-\varepsilon) \leq F(x) \quad \text{pour tout } \varepsilon > 0.$$

Et comme $G(x)$ est continu à gauche :

$$G(x) \leq F(x).$$

Mais on verrait de même que :

$$F(x) \leq G(x);$$

donc : $F(x) = G(x)$.

Ainsi, s'il n'était pas évident que la condition 2° soit satisfaite, cela est maintenant démontré, au moins quand la définition de (L, L') est appliquée dans le cas où la distance de deux nombres aléatoires satisfait aux conditions ci-dessus.

Cas où $([X], [Y])$ était égal à l'écart quadratique moyen de X et Y .

Dans la suite, nous prendrons d'abord comme distance globale de deux nombres aléatoires X et Y , leur écart quadratique moyen

$$\sqrt{\mathcal{M}_H(X - Y)^2}.$$

(Pour préciser, nous écrirons \mathcal{M}_H pour rappeler qu'il s'agit de la moyenne quand la fonction de répartition du couple (X, Y) est $H(x, y)$).

Alors la distance Δ des lois de probabilité L, L' caractérisées par les fonctions de répartition $F(x)$ et $G(y)$ pourra être représentée par la formule :

$$\Delta^2 = \text{Inf.}_H \mathcal{M}_H(X - Y)^2$$

Observons qu'on a :

$$(X - Y)^2 \leq 2(X^2 + Y^2) \quad \text{d'où : } \mathcal{M}_H(X - Y)^2 \leq 2\mathcal{M}_H X^2 + 2\mathcal{M}_H Y^2.$$

Donc si $\mathcal{M}_H(X - Y)^2$ est infini, l'une au moins des quantités :

$$\int_{-\infty}^{+\infty} x^2 dF(x) = \mathcal{M}_H X^2 \quad \int_{-\infty}^{+\infty} y^2 dG(y) = \mathcal{M}_H Y^2, \text{ est infinie.}$$

Pour éviter des complications, nous supposons que $F(x)$ et $G(x)$ sont les fonctions de répartition de nombres aléatoires ayant chacun un écart quadratique moyen fini. Alors $\mathcal{M}_H(X - Y)^2$ sera nécessairement fini. De plus, X et Y pris séparément, ont, sous la même hypothèse, des moyennes finies et déterminées :

$$a = \mathcal{M} X, \quad a' = \mathcal{M} Y;$$

soient aussi σ, σ' , leurs écarts quadratiques moyens. Alors,

$$Z = \frac{X - a}{\sigma} \quad T = \frac{Y - a'}{\sigma'}$$

seront les nombres aléatoires réduits correspondant à X et Y . C'est-à-dire que les moyennes de Z et T sont nulles et leurs écarts quadratiques moyens égaux à l'unité. On aura :

$$\begin{aligned} \mathcal{M}_H(X - Y)^2 &= \mathcal{M}(a - a' + \sigma Z - \sigma' T)^2 \\ &= (a - a')^2 + 2(a - a')(\sigma \mathcal{M} Z - \sigma' \mathcal{M} T) + \sigma^2 \mathcal{M} Z^2 - 2\sigma\sigma' \mathcal{M} ZT + \sigma'^2 \mathcal{M} T^2 \end{aligned}$$

qui, d'après ce qui précède, se réduit à

$$(a - a')^2 + \sigma^2 - 2\sigma\sigma' r + \sigma'^2 \quad (1)$$

en désignant, quand $\sigma\sigma' \neq 0$, par r le coefficient de corrélation (linéaire) de X et de Y . On sait que $|r| \leq 1$. On peut donc poser $r = \cos \theta$, d'où

$$(5) \quad \sqrt{(\mathcal{M}_H(X - Y)^2)} = \sqrt{(a - a')^2 + (\sigma^2 - 2\sigma\sigma' \cos \theta + \sigma'^2)}.$$

Cette quantité peut donc être décrite comme l'hypoténuse d'un triangle rectangle, dont un côté a pour longueur l'écart $|a - a'|$ entre les moyennes de X et de Y et dont l'autre côté est le troisième côté d'un triangle dont deux côtés ont pour longueurs σ et σ' , l'angle compris entre ces deux côtés étant égal à θ (ou encore le cosinus de cet angle étant égal au coefficient de corrélation (linéaire) de X et de Y).

