JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

PIERRE THIONET

L'école moderne de statisticiens italiens

Journal de la société statistique de Paris, tome 86 (1945), p. 245-255

<http://www.numdam.org/item?id=JSFS_1945__86__245_0>

III

L'ÉCOLE MODERNE DE STATISTICIENS ITALIENS

MONSIEUR LE PRÉSIDENT, MES CHERS CONFRÈRES,

On m'a demandé de remplacer M. Robert Satet, dans l'impossibilité de faire sa communication. On m'a laissé le choix de mon sujet. Je ne pourrai, comme certains d'entre vous, parler d'un récent voyage ou séjour à l'étranger; mais peut-être bénéficierai-je malgré tout de ce même intérêt que nous portons actuellement à tout ce qui peut nous dépayser. Je vous transporterai en effet en Italie, pour vous rendre compte d'une exploration que je fis, il y a déjà deux ans, sans quitter les locaux de la S. G. F., à travers les publications italiennes. Je cherchais à apprécier l'importance des travaux de statistique théorique de nos confrères italiens. M. Huber souhaitait que fût précisée la contribution de Corrado Gini et de ses élèves à la théorie mathématique des *moyennes ;* mais, comme vous le voyez, j'avais immédiatement élargi le sujet. Je fus ensuite amené à quitter Paris durant quelques mois; puis la Libération survint et les statisticiens italiens furent complètement abandonnés.

J'aborde aujourd'hui le sujet avec un certain recul: mon principal souci sera de ne pas ennuyer. Faire un cours de mathématiques sur ces questions serait déplacé ici, d'autant qu'il y faudrait bien un trimestre. Aussi la conférence sera-t-elle complétée par des Annexes plus spécialement mathématiques.

Tout d'abord, je ne serai pas aussi actuel que je le voudrais. La S. G. F. n'a rien reçu d'Italie depuis longtemps; d'après la Revue de l'I. I. S. de 1940 à 1943 (Comptes-rendus des communications faites à la Société italienne de Statistique), Gini s'occupait alors des indices de transvariation, qui constituent un sujet massif que nous avions renoncé à étudier; d'après le numéro de 1944, Gini bataillait toujours contre les théories anglo-saxonnes. Nous nous occuperons donc de travaux antérieurs à 1940.

Plusieurs d'entre vous ont bien connu Gini. Il semble bien qu'il ait eu le tempérament d'un chef d'école. Son intransigeance, touchant ses idées, se signalait dans les réunions d'experts statisticiens de la S. D. N. ou aux sessions de l'I. I. S. Il paraît d'ailleurs n'avoir rien ignoré des travaux d'autrui dont il se méfiait tant. Il n'hésitait pas à écrire (en 1939) : « *Suivant mon habitude d'étudier les questions ex-novo, sans me préoccuper des résultats acquis, sauf à en prendre connaissance ensuite*, je suis parvenu à des conclusions qui, apparemment, sont totalement différentes des propositions..... »

J'ajouterai qu'il montre — et plus encore son disciple Pietra — un sens très vif de la réclame et témoigne d'un profond mépris pour ses grands rivaux, les statisticiens anglais qui, il faut bien le dire, l'ignorent d'ailleurs à peu près complètement. Nous renverrons à un article de Gini, paru en 1926 dans le *Journal of the Royal Statistical Society* (de Londres), et à ceux de Pietra dans la Revue de l'I. I. S. de 1940 (et de 1944). Sous les longues phrases sonores propres à nos voisins latins, on trouve un catalogue des publications de l'École italienne classées systématiquement, ainsi que des détails précieux pour qui veut comprendre cette œuvre.

Tout d'abord, c'est de la Faculté de Droit que viennent la plupart des statisticiens italiens (plutôt que de celle des Sciences). On sait qu'en France l'esprit juridique et l'esprit scientifique font rarement bon ménage. « La méthode favorite des juristes est la déduction. En présence d'une difficulté donnée, l'esprit juridique consiste à rechercher quelle solution découle de la combinaison des principes et des textes...... N'hésitons pas à déclarer qu'un tel art demande plus d'habileté que de bonne foi, plus d'aptitude à la déduction que de patience dans l'observation et de scrupule dans l'affirmation. Et c'est dire que les qualités qui font le bon juriste ne sont pas nécessairement celles qu'on exige du savant. » (Gaétan Pirou, *Les Facultés de Droit ; l'Enseignement économique en France et à l'étranger*. Numéro du « Cinquantenaire de la Revue d'Économie politique »). Naturellement, on fera, dans ces propos, la part de ce qui revient au juriste Pirou, soutenant la thèse de l'incompatibilité de l'esprit scientifique et de l'esprit juridique.

Ainsi les juristes aiment bâtir des théories, des systèmes; et l'on a bien l'impression que Gini rebâtit la Statistique pour son propre compte, que c'est un véritable esprit de système qui le pousse à multiplier les *indices* ou les *moyennes*.

D'autre part, « l'une des caractéristiques de la Statistique italienne est que beaucoup de ses adeptes sont ou désirent faire partie de l'Université, comme professeurs ou maîtres de conférences. Ceci est dû au remarquable prestige qui est attaché au professeur de Faculté, en Italie comme en Allemagne, et à la liberté d'exprimer sa pensée à l'intérieur comme à l'extérieur de l'École, dont il jouissait encore récemment et qui n'est pas encore, même actuellement, entièrement abolie », écrivait Gini en 1926. (J. R. S. S. 1926, p. 703-4.)

Avant 1914, la Statistique était matière obligatoire à la Faculté de Droit au même titre que les Sciences économiques ou financières. L'examen de statistique était obligatoire pour le doctorat en droit. Dans huit écoles supérieures de commerce, il en allait de même; le cours obligatoire durait deux ans. « Ainsi il existait vingt-neuf chaires d'enseignement supérieur ouvertes aux professeurs de statistique et c'était un débouché suffisant pour assurer le placement de tout homme jeune, se consacrant à cette science avec intelligence et assiduité ». Le nombre de chaires s'est encore accru depuis, d'ailleurs : écoles de statistique, école de perfectionnement de Padoue, et même Faculté de Statistique de Rome.

