I. ETUDES SCIENTIFIQUES D'ORDRE STATISTIQUE.

COMPARAISON DES DIVERSES MESURES DE LA DISPERSION

par Maurice Fréchet.

INTRODUCTION.

Mesure de la dispersion. Sous le terme vague mais intuitif de dispersion d'une variable aléatoire X, on entend la dissémination des valeurs possibles que peut prendre cette valeur dans les différentes épreuves, étant entendu cependant qu'une valeur (ou un intervalle de valeur) possible interviendra d'autant moins dans l'évaluation de la dispersion que sa probabilité sera plus petite.

Pour chiffrer cette dispersion, une convention est nécessaire; la mesure [1]) de la dispersion comporte donc un certain arbitraire. Mais cette convention devra se conformer à notre sentiment intuitif de la dispersion, ce qui réduit considérablement cet arbitraire.

En fait, un petit nombre seulement de définitions distinctes de la mesure de la dispersion ont été proposées. Nous allons ici délimiter un intervalle où doit rester une mesure de cette dispersion, effectuée conformément à une de ces définitions, quand on connait une mesure de la même dispersion, calculée suivant une autre de ces définitions. Nous arriverons même, généralement, à déterminer dans chaque cas l'intervalle le plus strict possible.

Rappel de définitions. Parmi les mesures les plus utilisées de la dispersion figurent celles qui vont être définies ci-dessous.

Fixons d'abord les notations.

Désignons par $\mathcal{M} X$ ou $\overline{X}$, la valeur moyenne (ou espérance mathématique) de X. Appelons fonction de répartition de X, la fonction

$$F(x) = \text{Prob}\left\{ X < x \right\} \text{ [2])}$$

Appelons valeur probable de X, toute valeur p telle que

$$F(x) \leq \tfrac{1}{2} \quad \text{si} \quad x < p$$
$$F(x) \geq \tfrac{1}{2} \quad \text{si} \quad x > p$$

On définirait de même les quartiles de X, q et q_1, obtenus en remplaçant $\frac{1}{2}$ par $\frac{1}{4}$ et par $\frac{3}{4}$.

---

[1]) En fait, il s'agit plutôt de repérage de la dispersion que de sa mesure.

[2]) Nous avons pris l'habitude de désigner dans nos différentes publications les variables aléatoires par de grandes lettres, les nombres certains par de petites lettres. Cet usage, qui a été adopté par divers auteurs, mériterait peut-être d'être généralisé.

Les mesures les plus courantes de la dispersion sont, dans l'ordre de simplicité:

l'écart probable de X: $E_x = \frac{q_1 - q}{2}$ [1]),

l'écart moyen de X: $\Theta_x = \mathcal{M} \mid X - p \mid = \int\limits_{-\infty}^{+\infty} \mid x - p \mid d\,F(x)$,

l'écart quadratique moyen de X: $\sigma_x = \sqrt{\mathcal{M}(X - \overline{X})^2} =$

$$= \sqrt{\int\limits_{-\infty}^{+\infty} (x - \overline{X})^2\, d\,F(x)}.$$

Nous commencerons donc par la comparaison des valeurs de ces trois écarts, ce qui constituera la Première Partie.

Nous continuerons en comparant à ces trois écarts une mesure peu usitée qui pourrait consister dans une quantité proportionnelle à l'inverse de la borne supérieure de la „densité moyenne" de la probabilité, c'est-à-dire du rapport

$$\frac{F(x_2) - F(x_1)}{x_2 - x_1} = \frac{\text{Prob}\,\{ x_1 \leq X < x_2 \}}{x_2 - x_1} \qquad (x_1 < x_2)$$

En appelant Δ la borne supérieure (finie ou infinie) de ce rapport quand x_1 et x_2 varient, nous poserons (pour simplifier certaines formules grâce à l'introduction du facteur $\frac{1}{4}$)

$$e_x = \frac{1}{4\,\Delta}.$$

Cette quantité répond jusqu'à un certain point à notre sens intuitif de la dispersion. Elle en constitue une mesure assez peu satisfaisante, mais notre comparaison de e_x avec les autres mesures a, en réalité, surtout pour objet d'établir des limitations, utiles dans certaines questions, de σ, Θ et E quand on connait Δ.

Enfin il ne sera pas sans intérêt de comparer à ces écarts, une fonction introduite par M. Paul Lévy sous le nom de fonction de dispersion. Elle a l'inconvénient de remplacer le nombre, qu'est un écart, par l'entité plus complexe qu'est une fonction, mais elle serre certainement de plus près que ces écarts, notre notion intuitive de dispersion. Nous présenterons d'ailleurs cette notion nouvelle sous une forme légèrement différente et avec un nom différent. Nous appellerons écart maximum à une probabilité ε près, ou, plus brièvement, écart à ε près, la borne inférieure des valeurs de l telles qu'il existe pour chacune d'elles un nombre x pour lequel

$$\left[\text{Prob}\,\{ \mid X - x \mid \geq l \} \right] \leq \varepsilon.$$

Et nous représenterons ce nouvel écart par $L_x(\varepsilon)$.
Il est d'ailleurs clair que c'est une fonction non croissante de ε et que $L_x(1) = 0$.