On observera que, dans l'expression (5), a, a', σ, σ' sont déterminés connaissant F et G et c'est seulement $r = \cos \theta$ qui dépend de $H(x, y)$. Le minimum de $\mathcal{M}_H(X - Y)^2$ aura donc lieu quand $r = \cos \theta$ sera le plus grand possible. Plus précisément la distance entre les deux lois de probabilités est donc :

$$(6) \quad (L, L') = \sqrt{(a - a')^2 + (\sigma^2 - 2\sigma\sigma'\rho + \sigma'^2)}$$

(1) Cette expression a été donnée déjà dans le cas de nombres aléatoires discrets par Salvemini dans: "L'indice quadratico di dissomiglianza era distribuzione gaussiana". Atti della XIV Reunione Scientifica, Societa Italiana di Statistica, 1954.

où ρ est la borne supérieure du coefficient de corrélation (linéaire) entre X et Y , quand la loi de probabilité du couple X, Y varie de façon à laisser fixes les lois de probabilité de X et de Y pris séparément.

Observons d'ailleurs que :

$$r = \frac{\iint (x-a)(y-a') dH(x,y)}{\sqrt{\left[\int (x-a)^2 dF\right] \left[\int (y-a')^2 dG\right]}}$$

Posons : $h(z,t) = H(a + \sigma z, a' + \sigma' t)$. Alors : $r = \frac{\iint \sigma \sigma' z t dh(z,t)}{\sqrt{\dots}}$

$$(7) \quad r = \iint_{\mathcal{P}} z t dh(z,t).$$

De sorte que r est aussi le coefficient de corrélation (linéaire) de Z et de T et est exprimé par la formule (7) où $h(z,t)$ est une fonction de répartition dont les marges sont données et identiques à

$$\varphi(z) \equiv F(a + \sigma z) \quad \psi(t) \equiv G(a' + \sigma' t)$$

Tout le problème est ramené à trouver la borne supérieure de l'expression :

$$(8) \quad r = \iint_{-\infty}^{+\infty} z t dh(z,t)$$

quand h est une fonction de répartition dont les marges $\varphi(z)$ et $\psi(t)$ sont données. On sait, que $r = \cos \theta = +1$ dans le cas où il y a une relation fonctionnelle linéaire croissante

$$Y = mX + p \quad \text{avec } m > 0,$$

entre X et Y . Dans ce cas, on a donc plus simplement

$$(9) \quad (L, L') = \sqrt{(a - a')^2 + (\sigma - \sigma')^2}.$$

Dans l'article de 1954 cité en note plus haut, Salvemini avait conjecturé que le minimum de :

$$J = \iint_{-\infty}^{+\infty} (x-y)^2 d_x d_y H(x,y)$$

est atteint pour H identique à la fonction $H_1(x,y)$ citée plus haut.

En 1955, Bass a énoncé dans sa note citée plus haut, ce résultat équivalent (d'après la formule (6)), que le maximum du coefficient de corrélation (linéaire) est obtenu pour $H \equiv H_1$. Sa démonstration (qu'il a bien voulu me communiquer par écrit) suppose que X et Y sont bornées.

L'année suivante, dans son article cité plus haut, Dall'Aglio démontre que J est minimum pour $H = H_1$, dans l'hypothèse plus générale où

$$0 = \lim_{x \rightarrow -\infty} x^\delta F(x) = \lim_{y \rightarrow -\infty} y^\delta$$

$$0 = \lim_{x \rightarrow +\infty} x^\delta [1 - F(x)] = \lim_{y \rightarrow +\infty} y^\delta [1 - G(y)]$$

pour $\delta \geq 2$.

Il donne même l'expression suivante du minimum M soit :

$$(10) \quad M = 2 \int_{-\infty}^{+\infty} dy \int_y^{+\infty} \text{Sup} \{ G(y) - F(x), 0 \} dx +$$

$$+ 2 \int_{-\infty}^{+\infty} dx \int_x^{+\infty} \text{Sup} \{ F(x) - G(y), 0 \} dy .$$

Mais on peut donner une expression plus commode de ce minimum (qui est égal à Δ^2).