On voit donc que l'Italie présentait, au début du xx^e siècle, des circonstances tout à fait favorables à l'éclosion d'une pléïade de statisticiens. Il faut mettre en regard la situation existant en France, où la statistique ne s'enseigne qu'à Paris (il existe en outre un Institut d'Actuaires à Lyon). Ainsi, un peu avant 1914, Corrado Gini et quelques amis s'installaient dans les chaires que l'on venait de créer et de mettre au concours, et devaient faire école.

Dès 1922 paraissait la revue *Metron*, dirigée par Gini, qui entendait rivaliser avec les grandes revues anglo-saxonnes (*Biometrika*, *Annals of Eugenics*, *Journal of the R. S. S. ou A. S. A.*), — qui ouvrit ses colonnes aux mathématiciens étrangers (Georges Darmois, Lotka, Slutsky, Tchouproff, Charlier, R. A. Fisher) et italiens (Cantelli), tout en servant essentiellement de moyen d'expression à Gini, Pietra et leurs élèves : Galvani, Boldrini, Cisbani, Livi, Bresciano-Turroni et, surtout, de Finetti; enfin, à d'autres statisticiens : Vinci, Savorgnan, pour l'Italie; March, Bunle pour la France, etc...

Les candidats au doctorat de Statistique sont très souvent employés dans les offices de Statistique des grandes villes ou au Bureau central de Statistique du Royaume; professeurs de statistique et bureaux de statistique sont en contacts étroits, de sorte qu'il n'y a pas de cloison étanche, comme en France, entre l'art de rassembler les données, de les critiquer et les interpréter, et la science de la statistique mathématique. Savorgnan, par exemple, appartient à l'Administration.

Nous pensons avoir ainsi suffisamment suggéré sous quels astres était née l'École moderne de statisticiens italiens, sans oublier les conditions politiques singulièrement favorables à un mouvement national, abstraction faite de sa valeur. Les Italiens furent poussés à promettre beaucoup plus qu'ils ne tiendront et surtout à critiquer autrui, ce qui ne serait pas moins intéressant à étudier que la partie constructive de leur œuvre dont nous nous occupons à présent.

Gini reconnaît que la statistique théorique s'est développée grâce à Quételet, Galton, Lexis, Pearson, Yule, Bowley; mais ces statisticiens avaient une formation scientifique et les sciences physiques leur ont servi de modèle; c'étaient des astronomes, biologistes, anthropologistes, qui faisaient en outre de la statistique. Aussi ont-ils abusé de l'hypothèse que les populations de nombres suivent une loi de Gauss-Laplace ou des lois à peine plus générales, cas particuliers qu'ils rencontraient effectivement. L'emploi, pour des séries statistiques quelconques, des méthodes déduites de ces hypothèses est totalement injustifié : à savoir les calculs de moyenne, écart type, coefficients β, rapport de l'écart type à l'écart moyen, et finalement les carrés latins ou le test de $\mathcal{X}^2$. D'autre part, les manuels de Statistique sont un assemblage de chapitres disparates : méthode d'interpolation, méthodes d'ajustement, calcul des indices de corrélation, etc... Selon Gini, « ils ressemblent davantage à une collection d'intéressantes monographies qu'à un exposé ordonné d'une branche indépendante de la science ».

L'Italie voit naître, au contraire, une génération de *purs statisticiens*, qui font de la Statistique et de ses applications le but principal de leurs travaux, au lieu de l'étudier comme une science auxiliaire de l'Anthropologie, de la Médecine, de l'Économie ou de la Science financière. En outre, ils ont tous à l'exposer à leurs élèves. Leur tâche doit donc être de « *compléter et coordonner la science statistique* ». Remarquons que l'expression de Gini n'est pas en réalité *science statistique*, mais *méthodologie statistique*. Et chacun sait que la méthodologie des sciences fait partie, en France, du programme de philosophie.

Soit par exemple les indices des prix, les indices de quantum, la méthode de la population-type et quelques autres méthodes de démographie. Gini montre qu'on peut définir les *méthodes d'élimination*, qui renferment toutes ces méthodes usuelles et d'autres qui le sont moins. Voilà un bel exemple de systématisation dont le gain le plus clair doit être pour les professeurs italiens.

De même, soit des nombres a, b, c,... On emploie depuis longtemps leurs moyennes arithmétique, géométrique et harmonique. Une généralisation due à Dunkel (1909-1910) a permis déjà d'envisager ce que Gini appelle les moyennes *potentielles*. Par exemple :

celle d'ordre 5 : $\sqrt[5]{\dfrac{a^5 + b^5 + c^5}{3}}$, celle d'ordre — 2 : $\dfrac{\sqrt{3}}{\sqrt[2]{\dfrac{1}{a^2} + \dfrac{1}{b^2} + \dfrac{1}{c^2}}}$

Signalons à ce sujet que, si V désigne la vitesse du vent, la puissance moyenne d'un aéromoteur se calcule par une formule $P = \frac{k}{n} \sum_1^n V_i^3$ (n étant le nombre de mesures à intervalles réguliers de la vitesse du vent). On pose $P = k\,W^3$ où W est la moyenne potentielle d'ordre 3 des V_i, et le vent moyen au point de vue des aéromoteurs; W est, d'après un théorème classique dû à Dunkel, très supérieur à la vitesse moyenne du vent $\frac{1}{n} \sum V_i$.

On voit donc que, par pure commodité de langage, on introduit des moyennes potentielles sans en connaître le nom.