[1]) Exceptionnellement q_1 peut être indéterminé sur un segment, de même pour q. Dans un tel cas, l'écart probable reste indéterminé entre „le plus grand" et „le plus petit" écart probable.

On voit que si X est une variable bornée, c'est-à-dire si elle reste dans toutes les épreuves possibles entre deux nombres fixes $c - \lambda, c + \lambda$, on aura

$$\left[\text{Prob} \left\{ | X - c | \geq l \right\} \right] = 0 \leq \varepsilon$$

pour $l > \lambda$. Donc $L_x(\varepsilon) \leq \lambda$ quel que soit ε et, par suite $L_x(\varepsilon)$ est une fonction bornée de ε. Elle est en fait bornée par l'écart maximum des bornes de X avec leur milieu. Une généralisation théorique consiste à observer que ce résultat subsiste quand X est „presque sûrement" bornée, c'est-à-dire quand il y a une probabilité nulle que X dépasse certaines bornes.

On retrouve cette condition suffisante comme condition nécessaire pour que $L_x(\varepsilon)$ soit borné quand ε varie. C'est évident si l'on observe que $L_x(\varepsilon) \leq L_x(0)$. Si $L_x(\varepsilon)$ est borné pour $\varepsilon \geq 0$, $L_x(0)$ est fini et

$$\text{Prob} \left\{ | X - x | \geq l \right\} \leq 0$$

pour au moins un intervalle $x - l$, $x + l$ de longueur finie $2\, l \geq 2\, L_x(0)$. Et par suite, il y a une probabilité nulle que $X < x - l$ ou $X > x + l$: X est presque sûrement borné.

Il est même clair que si x', x'' sont les bornes supérieure et inférieure des nombres x_1, x_2, tels que

$$\text{Prob} \left\{ X < x_1 \right\} = 0 = \text{Prob} \left\{ X > x_2 \right\},$$

on aura pour la borne supérieure $L_x(0)$ de $L_x(\varepsilon)$,

$$L_x(0) = \frac{x'' - x'}{2}.$$

R e m a r q u e. Avant de comparer les grandeurs de ces différents écarts, il nous paraît indispensable de signaler qu'à côté d'une grande qualité de l'écart probable, celle de se prêter à un calcul rapide, souvent quasi instantané, cet écart présente un grand défaut: celui de pouvoir être nul sans qu'au sens intuitif la dispersion soit vraiment nulle. Si par exemple X prend les valeurs $-a$, 0, $+a$ avec les probabilités respectives $\frac{1}{8}$, $\frac{3}{4}$, $\frac{1}{8}$, son écart probable est nul et pourtant, dans des épreuves de probabilité totale $\frac{1}{4}$ non négligeable, il peut présenter des valeurs $-a$, $+a$ dont la différence est, avec le nombre arbitraire a, aussi grande que l'on veut.

Au contraire, que Θ_x ou que σ_x soit nul on sait qu'alors X reste égal à un nombre fixe sauf dans des cas de probabilité nulle: la dispersion de X est nulle aussi bien au sens intuitif que lorsqu'on la mesure par Θ_x ou σ_x. D'autre part, en vertu de la formule qui vient d'être établie, on voit que si $L_x(\varepsilon)$ est nul quel que soit ε, X est encore égal „presque sûrement" à un nombre fixe. (Par contre, bien entendu, quand X est un nombre certain — ou „presque sûrement" certain —, les trois écarts Θ_x, σ_x, E_x sont nuls et $L_x(\varepsilon)$ est nul quel soit ε).

PREMIÈRE PARTIE.

C o m p a r a i s o n d e s m e s u r e s c l a s s i q u e s d e l a d i s p e r s i o n. Nous nous proposons ici d'établir les inégalités les plus strictes possibles entre les mesures σ_x, Θ_x, E_x de la dispersion de X.

On sait déjà que l'on a toujours $\sigma_x \geq \Theta_x$, de sorte que

$$(1) \qquad 0 \leq \frac{\Theta_x}{\sigma_x} \leq 1.$$

Ainsi: $\frac{\Theta_x}{\sigma_x}$ reste compris entre deux bornes absolues, c'est-à-dire indépendantes de la variable aléatoire X considérée. Ces bornes ne peuvent être améliorées. La borne supérieure d'abord puisqu'elle est atteinte quand X est, par exemple, une variable aléatoire X_0 qui ne peut prendre que deux valeurs également probables. Dans ce cas, elle atteint son maximum. Si l'on assujettissait X à avoir partout une densité continue de probabilité, cette valeur 1 ne serait pas atteinte, mais resterait une borne supérieure, car il serait facile d'obtenir une loi de répartition, à densité partout continue, de façon que cette loi soit voisine de celle de X_0.