On peut arriver de deux façons à cette autre expression de Δ^2 . Faisons le calcul de Δ^2 dans le cas où $F(x)$, $G(y)$ sont partout continus et ne sont constants, l'un ou l'autre, dans aucun intervalle. Alors l'équation :

$$F(x) = G(y)$$

représente une courbe Γ qui est coupée en un point et un seul par toute parallèle à l'un des axes et sur laquelle y croît avec x et réciproquement. On a $H_1(x, y) = F(x)$ au-dessus (vers les y positifs) de Γ et $H_1(x, y) = G(y)$ au-dessous de Γ .

Considérons une suite de nombres $\dots x_{-1}, x_0, x_1, \dots x_k \dots$ croissent de $-\infty$ à $+\infty$; les nombres y_j correspondants, tels que :

$$F(x_j) = G(y_j)$$

croîtront aussi de $-\infty$ à $+\infty$. La courbe Γ sera contenue dans la suite S de rectangles :

$$x_j \leq x < x_{j+1} \quad ; \quad y_j \leq y < y_{j+1} .$$

Au-dessus de ces rectangles, $H_1(x, y) = F(x)$ et au-dessous $H_1(x, y) = G(y)$. Dès lors, l'intégrale :

$$\iint_B (x - y)^2 d_x d_y H_1(x, y)$$

étendue à la région B extérieure aux rectangles de S est nulle et on a

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - y)^2 d^2 H_1(x, y) = \iint_S (x - y)^2 d^2 H_1(x, y)$$

On a donc :

$$\begin{aligned} \mathcal{M}_{H_1}(X-Y)^2 &= \text{Lim} \sum_{i=-\infty}^{i=+\infty} \sum_{j=-\infty}^{j=+\infty} (x_i - y_j)^2 \Delta_{ij} H_1 \\ &= \text{Lim} \sum_i (x_i - y_i)^2 [H_1(x_{i+1}, y_{j+1}) - H_1(x_{i+1}, y_i) - H_1(x_i, y_{i+1}) + H_1(x_i, y_i)] . \end{aligned}$$

Or $H_1(x_{i+1}, y_{i+1}) = F(x_{i+1}) = G(y_{i+1})$

$H_1(x_i, y_i) = F(x_i) = G(y_i)$

$H_1(x_{i+1}, y_i) = G(y_i) , \quad H_1(x_i, y_{i+1}) = G(y_i) = F(x_i) .$

Donc le crochet est égal à :

$$F(x_{i+1}) - F(x_i)$$

et par suite :

$$\mathcal{M}_{H_1}(X - Y)^2 = \text{Lim} \sum_i (x_i - y_i)^2 [F(x_{i+1}) - F(x_i)] .$$

Puisque l'équation

(11) $F(x) = G(y)$

a une solution continue et croissante en y :

$y = \lambda(x) , \quad \text{on voit qu'on aura :}$

$$\Delta^2 = \mathcal{M}_{H_1}(X - Y)^2 = \int_{-\infty}^{+\infty} [x - \lambda(x)]^2 dF(x) .$$

En tirant, au contraire, de (11) $x = \mu(y)$

on pourrait aussi écrire :

$$\Delta^2 = \int_{-\infty}^{+\infty} [\mu(y) - y]^2 dG(y) .$$

De la même façon, la borne supérieure ρ de r est déterminée par la formule

$$\sigma \sigma' \rho = \int_{-\infty}^{+\infty} (x - a) [\lambda(x) - a] dF(x) = \int_{-\infty}^{+\infty} (y - a') [\mu(y) - a] dG(y) .$$

On peut obtenir encore plus facilement ces expressions en observant que si la fonction de répartition est $H_1(x, y)$, il y a presque certainement une relation fonctionnelle entre les deux nombres aléatoires X et Y , à savoir :