Dunkel a inventé également la moyenne combinatoire, à savoir :

$$\frac{a+b+c}{3}; \qquad \sqrt[2]{\frac{ab + bc + ca}{3}}; \qquad \sqrt[3]{a\,b\,c.}$$

La moyenne arithmétique est à la fois potentielle et combinatoire (d'ordre 1).

La moyenne géométrique est à la fois potentielle (d'ordre 0) et combinatoire (d'ordre n).

Gini a généralisé ces formules en introduisant les moyennes combinatoires potentielles telles que :

Combinatoire d'ordre 2
Potentielle d'ordre 3 : $\sqrt[6]{\dfrac{a^3b^3 + b^3c^3 + c^3a^3}{3}}$

Il a défini enfin les moyennes biplanes, c'est-à-dire, par exemple :

$$\frac{ab + bc + ca}{a + b + c}\,; \quad \left(\frac{a^4 + b^4 + c^4}{3\,abc}\right)^{1/4}; \quad \frac{a^3 + b^3 + c^3}{a^2 + b^2 + c^2}\,; \quad \text{etc...}$$

On comprendra mieux le peu de portée pratique de ces formules quand on saura qu'une moyenne biplane peut très bien être extérieure à l'intervalle des observations.

Lorsque la puissance du numérateur tend vers celle du dénominateur, on obtient des formes-limites très curieuses, généralisant la moyenne géométrique.

Les moyennes monoplanes rentrent dans une définition générale due à Bonferroni, à savoir : ξ tel que : $(\Sigma\, p_i)\, f\,(\xi) = \Sigma\, p_i . f\,(x_i)$.

Les moyennes biplanes exigent une définition générale plus large, due à Chisini ou à de Finetti (1926), à savoir : f étant une fonction homogène, ξ tel que :

$$f\,(\xi\xi\ldots\xi) = f\,(x_1\, x_2 \ldots\ldots x_n)$$

et les cas particuliers étudiés supposent en outre f symétrique, le numérateur et le dénominateur étant fonction symétrique élémentaire.

Pizzetti a envisagé, de son côté, une généralisation des moyennes potentielles (qu'il appelle basales) en moyennes exponentielles, radicales, baso-exponentielles, basoradicales :

$$X^{\frac{1}{p}} = \frac{1}{n} \sum_i x_i^{\,p} \qquad \left\{ \begin{array}{l} C^{X} = \dfrac{1}{n}\, \Sigma_i\, C^{x_i};\; C^{\frac{1}{X}} = \dfrac{1}{n} \Sigma\; C^{\frac{1}{x_i}} \\[2ex] X^{X} = \dfrac{1}{n}\, \Sigma_i\, x_i^{x_i};\; X^{\frac{1}{X}} = \dfrac{1}{n} \Sigma\; x_i^{\frac{1}{x_i}} \end{array} \right.$$

Parmi bien d'autres généralisations, Cisbani (1938) a étudié les moyennes potentielles pondérées dont les poids ne sont pas tous de même signe et donné une condition pour que la moyenne reste dans l'intervalle des observations, condition nécessaire et suffisante déjà trouvée par Bonferroni en 1927. En pratique, Giglio a calculé les altitudes moyennes des provinces d'Italie, en pondérant celle des communes par l'accroissement (> 0 ou < 0) de population ou par l'excédent (> 0 ou < 0) des naissances sur les décès; on rencontre alors des altitudes moyennes négatives (dans les provinces de forte émigration); l'intérêt de ce calcul serait de mettre en lumière l'influence de l'altitude sur l'émigration et sur la natalité. On peut d'ailleurs penser que l'altitude crée l'isolement et, par suite, certaines conditions de vie entraînant à leur tour natalité et immigration; l'isolement existant à d'autres altitudes produirait donc les mêmes effets et le calcul effectué semble de peu d'intérêt.

Beaucoup plus intéressante pour la théorie (sinon la pratique) nous a paru l'assimilation tentée par Gini entre *caractères quantitatifs* et *caractères qualitatifs*. La série statistique indiquant combien de personnes ont la taille comprise dans des intervalles successifs se prête au calcul d'une taille moyenne, d'une dispersion des tailles, etc... La série indiquant le nombre de femmes mariées ayant eu 0, 1, 2, 3,... enfants se prête encore aux mêmes calculs, même si l'on trouve par exemple 1,9 enfants en moyenne par femme. Dans ces deux cas les caractères étudiés (taille, nombre d'enfants) sont *quantitatifs* (l'un continu, l'autre discontinu).

Mais il existe bien d'autres caractères, qui ne sont pas quantitatifs mais *qualitatifs ;* par exemple le lieu de naissance peut être à la rigueur défini par deux coordonnées et il a été possible de déterminer le centre de gravité de la population de l'Italie après chaque recensement; mais la profession d'un individu ne peut vraiment pas être considérée autrement que comme un caractère qualitatif. Plusieurs cas sont distingués :

D'abord les *séries rectilignes*, pour lesquelles il y a à la fois un ordre naturel de succession des qualités (ou modalités) et deux qualités extrêmes (par exemple la couleur des yeux ira du bleu céleste au noir en passant par le gris, le châtain, le brun, quand on envisagera une population classée suivant la couleur des yeux) ;

Puis les *séries cycliques*, pour lesquelles un ordre de parcours s'impose, mais non la qualité par laquelle l'énumération doit commencer (par exemple les mois de l'année, les jours de la semaine forment des cycles; la répartition d'une population suivant le jour de la semaine où les individus sont nés constitue une série cyclique).

Enfin les *séries non ordonnées*, comprenant toutes les autres.

Si une qualité est représentée par un point ou un intervalle, une série rectiligne correspondra à des points alignés ou à des segments consécutifs de droite, — une série cyclique à des points sur un cercle ou à un cercle partagé en arcs. Il s'agit là, d'ailleurs, d'une représentation idéale.