Il est de même pour la borne inférieure de $\frac{\Theta_x}{\sigma_x}$. On sait en effet que σ_x peut être infini alors que Θ_x est fini. C'est, par exemple, ce qui aurait lieu si X avait une densité de probabilité $\Phi(x)$ telle que

$$(2) \qquad \begin{cases} \Phi(x) = A, & \text{pour } |x| \leq 1 \\ \Phi(x) = \dfrac{A}{|x|^{\frac{5}{2}}}, & \text{pour } |x| > 1 \end{cases}$$

A étant choisi de sorte que $\int_{-\infty}^{+\infty} \Phi(x)\, d\, x = 1$.

En résumé, on a

$$(3) \qquad \boxed{0 \leq \frac{\Theta_x}{\sigma_x} \leq 1}$$

et les bornes 0 et 1 ne peuvent être améliorées, que X soit une variable aléatoire quelconque, ou que X ait partout une densité de probabilité. C'est là, en somme, un résultat déjà connu bien qu'il n'ait peut-être pas encore été énoncé explicitement sous cette forme.

Comparons maintenant Θ_x et E_x. On sait aussi que $\frac{E_x}{\Theta_x}$ peut être nul puisque Θ_x peut être infini, sans que E_x le soit. C'est, par exemple, ce qui a lieu si

$$(4) \qquad \begin{cases} \Phi(x) = B, & \text{pour } x \leq 1 \\ \Phi(x) = \dfrac{B}{|x|^{\frac{3}{2}}}, & \text{pour } x \geq 1 \end{cases}$$

B étant choisi de sorte que $\int_{-\infty}^{+\infty} \Phi(x)\, d\,(x) = 1$.

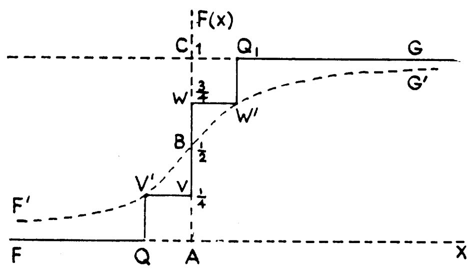
Passons à la borne supérieure. L'expression

$$\Theta_x = \int_{-\infty}^{+\infty} |x - p|\; d\, F(x)$$

où p est une valeur probable peut être interprétée géométriquement comme la somme de deux aires qu'on peut désigner dans la figure 1 par $B\ C\ G\ G'\ B$ et $B\ A\ F\ F'\ B$. Il en résulte que Θ_x est au moins égal à la somme des aires rectangulaires $A\ V\ V'\ Q$, $C\ W\ W'\ Q_1$, somme qui est égale à $\frac{1}{4}(C\ Q_1 + A\ Q) = \frac{1}{4}(2\ E_x)$.

D'où
$$\Theta_x \geq \frac{E_x}{2} \text{ [1]}$$

FIG. 1



Il est clair que l'égalité aura lieu quand la courbe $y = F(x)$ sera remplacée par la ligne polygonale discontinue formée par $F\ Q$, $V'\ V$, $W\ W'$, $Q_1\ G$, c'est-à-dire quand X se réduit à une variable aléatoire T ne possédant que trois valeurs, qui, rangées par ordre de grandeur, ont les probabilités respectives $\frac{1}{4}$, $\frac{1}{2}$ et $\frac{1}{4}$. Dans ce cas, le rapport $\frac{E_x}{\Theta_x}$ atteint son maximum, 2, sur l'ensemble de toutes les variables aléatoires.

Si on veut se restreindre aux variables aléatoires qui ont partout une densité continue de probabilité, la valeur 2 sera encore une borne supérieure de $\frac{E_x}{\Theta_x}$ (sans en être le maximum atteint) en remplaçant la ligne discontinue L qui représente la fonction de répartition de T par une courbe douée partout d'une tangente continue suffisamment voisine de L pour que Θ_T et E_T soient aussi voisins de Θ_x et E_x qu'on le veut, ce qui est possible.

En définitive, on a

$$\boxed{0 \leq \frac{E_x}{\Theta_x} \leq 2} \tag{5}$$

et ces inégalités ne peuvent être améliorées, soit qu'on les applique à toutes les variables aléatoires X, soit seulement à celles qui ont en tout point une densité de probabilité.