$Y = \lambda(X) .$

Dès lors :

$$\mathcal{M}_{H_1}(X - Y)^2 = \mathcal{M}_{H_1}[\lambda(X) - X]^2$$

$$(12) \quad (L, L')^2 = \int_{-\infty}^{+\infty} [\lambda(x) - x]^2 dF(x).$$

En posant $W = F(X) = G(Y)$, X et Y sont deux fonctions continues croissantes de W :

$$X = \ell(W) \quad ; \quad Y = m(W)$$

et W a une densité de probabilité constamment égale à 1 sur $0,1$. On a alors

$$(13) \quad (L, L')^2 = \int_0^1 [\ell(t) - m(t)]^2 dt.$$

Dans les notations précédentes,

$$\Delta^2 = (a' - a)^2 + (\sigma^2 - 2\sigma\sigma'\rho + \sigma'^2)$$

où

$$\rho = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} zt dh(z, t)$$

on aura :

$$\rho = \int_{-\infty}^{+\infty} zv(z) d\varphi(z)$$

si $t = v(z)$ est équivalent à : $\varphi(z) = \varphi(t)$.

Remarque :

La formule (10) de Dall'Aglio et les formules (12), (13) résolvent le problème posé (en note) par Salvemini dans son article cité plus haut.

Exemple

Supposons que X et Y soient telles que leurs variables "réduites" Z et T aient la même loi de probabilité. Alors, parmi les couples X, Y il y aura un couple pour lequel $Z = T$. Alors, il y a une relation fonctionnelle linéaire croissante entre X et Y pour $H = H_1$ et on aura encore comme à la formule (9).

$$(14) \quad \Delta = \sqrt{(a - a')^2 + (\sigma - \sigma')^2}.$$

On a ainsi un résultat particulièrement simple. La formule (14) convient, quand, par exemple, les marges ont une distribution dite normale. Mais elle convient aussi quand les marges vérifient toutes deux la lère loi de Laplace, ou sont toutes deux uniformes, etc...

Elle convient encore, plus généralement, quand X et Y ont une même loi réduite de probabilité comportant un écart quadratique moyen fini.

Généralisation :

Dall'Aglio a montré que si l'on prenait

$$([X], [Y])_H = \sqrt[s]{\mathcal{N}_H |X - Y|^s}$$

le minimum est, pour $s \geq 1$ atteint pour $H \equiv H_1$.

On voit comme plus haut, $p. =$ que si $F(x)$ et $G(y)$ sont continus et croissants, l'on a alors

$$(L, L')^s = \int_{-\infty}^{+\infty} |\lambda(x) - x|^s dF(x) \text{ ou } (L, L')^s = \int_{-\infty}^{+\infty} |\ell(t) - m(t)|^s dt.$$

UNE QUATRIÈME DÉFINITION DE LA DISTANCE DE DEUX LOIS DE PROBABILITÉ

La première des trois définitions de (L, L') dues à Lévy est très séduisante. Car elle explicite d'une façon très simple et très naturelle la notion intuitive que nous pouvons avoir de la distance (L, L') de deux lois de probabilité L, L' .

Son plus grand inconvénient, c'est que, variant avec la définition adoptée comme point de départ de la distance globale de deux nombres aléatoires, elle exigera à chaque fois la détermination préalable du minimum de cette distance quand varie la corrélation de ces deux nombres.

Une quatrième définition :

Nous allons donc proposer une quatrième définition de (L, L') , pour esquiver cette difficulté. Nous ne considérons donc pas notre définition comme meilleure mais seulement comme d'application plus immédiate. Nous verrons d'ailleurs ensuite que les deux définitions sont équivalentes dans un cas étendu. Elles sont probablement équivalentes dans des cas plus généraux encore, sinon dans tous les cas.

Nouvelle définition de (L, L') :

Prenons encore comme point de départ, l'une des définitions admissibles, de la distance "globale", $([X], [Y])_H$, de deux nombres aléatoires X, Y - déterminés simultanément dans chaque épreuve d'une catégorie déterminée, correspondant à la fonction de répartition $H(x, y)$ du couple X, Y .