Parmi les séries non ordonnées, un cas particulièrement simple est celui des *séries non* *connexes* où les s modalités peuvent se représenter par les s sommets d'un polyèdre régulier dans un espace à $(s - 1)$ dimensions (par exemple : triangle équilatéral, tétraèdre régulier, etc.); mais il est bien facile d'imaginer des répartitions non ordonnées d'un autre espace; telle est la répartition d'une population suivant le sexe, l'état (mariés, célibataires), le statut (Français, étrangers), laquelle est manifestement représentable par les huit sommets d'un cube. Mais les travaux de Gini se limitent aux séries non connexes.

Pour mieux comprendre le genre des difficultés rencontrées, il suffit de penser : 1° aux *couleurs*, qualités pour lesquelles l'arc-en-ciel fournit depuis toujours une représentation linéaire allant du violet au rouge et que la physique a transformées en caractères repérables par des longueurs d'onde moyennes; 2° aux *couleurs politiques*, qui se sont longtemps réparties linéairement dans l'hémicycle de la Chambre jusqu'au jour où un nouveau parti (M. R. P.) n'a pu réussir à s'y intercaler.

L'exemple de distribution non connexe donné par Pietra est celui de la répartition d'une certaine denrée importée par pays de provenance.

Le problème qui s'est posé en Italie est de pouvoir calculer la moyenne, l'écart-type et tous autres indices imaginés par Gini, pour les séries qualitatives comme pour les séries quantitatives.

S'il s'agit d'une *série linéaire*, on choisit très arbitrairement une unité sur l'axe représentatif pour mesurer les écarts entre modalités et repérer celles-ci. Par exemple :

COULEUR DES CHEVEUX

Blond	0	0	−1
Rouge	1	0	−1
Châtain	2	1	1
Noir.	3	2	2
	(Gini)	(Ammon)	(Bed doc)

ÉTAT PILEUX	
Nul	0
Rare.	1/2
Modéré	1
Fort.	2
Exceptionnel . . .	3
(Ammon)	

RÉCOLTE	
Très mauvais. . .	1
Mauvais	2
Moyen.	3
Bon	4
Très bon.	5

Il s'agit de nombres adoptés effectivement en anthropologie ou pour les statistiques agraires.

On peut de même repérer les notes de musique (en ton majeur) comme suit :

do	ré	mi	fa	sol	la	si
2	4	6	7	9	11	13

Le calcul de la moyenne et des autres indices étant alors le même que pour une série quantitative, on n'insistera pas davantage sur ce cas.

Avec *une série cyclique*, on opère de même pour repérer les modalités; mais ces *abscisses* sont déterminées à une période près. Enfin pour *une série non connexe*, tous les écarts entre modalités sont supposés égaux à 1.

Le procédé utilisé en Italie, pour définir la moyenne, la médiane, etc., dans de pareilles conditions, est tout simplement le *principe de conservation des propriétés formelles*, bien connu quand on raisonne dans le domaine des imaginaires.

On reprend les *propriétés formelles* de la moyenne, de la médiane,... et on les transpose pour une série qualitative (cyclique ou non-connexe), comme *définition* de la moyenne, de la médiane, etc.

Pour la moyenne, on a le choix entre les propriétés suivantes :

1° Le point moyen est le barycentre des modalités;
2° La somme algébrique pondérée des écarts entre modalités et moyenne est nulle;
3° La somme pondérée des carrés des écarts entre modalités et moyenne est minimum.

La première propriété, si elle servait de définition à la moyenne d'une série cyclique, donnerait un point extérieur au cercle; aussi l'abandonne-t-on.

En revanche, pour une série cyclique, chacune des autres propriétés, utilisée comme définition, procure des valeurs moyennes qui, sauf exception, concordent. Mais une distribution cyclique donnée aura plusieurs valeurs moyennes.

Par exemple, si les modalités sont les douze mois de l'année représentés par 12 points équidistants sur un cercle, la série dont les 12 fréquences sont égales aura 24 moyennes à savoir les 12 modalités et les points qui s'en déduisent par une rotation de $\frac{\pi}{12}$, ce qui s'explique par une symétrie évidente (remarquez que les modalités satisfont à la 2e, non à la 3e propriété).

Pour une série non connexe, au contraire, Gini a été finalement amené à considérer que le barycentre des modalités était le point moyen.

En ce qui concerne la médiane, en supposant tous les x_i distincts, on admet généralement qu'il existe un intervalle médian entre le n^e et le $(n + 1)^e$ élément lorsque la distribution porte sur un nombre pair 2 n d'éléments. Mais Jackson a proposé de définir la médiane comme la limite, lorsque p tend vers 1, de la valeur x rendant minimum la somme

$$S_p(x) = \sum_1^n |x - x_i|^p$$

Ceci conduit bien au $(n + 1)^e$ élément, si le nombre de ceux-ci est $(2n + 1)$, — mais en outre définit de façon univoque la médiane, à l'intérieur de l'intervalle médian, comme racine de l'*équation de Jackson* (racine unique dans cet intervalle) :

$$(x - x_1).(x - x_2)\ldots\ldots(x - x_n) = (x_{n+1} - x).(x_{n+2} - x)\ldots\ldots(x_{2n} - x).$$

Mais généralement le nombre $(2n)$ d'éléments est supérieur au nombre s de modalités. Soit fi la fréquence de x_i. Aucun problème ne se pose, à moins que l'intervalle médian ne joigne 2 modalités consécutives $x_{r-1}\ x_r$: $\left(\sum_1^{r-1} fi = \sum_r^s fi\right)$; l'équation de Jackson devient finalement :

$$(x - x_1)^{f1}(x - x_2)^{f2}\ldots(x - x_{r-1})^{fr-1} = (x_r - x)^{fr}\ldots(x_{r+1} - x)^{fr+1}\ldots\ldots(x_s - x)^{fs}$$

Cette formule a été adoptée comme définition formelle de la médiane d'une série cyclique et conduit à plusieurs points médians.

