JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

R. RISSER

Les principes de la statistique mathématique

Journal de la société statistique de Paris, tome 77 (1936), p. 337-379

<http://www.numdam.org/item?id=JSFS_1936__77__337_0>

JOURNAL

DE LA

SOCIÉTÉ DE STATISTIQUE DE PARIS

N° 10. — OCTOBRE 1936

I

LES PRINCIPES DE LA STATISTIQUE MATHÉMATIQUE

DEUXIÈME PARTIE

Corrélation. — Covariation.

CHAPITRE I

DÉPENDANCE ET CORRÉLATION — LES MÉTHODES ANCIENNES DE CALCUL DE L'INDICE DE DÉPENDANCE DE DEUX SÉRIES STATISTIQUES

Quel que soit le domaine où se poursuivent les recherches scientifiques, le rôle du chercheur consiste, après avoir étudié les effets d'un phénomène, d'en découvrir les causes; pour lui, la notion de dépendance causale intervient immédiatement, car la cause et la conséquence sont indissolublement liées, et la plus ou moins grande rigidité de la dépendance ne peut être mise en question.

Admettons qu'aux causes A et B soient associés d'une manière absolue les effets respectifs A′ et B′; il s'ensuit que le phénomène X formé par la conjonction de A et de B a pour conséquence un phénomène Y résultant de la combinaison des effets A′ et B′, et nous dirons alors que X et Y sont *en liaison indissoluble.*

Toutefois, si X, provenant toujours d'une combinaison de A et de B, Y était formé de A′ + B′ + C′, il est exact que Y ne pourra jamais être observé sans que X l'ait été auparavant, mais néanmoins X pourra fort bien être suivi d'un effet autre que Y, par exemple dans le cas où ce dernier effet serait la résultante de A′, B′ et D′. On peut aussi concevoir le cas inverse afférent à l'apparition de l'effet Y grâce à l'intervention d'une cause autre que X, et enfin étudier le cas où X et Y auraient les compositions respectives (A + B) et (A′ + C′), cas qui met en lumière pour la cause X des effets divers et aussi pour l'effet Y des causes diverses.

La rigidité de la dépendance entre X et Y et le coefficient qui la caractérise, tiennent en partie à la constitution de X et de Y ; plus grande est l'importance dans chacun de ces phénomènes des éléments causalement liés, plus rigide est la dépendance.

On conçoit, par exemple, que la dépendance héréditaire entre les pères et les fils soit plus accusée que celle se manifestant entre les grands pères et leurs petits-fils.

En définitive, les dépendances apparaissant dans le domaine statistique sont beaucoup moins faciles à déceler que celles se manifestant dans certains autres domaines de l'activité scientifique, en raison même de la multitude des phénomènes secondaires qui viennent se greffer sur les phénomènes principaux, et de l'influence de ces mêmes phénomènes secondaires à la fois dans le temps et dans l'espace. Si la rigidité plus ou moins affirmée semble une des caractéristiques des dépendances « non indissolubles » aux termes mêmes de Tschuprow, il reste à étudier les propriétés essentielles de ces dépendances. L'ensemble des recherches relatives à la détermination des coefficients de dépendance et de leurs propriétés limites constitue l'objet de la théorie de la covariation et de la corrélation ; l'étude des procédés qui ont précédé l'introduction de la liaison stochastique va dès maintenant appeler notre attention.

Les problèmes que nous allons étudier au cours de la deuxième partie peuvent être considérés comme des problèmes fondamentaux communs à toutes les disciplines statistiques correspondant soit au domaine des sciences naturelles, soit à celui des sciences sociales.

C'est à Francis Galton que nous devons tout ce mouvement scientifique, car c'est dans ses belles recherches de biométrie qu'il faut en réalité chercher les éléments des théories modernes statistiques.