Ceci étant, nous prendrons comme nouvelle définition de la distance de deux lois L, L' la quantité :

$$(15) \quad (L, L') = ([X], [Y])_{H_1}$$

où X, Y obéissent aux deux lois L, L' . Ainsi nous appelons distance de 2 lois de probabilité L, L' , la valeur que prend une distance "globale" de

deux nombres aléatoires X, Y , obéissant séparément aux lois L, L' , quand la fonction $H(x, y)$ de répartition du couple X, Y atteint sa valeur maximum $H_1(x, y)$ précisée par la formule (2).

Il faut maintenant prouver que cette quantité (L, L') est bien "une distance"; c'est-à-dire satisfait aux trois conditions bien connues, rappelées pages 1, 2, sous les n° 1, 2°, 3°. On aura ainsi une justification partielle a posteriori de la définition que nous venons de poser dogmatiquement.

Par hypothèse, pour toute fonction de corrélation $H(x, y)$ admettant pour marges les fonctions de répartition $F(x), G(y)$ des lois L, L' , la distance globale de $[X]$ et $[Y]$ vérifie les conditions :

$$1^\circ ([X], [Y])_H = ([Y], [X])_H \geq 0$$

2° $([X], [Y])_H$ est positif si X et Y sont "distincts" et inversement. (On considère, en général, comme non distincts deux nombres aléatoires qui sont "presque certainement" égaux, c'est-à-dire tels que :

$$\{\text{Prob.}(X=Y)\} = 1.$$

C'est aussi la condition nécessaire et suffisante pour que $F(x) \equiv G(x)$. Dès lors en appliquant au cas où $H \equiv H_1$

$$1^\circ (L, L') = (L', L) \geq 0$$

$$2^\circ (L, L') = 0 \text{ si } L \equiv L' \text{ et inversement.}$$

En ce qui concerne la condition 3°, nous devons faire intervenir trois lois de probabilité L, L', L'' , correspondant à trois fonctions de répartition $F(x), G(y), K(z)$. Introduisons une corrélation particulière entre trois nombres aléatoires X, Y, Z , obéissant respectivement aux trois lois L, L', L'' . Nous choisissons à cet effet la corrélation définie par la fonction de répartition :

$$E(x, y, z) = \text{Min}[F(x), G(y), H(z)].$$

On voit bien que si le système X, Y, Z a pour fonction de répartition $E(x, y, z)$, X, Y, Z auront bien $F(x), G(y), H(z)$ comme fonctions de répartition. Car, par exemple,

$$\begin{aligned} \text{Prob}(X < x) &= \text{Prob}(X < x, Y < +\infty, Z < +\infty) \\ &= E(x, +\infty, +\infty) = \text{Min}(F(x), +1, +1) = F(x). \end{aligned}$$

On aura de plus :

$$E(x, y, +\infty) = \text{Min}[F(x), G(y), +1] = H_1(x, y)$$

et de même les fonctions maximées de répartition de (X, Z) et de (Y, Z) seront

$$H_1'(x, z) = E(x, +\infty, z)$$

$$H_1''(y, z) = E(+\infty, y, z).$$

Ceci étant, on a :

$$(16) \quad ([X], [Y])_{H_1} = ([X], [Y])_{E(x, y, +\infty)} = ([X], [Y])_{E(x, y, z)}$$

Or, par hypothèse, pour n'importe quelle fonction de répartition à trois variables, $U(x, y, z)$, ayant les marges données F, G, K on a :

$$([X], [Y])_U \leq ([X], [Z])_U + ([T], [Z])_U.$$

Donc, en prenant en particulier E pour U et utilisant les deux égalités analogues à (4), on a :

$$([X], [Y])_{H_1} \leq ([X], [Z])_{H_1'} + ([Y], [Z])_{H_1''}$$

c'est-à-dire :

$$(L, L') \leq (L, L'') + (L', L'').$$

Equivalence :

Il résulte des démonstrations de Bass et de Dall'Aglio que dans le cas où l'on a pris :

$$([X], [Y])_H = \sqrt{\sigma_H^2 (X-Y)^2}$$

et sous les restrictions très générales et qui, d'ailleurs, ne sont peut-être pas nécessaires, la valeur de (L, L') reste la même qu'on emploie la première définition de Paul Lévy ou la nôtre.