Pour une série non connexe, c'est la modalité de fréquence maximum qui joue le rôle de médiane.

Quant aux autres indices définis par Gini, l'extension en a également été faite aux séries cycliques ou non connexes.

Pour bien comprendre le concept d'*indice*, il faut connaître l'idée que se fait Gini de l'emploi des mathématiques en statistique. Ce sont surtout Pietra et ses élèves qui emploient le calcul intégral; Gini s'en tient toujours au discontinu et à l'algèbre ordinaire. Quant au calcul des Probabilités, il n'en fait presque aucun usage. Voici ce qu'il dit aux Anglais :

« Il est universellement admis qu'en statistique, les mathématiques doivent être considérées comme un moyen pour présenter un sujet sous une forme plus ou moins élégante, mais leur mérite ne doit pas nous faire perdre de vue cette vérité fondamentale que, pour la Statistique, elles ne sont jamais qu'un moyen.

« On doit admettre que, tous les problèmes statistiques étant des problèmes quantitatifs, ils peuvent en théorie, être traités mathématiquement; et l'on doit admettre aussi que, si le sujet se prête à une solution théorique, la solution donnée par les mathématiques sera plus rigoureuse et plus précise que toute autre. Cela ne veut pas dire toutefois qu'il soit toujours indiqué d'utiliser cette méthode; la matière statistique à notre disposition peut être trop grossière pour se prêter à l'application de procédés perfectionnés; ou bien, pour les buts poursuivis, il peut sembler inutile de rechercher la précision au delà d'une certaine limite.....

« Sa devise (au statisticien), au lieu d'être « Statistique avec mathématiques », devrait être « Statistique avec le minimum de moyens mathématiques ». (J. R. S. S. 1926, p. 706-7).

Nous ne saurions partager pareille conception si elle signifie, comme c'est bien en effet le cas, que la moindre formule de Calcul des Probabilités est une faute, alors que Gini se permettra des formules (d'algèbre très élémentaire) effroyablement compliquées et inutiles; on peut véritablement parler alors d'esprit de système; mais cette extrême méfiance de notre auteur à l'égard de toute hypothèse supplémentaire gratuite, de tout *schéma* (surtout probabilité), caractérise véritablement ses travaux et son école.

L'Algèbre va être employée pour élaborer des *coefficients* et des *indices*. Par *indices*, il ne faut pas entendre « *index numbers* » (indice des prix, par exemple), Gini se référant à ces derniers à l'aide du vocable « Méthodes d'élimination », — ni donner à ce nom le sens très général qu'il a pour M. Hostelet (indicateur d'un autre fait non directement observé ou observable). Pour Gini un *coefficient* (ou *indice*) est un nombre destiné à repérer une certaine notion abstraite, telle que la variabilité, la concentration, etc. Après avoir défini le maximum d'un coefficient, on en déduit parfois un *indice* variant de 0 à 1, ou de — 1 à + 1 selon les cas. Le choix d'un coefficient ou indice est fortement arbitraire, mais l'Algèbre (c'est-à-dire les propriétés des formules) y joue un grand rôle.

Pourtant il faut, avant tout, savoir de quelle notion on s'occupe, afin de choisir le coefficient le mieux adapté. D'où quelques controverses, notamment quand Gini appelle *connexion* ce que tout le monde nomme *corrélation* et quand, en outre, ayant défini la *concordance* et ses multiples indices, notre auteur appelle l'un d'eux indice de corrélation, par un excès de logique (: dans corrélation, il y a relation, dit-il). On s'explique aisément la confusion où ceci peut jeter les collègues étrangers qui le lisent.

S'agit-il, par exemple, de la notion de dispersion. Pour Gini ce mot n'a de sens que pour les distributions gaussiennes ou voisines, car il s'agit de dispersion autour de la moyenne, c'est-à-dire de la valeur centrale. Pour une distribution asymétrique, on peut prendre la médiane comme seconde valeur centrale; on peut surtout n'en prendre aucune, — puisqu'il n'y en a pas, à proprement parler, — et calculer la différence moyenne, moyenne des différences des valeurs de la série prises deux à deux.

Pour la série a, b, c, on formera donc :

$$\begin{aligned} \mathrm{N} &= |a-b| + |b-c| + |c-a| + |b-a| + |c-b| + |a-c| \\ &= 2|a-b| + 2|b-c| + 2|c-a| \end{aligned}$$

qu'on divisera par le nombre de différences, c'est-à-dire 6 si l'on compte les termes précédents et 9 si l'on compte en outre les différences $|a-a| + |b-b| + |c-c|$. Plus généralement, s'il y a n termes, on pose :

$$\Delta = \frac{\mathrm{N}}{n\,(n-1)},\ \Delta_{\mathrm{R}} = \frac{\mathrm{N}}{n^2} \qquad (\Delta_{\mathrm{R}} : \text{différence moyenne avec répétition}).$$

$$\text{avec } \Delta_{\mathrm{R}} = \frac{n-1}{n}\,\Delta \text{ (sensiblement égales si } n \text{ est grand).}$$

On définit pareillement ${}^2\Delta$ et ${}^2\Delta_{\mathrm{R}}$ à l'aide des carrés des différences; mais ces grandeurs n'ont pas du tout le même intérêt que les premières.

Au total, on considère, avec Δ, Δ_{R}, ${}^2\Delta$ et ${}^2\Delta_{\mathrm{R}}$, les grandeurs suivantes :

— écart moyen par rapport à la moyenne ${}^1\mathrm{S}_{\mathrm{A}}$, à la médiane ${}^1\mathrm{S}_{\mathrm{M}} = e$;

— écart quadratique moyen par rapport à la moyenne ${}^2\mathrm{S}_{\mathrm{A}} = \sigma$, à la médiane ${}^2\mathrm{S}_{\mathrm{M}}$.