Il résulte de (1) et (5) qu'on a toujours

$$\sigma_x \geq \Theta_x \geq \frac{E_x}{2} \geq 0 \tag{6}$$

et, par suite, que:

$$0 \leq \frac{E_x}{\sigma_x} \leq 2 \tag{7}$$

mais il ne s'en suit pas immédiatement que ces inégalités ne puissent être

[1]) Dans le cas où l'écart probable n'a pas une valeur unique, l'inégalité a lieu pour toutes ses valeurs; elle sera la plus stricte possible quand on prendra la plus grande.

améliorées. Cependant σ_X étant $\geq \Theta_X$ la borne inférieure de $\frac{E_X}{\sigma_X}$ est au plus égale à celle de $\frac{E_X}{\Theta_X}$ qui est nulle. Par contre, en ce qui concerne la borne supérieure, il faut raisonner directement. Il faut comparer à E_X, $\sigma_X = \sigma_X(x_0)$, où x_0 est la valeur moyenne de X. Si x_0 est entre les deux quartiles, le raisonnement fait pour $\frac{E_X}{\Theta_X}$ peut être répété en remplaçant p par x_0.

On a alors (fig. 1 où l'abscisse commune à C et A serait non plus p mais x_0)

$$(\sigma_X)^2 \geq \tfrac{1}{4}(C\,Q_1)^2 + \tfrac{1}{4}(A\,Q)^2 \geq$$
$$\tfrac{1}{8}\left\{(C\,Q_1 + A\,Q)^2 + (C\,Q_1 - A\,Q)^2\right\} \geq \tfrac{1}{8}(C\,Q_1 + A\,Q)^2 =$$
$$\tfrac{1}{8}(2\,E_X)^2 = \tfrac{1}{2}(E_X)^2.$$

D'où

$$\frac{\sigma_X}{E_X} \geq \frac{1}{\sqrt{2}}\ (> \tfrac{1}{2}) \qquad \frac{E_X}{\sigma_X} \leq \sqrt{2}$$

Si x_0 est en dehors des quartiles, par exemple à gauche, on a (fig. 2):

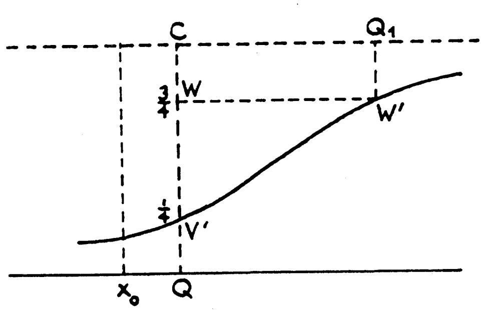
$$(\sigma_X{}^2) = \int_{-\infty}^{x_0} (x_0 - x)^2\, d\,F + \int_{x_0}^{+\infty} (x - x_0)^2\, d\,F = \sigma'^2 + \sigma''^2.$$

Or

$$\sigma''^2 \geq \int_{q_1}^{+\infty} (x - x_0)^2\, d\,F \geq (q_1 - q)^2 \int_{q_1}^{+\infty} d\,F = (2\,E_X)^2\,\tfrac{1}{4} = (E_X)^2 \geq \frac{(E_X)^2}{2}$$

en appelant q, q_1 les abscisses des quartiles. Ainsi, on a $(\sigma_X)^2 \geq \frac{(E_X)^2}{2}$ quelle que soit la position de la moyenne x_0 de X et par suite [1] $\frac{E_X}{\sigma_X} \leq \sqrt{2}$ (< 2).

FIG. 2



Cette borne supérieure (qui est plus petite que celle donnée par (7)) est, cette fois, atteinte quand, par exemple, X est une variable aléatoire ne prenant que les valeurs -1, 0, 1 avec les probabilités $\frac{1}{4}$, $\frac{1}{2}$ et $\frac{1}{4}$. Et cette borne supérieure resterait une borne supérieure (sans être atteinte), si l'on ne considérait que les variables X ayant partout une densité continue de probabilité. Ainsi on a

[1] Même observation qu'en note de la p. 5 pour le cas où E_X n'a pas une valeur unique.

$$(8) \qquad \boxed{0 \leq \frac{E_X}{\sigma_X} \leq \sqrt{2}}$$

et ces inégalités ne peuvent être améliorées quand on les étend, soit à toutes les variables aléatoires, soit seulement à toutes celles qui ont une densité continue de probabilité.