Calcul général de (L, L') :

Considérons le cas où $F(x)$ et $G(y)$ sont des fonctions partout continues et croissantes. Alors la courbe Γ :

$$F(x) = G(y)$$

va séparer la région R où $H_1 = F(\leq G)$ de celle, Q_1 , où $H_1 = G(\leq F)$. Si l'on pose :

$$F(x) = G(y) = t$$

on voit comme plus haut que x et y seront des fonctions continues et croissantes de t , soit :

$$x = \ell(t) \qquad y = m(t).$$

Quand la fonction de répartition est : $H_1(x, y) = \text{Min} [F(x), G(y)]$, la différence seconde de H_1 sera nulle à l'intérieur de R et à l'intérieur de Q_1 , c'est-à-dire que le point (X, Y) est presque certainement sur Γ .

Autrement dit, il y a dans ce cas une relation fonctionnelle presque certaine entre X et Y , à savoir :

$$F(x) = G(y).$$

Le nombre aléatoire T égal à la valeur commune de F et G sur Γ reste entre 0 et 1 avec une densité de probabilité constante et égale à 1.

Il en résulte l'expression de (L, L') :

$$(L, L') = ([\ell(T)], [m(T)]).$$

Exemple :

Par exemple, si l'on part de l'expression :

$$([X], [Y])_H = \sqrt[\alpha]{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |x - y|^\alpha d_x d_y H(x, y)} = \sqrt[\alpha]{\mathcal{N}_H |X - Y|^\alpha}$$

avec

$$\alpha \geq 1,$$

on aura :

$$(L, L')^\alpha = \mathcal{N} |l(T) - m(T)|^\alpha, \quad \text{d'où :}$$

$$(L, L')^\alpha = \int_0^1 |l(t) - m(t)|^\alpha dt.$$

On peut aussi résoudre :

$$F(x) = G(y)$$

en $y = \lambda(x)$,

où $\lambda(x)$ sera, comme F et G , une fonction continue croissante. On pourra alors écrire :

$$(L, L')^\alpha = \int_{-\infty}^{+\infty} |x - \lambda(x)|^\alpha dF(x)$$

L'INDICE DE DISSEMBLANCE DE GINI

Pour deux suites également nombreuses de nombres non décroissants;

$$a_1 \leq a_2 \leq \dots \leq a_n, \quad b_1 \leq b_2 \leq \dots \leq b_n,$$

cet indice est posé égal à

$$I_\alpha = \sqrt[\alpha]{\frac{1}{n} \sum_{h=1}^n |a_h - b_h|^\alpha} \quad \begin{array}{l} \alpha = 1: \text{indice simple} \\ \alpha = 2: \text{indice quadratique} \end{array}$$

Mais on peut prendre n'importe quel $\alpha \geq 1$.

Les a n'étant pas nécessairement distincts, leurs valeurs distinctes sont répétées, r_1 fois, ... r_n fois. Elles peuvent être considérées comme les valeurs prises par une grandeur statistique X dans n épreuves. Les valeurs distinctes de X : $x_1 < x_2 < \dots$ auront les fréquences :

$$f_1 = \frac{r_1}{n}, \quad f_2 = \frac{r_2}{n}, \dots \quad \text{La suite } \begin{cases} x_1, x_2, \dots \\ f_1, f_2, \dots \end{cases}$$

définit ainsi une loi de fréquence L . De même, la suite des valeurs dis-

tinctes y_1, y_2, \dots prises par les b_j avec les fréquences: $\varphi_1 = \frac{s_1}{n}, \varphi_2 = \frac{s_2}{n}, \dots$ définit une loi de fréquence L' .

En permuttant les b_h , on obtient une suite b'_1, b'_2, \dots qui n'est plus nécessairement non décroissante. Si chaque épreuve associait à un a_h, b_h , on aurait une loi H de fréquence du couple X, Y . En associant b_h à a_h , on a une loi particulière, H' , de fréquence de ce couple. Nous allons montrer que $H' = H_1$, de sorte que l'indice de dissemblance de Gini deviendra identique à ce à quoi se réduit notre distance de L et de L' dans le cas particulier où L, L' sont les lois de deux grandeurs statistiques (ayant chacune un nombre fini de valeurs).