Toutes ces expressions s'appellent les indices de *variabilité* (séries quantitatives) ou de *commutativité* (séries qualitatives). Du fait qu'un indice de variabilité est un peu plus grand pour une série que pour une autre, on ne peut conclure que celle-ci est moins dispersée que la première, un autre indice pouvant fournir une inégalité en sens contraire. D'où l'emploi de 5 ou 6 indices destinés à se recouper. Enfin, en aucun cas, on ne devra, pour calculer un indice relatif à une autre notion, se servir des indices de variabilité, ce qui reviendrait à substituer à la série elle-même un seul nombre; on devra, au contraire, calculer tout nouvel indice à l'aide de la série elle-même. Par exemple, ne s'occuper que de la moyenne et de l'écart-type d'une série serait la confondre avec une série de Gauss, ce qui peut conduire à des erreurs considérables. Ces vues paraissent très raisonnables.

« La nécessité de multiplier les indices statistiques de façon à avoir à sa disposition l'indice approprié pour chaque recherche est depuis longtemps reconnue par les statisticiens avisés, qui ne proposent des méthodes qu'en vue de leur application pratique ou ne font des applications qu'en se rappelant des propriétés des méthodes employées. C'est une exigence qui n'est pas limitée à la statistique mais est commune à toute technique. Et ainsi les chirurgiens, les dentistes ou n'importe quels autres spécialistes, aux débuts de la spécialité, se contentaient de quelques instruments indifférentiés, bons à tout faire; tandis que, maintenant, ils disposent de tout un arsenal d'instruments spécialisés, chacun répondant aux exigences d'une ou de quelques-unes des nombreuses opérations qu'on leur demande.....

« A mon avis, il (le statisticien) se tromperait lourdement s'il prétendait contraindre la réalité multiple à coucher sur le lit de Procuste d'un ou de quelques indices choisis parce que le calcul en est facile ou bien en raison de leurs propriétés analytiques. »

Annexe I — Les moyennes (1)

Les Moyennes de Gini.

On prend la moyenne soit de $x_1 < x_2 < \ldots\ x_i \ldots < x_n$, soit de $a\ b\ c$.

Notations : $\dfrac{n!}{c!\,(n-c)!} = \dbinom{n}{c}$; $\sum\limits_{i=1}^{n} (x_i)^p = \mathrm{S}_p$ $\begin{cases} \text{Somme des produits } c \text{ à } c \text{ des } x_i = s_c. \\ \text{Somme des produits } c \text{ à } c \text{ des } x_i^p = s_c^p. \end{cases}$

Moyennes monoplanes : Combinatoires : $\overline{\mathrm{C}^c} = \sqrt[c]{\dfrac{s_c}{\binom{n}{c}}}$

potentielles : $\overline{\mathrm{P}^p} = \sqrt[p]{\dfrac{\mathrm{S}_p}{n}}$

$\overline{\mathrm{P}^0} = \overline{\mathrm{C}^n}$ (moyenne géométrique).

cómbinatoires-potentielles : $\overline{\mathrm{C}^c\,\mathrm{P}^p} = \sqrt[cp]{\dfrac{s_c^p}{\binom{n}{c}}}$ $\Big|$ $\overline{\mathrm{C}^c\,\mathrm{P}^0} = \overline{\mathrm{C}^n}$.

(1) Gini, Metron, XIII-2; Cisbani, Metron, XIII-2 et 3; Gini et Zappa, Metron, XIII-3; Pizzetti, Metron, XIII-4.

Moyennes biplanes : **combinatoires :** $\overline{C^c_d} = \sqrt[c-d]{\frac{\binom{n}{d}\, s_c}{\binom{n}{c}\, s_d}}$ **(Théorèmes :** $\overline{C^c_n} = \frac{1}{\overline{C^{n-c}}}$**)**

potentielles : $\overline{P^p_q} = \sqrt[p-q]{\frac{S_p}{S_q}}$.

Combinatoires-équipotentielles : $\overline{C^c_p\, P^p_q} = \sqrt[p(c-d)]{\frac{\binom{n}{d} s^p_c}{\binom{n}{c} s^q_d}}$.

Équicombinatoires-potentielles : $\overline{C^c_c\, P^p_q} = \sqrt[c(p-q)]{\frac{s^p_c}{s^q_c}}$.

Combinatoires-potentielles : $\overline{C^c_d\, P^p_q} = \sqrt[cp-dq]{\frac{\binom{n}{d} s^p_c}{\binom{n}{c} s^q_d}}$.

Cas limites : **1°) Posons $p = q + h$. Soit** $x^h = \frac{S_{q+h}}{S_q} = \frac{a^{q+h} + b^{q+h} + c^{q+h}}{a^q + b^q + c^q}$

$$\text{Formons } \frac{x^h - 1}{h} = \frac{\frac{a^{q+h} - a^q}{h} + \frac{b^{q+h} - b^q}{h} + \frac{c^{q+h} - c^q}{h}}{a^q + b^q + c^q}$$

D'où, à la limite : $Lx = \frac{a^q\, La + b^q\, Lb + c^q\, Lc}{a + b + c'}$; $x = \sqrt[a^q + b^q + c^q]{a^{aq}\, b^{bq}\, c^{cq}}$

d'où $\overline{P^q_q} = \sqrt[S_q]{\prod_{i=1}^{i=n} x_i^{x_i^q}}$ **(moyenne géométrique lorsque $q = 0$).**

2°) Cherchons de même l'expression de : $\overline{C^c_c\, P^q_q}$ **sur un cas particulier:**

Soit $(x^2)^h = \frac{(ab)^{q+h} + (bc)^{q+h} + (ca)^{q+h}}{(ab)^q + (bc)^q + (ca)^q}$ **; il vient de même, à la limite :**

$$x^2 = \sqrt[(ab)^q + (bc)^q + (ca)^q]{(ab)^{(ab)^q}\, (bc)^{(bc)^q}\, (ca)^{(ca)^q}}$$

$$\text{où } x^2 = \sqrt[(a^q b^q) + (b^q c^q) + (c^q a^q)]{a^{a^q (b^q + c^q)}\, b^{b^q (c^q + a^q)}\, c^{c^q (a^q + b^q)}}$$

On a jugé inutile de reproduire l'expression générale de $\overline{C^c_c\, P^q_q}$.