Comparaison de σ_X, Θ_X, E_X avec l'écart à ε près $L_X(\varepsilon)$. D'après la définition de l'écart maximum à une probabilité $\leq \varepsilon$ près, $L_X(\varepsilon)$ et en appliquant l'inégalité de Bienaymé généralisée

$$\left[\text{Prob}\ \{\, | X - x | \geq l \,\}\right] \leq \frac{\mathfrak{M}\, | X - x |^n}{l^n}$$

si l'on pose $\varepsilon = \dfrac{\mathfrak{M}\, | X - x |^n}{l^n}$, on aura quel que soit x

$$L_X(\varepsilon) \leq l = \sqrt[n]{\frac{\mathfrak{M}\, | X - x |^n}{\varepsilon}}.$$

Dès lors en prenant pour x la valeur probable pour $n = 1$, puis la valeur moyenne de X, pour $n = 2$, on a

$$(9) \qquad \boxed{L_X(\varepsilon) \leq \frac{\Theta_X}{\varepsilon}}$$

$$(10) \qquad \boxed{L_X(\varepsilon) \leq \frac{\sigma_X}{\sqrt{\varepsilon}}}$$

D'autre part, si a_1 et a_2 comprennent à leur intérieur tous les quartiles (de sorte que $a_2 - a_1 > 2\, E_X$), on aura

$$\text{Prob}\ \left\{\, \left| X - \frac{a_1 + a_2}{2} \right| \geq \frac{a_2 - a_1}{2} \,\right\} < \tfrac{1}{2}$$

d'où $L_X(\tfrac{1}{2}) \leq \dfrac{a_2 - a_1}{2}$, et par suite $L_X(\tfrac{1}{2}) \leq E_X$, donc aussi

$$(11) \qquad \boxed{L_X(\varepsilon) \leq E_X \ \text{ pour } \varepsilon \geq \tfrac{1}{2}\ ^{1)}}$$

On vient de prouver que

$$(11') \qquad \frac{\varepsilon\, L_X(\varepsilon)}{\Theta_X}, \quad \frac{\sqrt{\varepsilon}\, L_X(\varepsilon)}{\sigma_X}, \quad \frac{L_X(\frac{1}{2})}{E_X}$$

sont bornés inférieurement et supérieurement chacun par deux nombres indépendants de ε et de la variable aléatoire X. Les bornes inférieures sont même exactement égales à zéro. En ce qui concerne les deux premiers rapports, il suffit de montrer que $\varepsilon L_X(\varepsilon)$ peut être fini quand Θ_X est infini (et par suite

---

[1]) On observera que dans le cas où il y a plusieurs quartiles, on peut pendre dans (11) pour E_X une quelconque des valeurs de l'écart probable.

aussi σ_{x}). Or si l'on prend une variable aléatoire X dont l'écart moyen est infini, et si ε est un nombre positif arbitraire, il suffit de prendre x_1 et x_2 assez éloignés vers $-\infty$ et ∞ pour que Prob $\left\{x_1 < X < x_2\right\}$ soit $\geq 1-\varepsilon$. Alors $L_{\text{x}}(\varepsilon)$ sera fini et $\leq x_2 - x_1$, donc $\varepsilon\, L_{\text{x}}(\varepsilon)$ sera aussi fini. Pour le troisième rapport, prenons $X = 0$ ou 1 avec les probabilités respectives $\frac{3}{4}$ et $\frac{1}{4}$. On aura $2\,E_{\text{x}} = 1$ et $L_{\text{x}}(\frac{1}{2}) = 0$. Donc la valeur 0 peut être aussi atteinte par le troisième rapport.

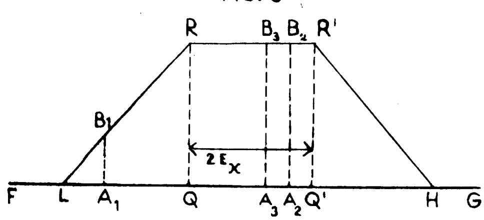
Passons aux bornes supérieures.

On ne peut améliorer l'inégalité

$$\frac{L_{\text{x}}(\frac{1}{2})}{E_{\text{x}}} \leq 1,$$

car la borne supérieure est atteinte pour la variable aléatoire U dont la courbe de densité de probabilité est représentée dans la fig. 3, par la ligne polygonale $F\ L\ R\ R'\ H\ G$ où les aires triangulaires $L\ Q\ R$, $R'\ Q'\ H$ sont égales chacune à la moitié de l'aire rectangulaire $Q\ Q'\ R'\ R$, de sorte que Q et Q' représentent les quartiles.

FIG. 3



En effet, si

(12) $$\text{Prob}\left\{x_1 < X < x_1 + 2\,l = x_2\right\} \geq \tfrac{1}{2},$$

ou, par exemple: aire $A_1\ B_1\ R\ B_2\ A_2 \geq \frac{1}{2} = A_1\,B_1\,RB_3\,A_3 =$ aire $Q\ R\,R'\ Q'$, il est visible que $A_1\ Q$ devra être plus grand que $A_3\ Q'$, c'est-à-dire que le minimum de $2\,l$ dans (12) quand x_1 varie est obtenu quand $2\,l$ est égal à $Q\,Q'$. Donc

$$2\ L_U(\tfrac{1}{2}) = 2\ E_U, \text{ ou } \frac{L_U(\frac{1}{2})}{E_U} = 1.$$

Peut-on améliorer (9) et (10)?