En effet, supposons d'abord :

$$a_1 < x \leq a_n \quad , \quad b_1 < y \leq b_n$$

alors il y aura un h et un j tels que

$$a_h < x \leq a_{h+1} \quad , \quad b_j < y \leq b_{j+1} .$$

On aura

$$H'(x, y) = \text{Prob}(X < x, Y < y) = \text{Prob}(X < a_{h+1}, Y < b_{j+1})$$

et puisque H' correspond au cas où b_j est associé à a_j , alors si $j < h$

$$H'(x, y) = \text{Prob}(Y < b_{j+1}) = \text{Prob}(Y < y) = G(y)$$

et $F(x) = \text{Prob}(X < a_{h+1}) \geq \text{Prob}(Y < b_{j+1}) = \text{Prob}(Y < y)$.

Donc : si $j \leq h$

$$H'(x, y) = G(y) \leq F(x)$$

$$H'(x, y) = \text{Min}[F(x), G(y)] = H_1(x, y) .$$

En répétant le même raisonnement pour $h \leq j$, on voit que dans les deux cas :

$$(1') \quad H'(x, y) = H_1(x, y) .$$

Dans le cas où $x \leq a_1$, on aurait évidemment :

$$H'(x, y) = F(x) = 0 \leq G(y)$$

et par suite (1') subsisterait; de même si $y \leq b$, raisonnement analogue pour $x > a_n$ ou $y > b_n$. Ainsi dans tous les cas

$$H'(x, y) \equiv H_1(x, y) . \quad (1)$$

L'indice simple ($\alpha = 1$) ou quadratique ($\alpha = 2$) de dissemblance, habilement choisi par Gini, se trouve ainsi être en même temps une "distance", de deux lois de fréquence. Plus précisément, il rentre dans

(1) Nous ne faisons ici que reprendre un peu plus en détail le raisonnement de Salvemini qui, du reste, ne s'occupe pas de la notion de distance.

la catégorie plus générale de la quatrième définition de la distance de deux lois de probabilité absolument quelconques (discrètes ou continues ou intermédiaires) précisée p. 11. Il présente, d'autre part, cet intérêt d'offrir un mode de calcul plus simple que celui qui résulterait directement de la formule :

$$I_\alpha = \sqrt[\alpha]{\int \int \pi \pi_h |X - Y|^\alpha}$$

En effet, supposons d'abord :

$$x < y \leq a$$

alors H correspond au cas où p est associé à x, alors si j < k

$$p < y \leq p_h$$

On voit

$$H(x, y) = \text{Prob}(X < x, Y < y) = \text{Prob}(X < x, Y < p_h)$$

et puisque H correspond au cas où p est associé à x, alors si j < k

$$H(x, y) = \text{Prob}(Y < p_h) = \text{Prob}(Y < y) = G(y)$$

$$F(x) = \text{Prob}(X < a) \geq \text{Prob}(Y < p_h) = \text{Prob}(Y < y)$$

Donc : si j < k

$$H(x, y) = G(y) \leq F(x)$$

$$H(x, y) = \min[F(x), G(y)] = H(x, y)$$

En réitérant le même raisonnement pour h < k, on voit que dans les deux cas :

$$H(x, y) = H(x, y)$$

Dans le cas où x < a, on a aussi évidemment :

$$H(x, y) = F(x) \leq G(y)$$

et par suite (1') subsiste : de même si y < b, raisonnement analogue pour x > a ou y > b. Ainsi dans tous les cas

$$(1) \quad H(x, y) = \min[F(x), G(y)]$$

L'indice simple (α = 1) ou quadratique (α = 2) de dissimilitude habituellement appelé par Gini, se trouve ainsi être en même temps une "distance" de deux lois de fréquence. Plus précisément, il tient dans

(1) Nous ne faisons ici que répéter ce que nous avons dit précédemment de la distance de Gini, ce n'est pas de la distance de Gini.