Quelques propriétés de ces moyennes.

1° *Moyenne monoplane.* $\overline{P^p}$ **croît avec p (proposition classique due à Dunkel).**
$\overline{C^c}$ **croît avec c.**
$\overline{C^c\, P^p}$ **croît avec p, c restant fixe, — ou décroît quand c croît p restant fixe.**

Pour démontrer cette dernière proposition, on écrit :

$$\overline{C^c\, P^p} = \sqrt[c]{\sqrt[p]{\frac{s^p_c}{\binom{n}{c}}}} = \sqrt[p]{\sqrt[c]{\frac{s^p_c}{\binom{n}{c}}}} \text{ avec } s^p_c = \Sigma\, a^p\, b^p \ldots = \Sigma\, (a\, b \ldots)^p$$

et on se sert des propriétés d'une moyenne combinatoire des a, b,... et d'une moyenne potentielle de $(ab...)$.

2° *Moyenne biplane potentielle :* **Tout d'abord on a :**

$$P_q^p = P_p^q \left(\text{car } \sqrt[p-q]{\frac{Sp}{Sq}} = \sqrt[q-p]{\frac{Sq}{Sp}}\right)$$

Théorème : $\overline{P_q^p}$ croît avec p, $(p - q)$ restant fixe.

Posons : $p - q = h$, $x^h = \dfrac{a^{q+h} + b^{q+h}}{a^q + b^q}$, $y = \mathrm{L}\, x^h = \mathrm{L}\,(a^{q+h} + b^{q+h}) - \mathrm{L}\,(a^q + b^q)$

$$\frac{dy}{dq} = \frac{a^{q+h}\,\mathrm{L}\,a + b^{q+h}\,\mathrm{L}\,b}{a^{q+h} + b^{q+h}} - \frac{a^q\,\mathrm{L}\,a + b^q\,\mathrm{L}b}{a^q + b^q}.$$

Cette dérivée sera positive si la dernière expression est une fonction croissante de q.

Soit : $$z = \frac{a^q\,\mathrm{L}\,a + b^q\,\mathrm{L}\,b}{a^q + b^q}$$

Or $$\frac{dz}{dp}\,(a^q + b^q)^2 = [a^q\,(\mathrm{L}\,a)^2 + b^q\,(\mathrm{L}\,b)^2]\,(a^q + b^q) - (a^q\,\mathrm{L}\,a + b^q\,\mathrm{L}\,b)^2.$$
$$= a^q\,b^q\,(\mathrm{L}\,a - \mathrm{L}\,b)^2 \quad \text{expression positive, } cqfd.$$

Théorème : $\overline{P_q^p}$ croît avec p, q restant fixe.

En effet : $$x^h = \frac{a^{q+h} + b^{q+h} + \dots}{a^q + b^q + \dots} = \frac{a^q}{Sq}\,a^h + \frac{b_q}{S_q}\,b^h + \dots$$

Donc x est la moyenne potentielle des a avec une certaine pondération; on sait que cette moyenne croît avec l'ordre h, q étant fixe, donc avec $p = q + h$ *cqfd.*

Limites : Si q est fixe et si p tend $\begin{cases} \text{vers } + \infty, & \overline{P_q^p} \text{ tend vers } x_n. \\ \text{vers } - \infty, & \overline{P_q^p} \text{ tend vers } x_1. \end{cases}$

3° *Moyenne biplane combinatoire :* Gini et Zappa ont établi que $\overline{C_d^c}$ décroît quand c croît, d restant fixe, — ou quand c croît, $c - d$ restant fixe.

4° *Moyenne biplane générale :* Une condition suffisante pour que $\overline{C_d^c\,P_q^p}$ soit intérieure à l'intervalle $(x_1\,x_n)$ est que $(c - d)\,p$ et $d\,(p - q)$ ne soient pas de signes contraires.

Un cas où la moyenne sort de l'intervalle est signalé : celui de $\overline{C_1^n\,P_1^{\frac{2}{n}}}$.

Cas de $pc = dq$: Si p tend vers $q\dfrac{d}{c}$, la moyenne biplane tend vers l'infini ou zéro selon que l'expression sous radical est plus grande ou plus petite que 1.

Moyennes de Cisbani et Galvani.

Moyenne potentielle pondérée : $\overline{p^p} = \sqrt[p]{\dfrac{\Sigma\, m_i\, x_i^p}{\Sigma\, m_i}}$ avec $\overline{p^0} = \sqrt[\Sigma m_i]{x_1^{m_1} \dots x_n^{m_n}}$

Distribution continue : $\overline{p^p} = \sqrt[p]{\dfrac{\mathrm{S}\, m\,(t)\,.\,x^p\,(t)\,.\,dt}{\mathrm{S}\, m\,(t)\, dt}}$

1. *Pondération égale à 1.*

Moyenne d'un segment (a, b) (Galvani) :
en posant $dt = \lambda\, dx$ . . . .
On s'assure que ces expressions sont bien continues.

$$\begin{cases} \dfrac{b^{p+1} - a^{p+1}}{(p+1)\,(b-a)} & \text{si } p \neq -1 \\[2ex] \dfrac{\mathrm{L}\,b - \mathrm{L}\,a}{b-a} & \text{si } p = -1 \\[2ex] \dfrac{1}{e}\left(\dfrac{b^b}{a^a}\right)^{\frac{1}{b-a}} & \text{si } p = 0. \end{cases}$$