Les deux premiers rapports de (11') prennent des valeurs $> \frac{1}{2}$ dans quelques cas simples [1]; leurs bornes supérieures quand X et ε sont arbitraires sont donc certainement ≤ 1 et $\geq \frac{1}{2}$.

DEUXIÈME PARTIE.

C o m p a r a i s o n d e s m e s u r e s d e l a d i s p e r s i o n e t d u m a x i m u m d e d e n s i t é d e p r o b a b i l i t é. Nous nous proposons de montrer que le fait

[1]) Quand X ne peut prendre que deux valeurs, chacune avec la probabilité $\frac{1}{2}$, ces premiers membres ont respectivement $\frac{1}{2}$ et $\frac{1}{\sqrt{2}}$ pour maximum quand ε varie.

pour une variable aléatoire X d'avoir une densité de probabilité bornée permet de déterminer une borne inférieure positive de la dispersion de X, qu'on mesure celle-ci par son écart quadratique moyen σ_x, son écart moyen Θ_x ou par son écart probable E_x. C'est un résultat intéressant en soi et qui en tout cas nous a été utile dans une autre étude publiée ailleurs [1]).

Plus précisément, nous allons démontrer que, pour toute variable aléatoire X admettant une „dominante", on a

$$\Delta\, \Theta_x \geq \frac{1}{4} \tag{13}$$

$$\Delta\, \sigma_x \geq \frac{1}{2\sqrt{3}} \tag{14}$$

$$\Delta\, E_x \geq \frac{1}{4} \tag{15}$$

en appelant Δ le maximum de la densité de probabilité.

On peut d'ailleurs abandonner toute limitation du choix de X en généralisant la signification de Δ.

On peut définir la densité moyenne de probabilité sur le segment x_1, x_2 comme étant le rapport

$$\frac{\text{Prob}\left\{x_1 \leq X \leq x_2\right\}}{x_2 - x_1} \quad \text{avec } x_1 < x_2.$$

On appellera Δ la borne supérieure (finie ou infinie) de cette densité moyenne quand on fait varier arbitrairement x_1, x_2, avec $x_1 < x_2$. On aura donc Prob $\left\{x_1 \leq X \leq x_2\right\} \leq \Delta\,(x_2 - x_1)$. En particulier, ceci montre que Δ ne peut être nul. Car pour x_1 et x_2 assez grands en valeur absolue, mais déterminés, le premier membre peut être rendu aussi voisin de 1 que l'on veut.

Ceci étant, on va prouver que

1°. pour toute variable aléatoire X, on a

$$\boxed{0 \leq \frac{1}{\Delta\, \Theta_x} \leq 4} \tag{13bis}$$

$$\boxed{0 \leq \frac{1}{\Delta\, \sigma_x} \leq 2\sqrt{3}} \tag{14bis}$$

$$\boxed{0 \leq \frac{1}{\Delta\, E_x} \leq 4} \tag{15bis}$$

2°. ces inégalités ne peuvent être améliorées, c'est-à-dire qu'on ne peut en resserrer aucune en remplaçant l'un des nombres limites par un nombre également indépendant de X.

---

[1]) Sur une limitation très générale de la dispersion de la médiane (Journal de la Société de Statistique de Paris, avril 1940).

Dans ces inégalités, si l'un des nombres Δ, Θ_x, σ_x est infini, on remplacera le facteur $\frac{1}{\Delta}$, $\frac{1}{\Theta_x}$, $\frac{1}{\sigma_x}$ correspondant par zéro. On a vu que $\frac{1}{\Delta}$ ne peut être infini. Ces inégalités auront alors un sens pour tout nombre X vraiment aléatoire. Pour X „presque sûrement" certain, Θ_x, σ_x, E_x seraient nuls et Δ infini, les inégalités sans être mises en défaut, se trouveraient privées de sens.

Démonstration. On traite immédiatement le cas des bornes inférieures dans les inégalités (13^{bis}), (14^{bis}), (15^{bis}). Il est évident que les seconds membres sont ≥ 0. Pour établir que la borne inférieure, zéro, ne peut être améliorée, c'est à-dire qu'il n'existe aucun nombre positif indépendant de X qui puisse remplacer 0 dans l'un des premiers membres, il suffit de montrer qu'en choisissant à chaque fois convenablement X, chacun des seconds membres peut s'annuler. En ce qui concerne (13^{bis}) et (14^{bis}), il suffit de prendre X, par exemple, selon la formule (1), de sorte que Θ_x et par suite σ_x soient infinis et que cependant Δ soit $\neq 0$. Quand à (15^{bis}) il suffit de prendre X de sorte que E_x soit $\neq 0$ et Δ infini. C'est ce qui aura lieu si l'on prend X de sorte que sa densité de probabilité $\varphi(x)$ vérifie, par exemple:

$$\varphi(x) = \frac{A}{\sqrt{|x|}} \text{ pour } |x| < 4 \,;\; \varphi(x) = 0 \text{ pour } |x| > 4,$$

A étant choisi de sorte que $\int_{-4}^{+4} \varphi(x)\, d\, x = 1$.

Occupons nous maintenant des bornes supérieures des inégalités (13^{bis}), (14^{bis}).

On a, par définition de Δ

$$\frac{F(x) - F(x_0)}{x - x_0} \leq \Delta, \tag{16}$$

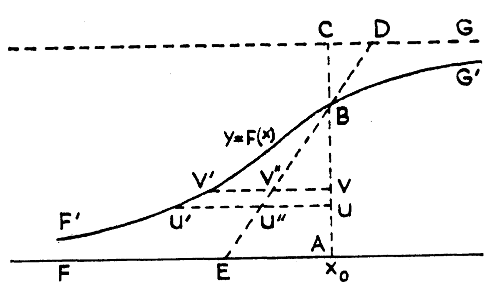
pour tout couple x, x_0, de sorte que

$$F(x) \leq F(x_0) + \Delta(x - x_0) \quad \text{si} \quad x \geq x_0$$
$$F(x) \geq F(x_0) + \Delta(x - x_0) \quad \text{si} \quad x \leq x_0$$

Autrement dit, la droite $y = F(x_0) + \Delta(x - x_0)$ traverse la courbe $y = F(x)$ au point x_0. Désignons par $\psi(x)$ la fonction égale à $F(x_0) + \Delta(x - x_0)$ quand cette dernière fonction est comprise entre 0 et 1 (soit pour $x' \leq x \leq x''$), égale à 0 pour $x \leq x'$ et égale à 1 pour $x \geq x''$. $y = \psi(x)$ est donc la ligne polygonale $F\ E\ B\ D\ G$ de la fig. 4. Et soit Z une variable aléatoire ayant $\psi(x)$ pour fonction de répartition

$$\psi(x) = \text{Prob}\,[Z < x].$$

FIG. 4



On aura

$$F(x) \geq \psi(x) \text{ pour } x \leq x_0,$$
$$F(x) \leq \psi(x) \text{ pour } x \geq x_0.$$

Dans l'hypothèse plus générale où $\Theta_X(x_0)$ est donné par $\int_{-\infty}^{+\infty} | x - x_0 | \, dF(x)$, on voit que $\Theta_X(x_0)$ est la somme des aires curvilignes $BCGG'$ et $BAFF'B$, tandis que $\Theta_Z(x_0)$ est la somme des aires triangulaires BCD, ABE. On a donc $\Theta_X(x_0) \geq \Theta_Z(x_0)$.

Or

$$\Theta_Z(x_0) = \int_{-\infty}^{+\infty} | x - x_0 | \, d\psi(x) = \int_{x'}^{x''} | x - x_0 | \, \Delta \, dx,$$

d'où

$$\Theta_Z(x_0) = \left[\frac{(x'' - x_0)^2}{2} + \frac{(x_0 - x')^2}{2} \right] \Delta = \frac{\beta^2 + \alpha^2}{2} \Delta. \tag{17}$$

Or on a $0 = F(x_0) + (x' - x_0)\Delta$; $1 = F(x_0) + (x'' - x_0)\Delta$, d'où $1 = (x'' - x')\Delta = (\beta + \alpha)\Delta$ et $\dfrac{\beta^2 + \alpha^2}{2}\Delta = \dfrac{\beta^2 + \alpha^2}{2(\beta + \alpha)^2}\dfrac{1}{\Delta} \geq \dfrac{1}{4\Delta}$.

On a donc

$$\Theta_X(x_0) \geq \frac{1}{4\Delta}. \tag{18}$$

En prenant pour x_0 une valeur probable de X, on obtient donc (13^{bis}).

Cette inégalité ne peut être améliorée, car la borne supérieure peut être atteinte.

Il suffit de prendre pour X la variable Z qu'on vient de faire intervenir. En effet, Θ_Z est le minimum de $\Theta_Z(x_0)$ quand x_0 varie. Or, d'après (17), ce minimum a lieu pour $x_0 = \dfrac{x' + x''}{2}$ ou $\beta = \alpha$, et prend alors la valeur $\dfrac{1}{4\Delta}$.