Même moyenne (Cisbani),
en posant $x^j = t$ . . . .
On s'assure encore de la continuité.

$$\begin{cases} \dfrac{b^{p+j} - a^{p+j}}{(p/j + 1)\,(b-a)} & \text{si } p \neq -j \\[2ex] \dfrac{\mathrm{L}\,b^j - \mathrm{L}\,a^j}{b^j - a^j} & \text{si } p = -j \\[2ex] \dfrac{1}{e^{\frac{1}{j}}}\left(\dfrac{b^{bj}}{a^{aj}}\right)^{\frac{1}{b^j - a^j}} & \text{si } p = 0. \end{cases}$$

REMARQUE : La moyenne de Galvani était la limite d'une moyenne relative à des nombres $a, x_1, x_2, \ldots x_n \Rightarrow b$ formant une progression arithmétique. Poser $t = x^j$ revient à prendre des nombres $a^j, x_1^j, x_2^j, \ldots x_n^j$ en progression arithmétique (puis à faire tendre la raison vers zéro).

Propriétés de la fonction Z (p, j) ainsi définie :
Si p tend vers $+\infty$, j restant fixe (ou j vers $+\infty$, p restant fixe), Z tend vers b.
C'est vers a que tend Z quand j tend vers $-\infty$.
Le plan $p = 0$ est diamétral conjugué des cordes de direction $2j + p = 0$ (échanger j et $p + j$: Z ne change pas).

$Z(p, j) \, . \, Z(-p, -j) = ab$. En particulier, si $2j + p = 0$, on a $Z(p, j) = \sqrt{ab}$. On a ainsi trouvé une droite sur la surface d'équation $Z = Z(p, j)$.

2. *Coefficients de pondération positifs ou négatifs.*

Conditions pour qu'une moyenne potentielle pondérée ne sorte pas de l'intervalle moyenné :

$$x^p(a) \leqslant \frac{\int_a^b x^q(t) \, . \, m(t) \, . \, dt}{\int_a^b m(t) \, dt} \leqslant x^p(b).$$

où l'on peut toujours supposer $\int_a^b m(t) \, dt > 0$.

Cette double inégalité se transforme comme suit par la formule de la moyenne d'Ossian-Bonnet :

$\int_o^u m(t) \, . \, dt \geqslant 0$ avec $a < u < b$	$\int_v^b m(t) \, dt \geqslant 0.$ $a < v < b$	Conditions nécessaires
qu'on peut aussi écrire : $\int_a^b m(t) \, dt \geqslant \int_a^u m(t) \, dt \geqslant 0.$		et suffisantes.

REMARQUE : Il suffira, pour que ces conditions soient remplies, qu'elles le soient pour *toute* valeur $u = t$ de l'intervalle (a, b). (Ces conditions *suffisantes* sont seules utilisables en pratique).

On peut passer de là aux inégalités relatives à une distribution finie.

Moyennes de Pizzetti :

Théorème de Bonferroni : La moyenne exponentielle est supérieure à la moyenne arithmétique si la base est supérieure à l'unité.

Démonstration : $$C^Z = \frac{C^{x_1} + \ldots + C^{x_n}}{n} = \frac{e^{x_1 LC} + \ldots + e^{x_n LC}}{n}$$

d'où
$$C^Z = 1 + \frac{x_1 + \ldots + x_n}{n} LC + \frac{x_1^2 + \ldots + x_n^2}{n} \frac{L^2C}{2} + \ldots$$

ou :
$$C^Z = 1 + \overline{P^1} \, LC + (\overline{P^2})^2 \frac{L^2 C}{2} + \ldots..$$

alors que $$C^{\overline{P^1}} = 1 + \overline{P^1} \, LC + (\overline{P^1})^2 \frac{L^2 C}{2} + \ldots..$$

Or on sait que : $\overline{P^1} < \overline{P^2} < \overline{P^3} < \ldots..$

Donc : $C^{\overline{P^1}} < C^Z$,

Donc : $\overline{P^1} < Z$ *cqfd.*

On pose alors $\overline{P^1} = A$, moyenne arithmétique.

Théorèmes de Pizzetti : Soit $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_k \leqslant A \leqslant x_{k-1} \leqslant \ldots \leqslant x_n$ la série statistique étudiée.

On démontre que : 1) $\sum_1^n x_i^{x_i} > \sum_1^n A^{x_i}$ $\left[\text{se ramène à } (\overline{P^{x_x}})^{x_x} > (\overline{P^1})^{x_x}\right]$

2) $\sum_1^n x_i^{x_i} > \sum_1^n x_i^A$ $\left[\text{se ramène à } (\overline{P^h})^h > (\overline{P^1})^h, h = 2, 3, 4 \ldots.\right]$

3) $\sum_1^n x_i^{\frac{1}{x_i}} < \sum_1^n A^{\frac{1}{x_i}}$

4) $\sum_{1}^{n} x_i^{x_i} > \sum_{1}^{n} (\overline{P}_r)^{x_i}$ (si $r < x_k$)

5) $\sum_{1}^{n} x_i^{x_i} > \sum_{1}^{n} (x)^{Pr}$ (sous certaines conditions pour r).

Soit, par définition : $C^z = \frac{1}{n} \Sigma C^{x_i}$, $C^{\frac{1}{v}} = \frac{1}{n} \Sigma C^{\frac{1}{x_i}}$, $y^y = \frac{1}{n} \Sigma x_i^{x_i}$, $u^{\frac{1}{u}} = \frac{1}{n} \Sigma x_i^{\frac{1}{x_i}}$,

On en tire les inégalités suivantes :

$$y^y > A^z \ ; \qquad y^y > x^A \ ; \qquad v^{\frac{1}{v}} < A^{\frac{1}{u}}$$

(*A suivre.*)

Pierre Thionet.