Pour prouver l'inégalité (14) le raisonnement n'est pas moins simple. Observons, en effet, que $[\sigma_X(x_0)]^2 = \int_{-\infty}^{+\infty} (x - x_0)^2 \, dF$ est la somme d'éléments tels que $(UU')^2 \times UV$ dans la fig. 4.

D'autre part, $[\sigma_Z(x_0)]^2 = \int_{-\infty}^{+\infty} (x - x_0)^2 \, d\psi(x)$, étant la somme d'éléments tels que $(UU'')^2 \times UV$, sera au plus égal à $[\sigma_X(x_0)]^2$. Or, $\psi(x)$ est la fonction de répartition de la variable aléatoire Z déjà considérée.

On a donc, d'une part

$$\sigma_X^2(x_0) \geq \sigma_Z^2(x_0)$$

et d'autre part

$$\sigma_Z^2(x_0) = \int_{x'}^{x''} (x - x_0)^2 \, \Delta \, dx = \frac{\beta^3 + \alpha^3}{3} \Delta =$$

$$= \frac{(\beta + \alpha)\Delta}{3} (\beta^2 + \alpha^2 - \alpha\beta) = \frac{1}{3} \left[\frac{(\beta + \alpha)^2 + 3(\beta - \alpha)^2}{4} \right] \geq \frac{(\beta + \alpha)^2}{12} = \frac{1}{12\Delta^2},$$

d'où

$$\sigma_X(x_0) \geq \frac{1}{2\Delta\sqrt{3}}.$$

Observons que, lorsque l'écart quadratique moyen est, comme nous le supposons ici fini, l'intégrale $\int_{-\infty}^{+\infty} | x - x_0 | \, d\, F$ sera finie, donc aussi l'intégrale $\int_{-\infty}^{+\infty} (x - x_0) \, d\, F$, les intégrales $\int_{x_0}^{+\infty} (x - x_0) \, d\, F$ et $\int_{-\infty}^{x_0} (x - x_0) \, d\, F$ seront absolument convergentes et par suite l'intégrale $\int_{-\infty}^{+\infty} x \, d\, F(x)$. Ainsi dans le cas actuel, la valeur moyenne $\int_{-\infty}^{+\infty} x \, [d\, F(x)]$ est bien déterminée. En prenant alors pour x_0 cette valeur moyenne, on a

$$\sigma_x \geq \frac{1}{2 \Delta \sqrt{3}}. \tag{19}$$

Enfin, cette dernière n e p e u t ê t r e a m é l i o r é e. En effet, on voit, comme pour (13), que l'inégalité (19) devient une égalité dans le cas où $X = Z$.

Enfin, on démontre immédiatement (15) par un raisonnement analytique puisqu'en appliquant l'inégalité (16) au cas où x et x_0 comprennent les deux quartiles, on a

$$\tfrac{1}{2} \leq F(x) - F(x_0) \leq \Delta (x - x_0)$$

et en faisant tendre x et x_0 vers ces quartiles, on trouve

$$\tfrac{1}{2} \leq \Delta \, 2 \, E_x. \tag{20}$$

Cette inégalité (20) ne peut non plus être améliorée. Elle devient, en effet, une égalité dans le cas où l'on suppose, par exemple, que la courbe de densité a la forme polygonale $F \, L \, R \, R' \, H \, G$ indiquée sur la fig. 3, $L \, H$ étant choisis de sorte que les aires $L \, Q \, R$ et $H \, Q' \, R'$ soient égales à $\frac{1}{4}$.

Summary: A comparison is made between several measures of dispersion of a random variable X, the mean absolute deviation Θ_x, the quadratic mean deviation σ_x and the probable deviation E_x. It is proved that

$$0 \leq \frac{\Theta_x}{\sigma_x} \leq 1, \quad 0 \leq \frac{E_x}{\Theta_x} \leq 2, \quad 0 \leq \frac{E_x}{\sigma_x} \leq \sqrt{2}. \tag{A}$$

Let Δ_x be the upper bound of the mean density of probability of X between x_1 and x_2:

$$\frac{\text{Prob}\left\{ x_1 < X < x_2 \right\}}{x_2 - x_1} \text{ when } x_1, x_2 \text{ vary.}$$

Then, Δ_x is compared with Θ_x, σ_x, E_x.

As below

$$0 \leq \frac{1}{\Delta \, \Theta_x} \leq 4, \quad 0 \leq \frac{1}{\Delta \, \sigma_x} \leq 2 \sqrt{3}, \quad 0 \leq \frac{1}{\Delta \, E_x} \leq 4. \tag{B}$$

Moreover, it is proved that in the twelve inequalities (A) and (B), none of the limits may be replaced by a more advantageous one.

A comparison is also made (see above formulae 9, 10, 11, p. 7) of Θ_x, σ_x, E_x with a functional measure of dispersion introduced by P a u l L é v y.