C'est en effet Galton qui — le premier — a appliqué la méthode statistique au problème de l'évolution des êtres organisés. La lecture des *Lettres sur la théorie des probabilités*, dit M. Yule, lui suggéra l'idée de cette application ; Galton doit donc de ce fait être regardé comme le continuateur de Quételet dans l'étude statistique des questions biologiques.

N'ayant point fait de mathématiques, son esprit éminemment curieux et ingénieux chercha à se représenter par des moyens simples les formules au moyen desquelles on fait une description brève des séries d'observations.

Il fut ainsi amené à réaliser la courbe des erreurs en utilisant ce jouet bien connu où des billes, tombant verticalement le long d'un plan hérissé de clous, se disposent d'elles mêmes en colonnes inégales ; il caractérisa les modes de distribution — et cela sans calculs laborieux — en introduisant la méthode des quartiles ou des percentiles.

A la tendance qu'avait eu Quételet de concentrer la méthode statistique sur les comparaisons de moyennes, Galton moins exclusif fit appel à l'étude de la variabilité, sans délaisser le procédé d'investigation du grand statisticien belge.

« On comprend difficilement, dit-il dans son principal ouvrage, pourquoi les statisticiens limitent d'ordinaire leurs recherches aux moyennes et n'envisagent pas les choses d'une manière plus compréhensive.

Et il ajoute que certains esprits ont horreur de la statistique, mais que, pour sa part, il la trouve remplie de beauté et d'intérêt.

« Traitée délicatement, sans brutalité, par des méthodes valables, son pouvoir d'analyse, dit-il, à l'égard des phénomènes compliqués est extraordinaire. »

En définitive, l'on doit à Galton le principe de l'une des méthodes les plus fécondes de la statistique, et considérer ce savant comme le créateur de la théorie de la corrélation; il a eu aussi l'idée et indiqué le moyen d'étendre aux observations qualitatives, les méthodes appliquées aux observations quantitatives.

I — *Les méthodes anciennes de calcul de l'indice de dépendance de deux séries statistiques.*

Ce qui intéresse d'une manière particulière le statisticien, c'est déterminer la dépendance entre les phénomènes donnant lieu à représentation statistique; ce problème l'a toujours préoccupé, mais il n'en a constaté l'importance considérable qu'à la suite des progrès rapides des méthodes statistiques dans le domaine des sciences naturelles.

Ainsi que le dit avec beaucoup de netteté Tschuprow dans son remarquable exposé (1), la théorie de la corrélation avec K. Pearson et ses disciples « prit dès le début des formes mathématiques qui furent une pierre d'achoppement pour les défenseurs des anciennes méthodes. Il en est résulté pour les statisticiens une scission dont le caractère a dépassé les limites normales. Ceux qu'on appelle les mathématiciens manifestent une tendance à rejeter dédaigneusement comme sans valeur les méthodes élémentaires des non mathématiciens, considérées par eux comme rudimentaires et insuffisamment approfondies. De leur côté, les non mathématiciens regardent les procédés mathématiques comme un jeu de calcul, scientifiquement stérile, qui, par l'éclat trompeur d'une précision pratiquement irréalisable, fait illusion sur les esprits non avertis, mais qui ne peut tenir devant l'examen du statisticien expérimenté ».

En réalité, on peut dire que dans les deux camps, l'on a exagéré, et même ajouter, sans crainte d'être réfuté, que la théorie moderne de la corrélation, dont on doit les grands principes aux spécialistes des sciences naturelles, apparaît comme un développement logique des anciens principes, et a ses racines profondes dans les travaux des statisticiens spécialisés dans les sciences sociales; c'est ce que vont nous faire apparaître l'examen de l'indice de Fechner et l'étude des indices caractérisant la covariation.

Jusqu'ici, nous n'avons considéré que des séries statistiques simples, que l'on a pu faire entrer dans l'un des types bien définis au point de vue de la représentation analytique, et nous nous sommes borné à rechercher si chacune de ces séries pouvait être caractérisée par un schéma donné, ou par des tirages dans une ou plusieurs urnes.

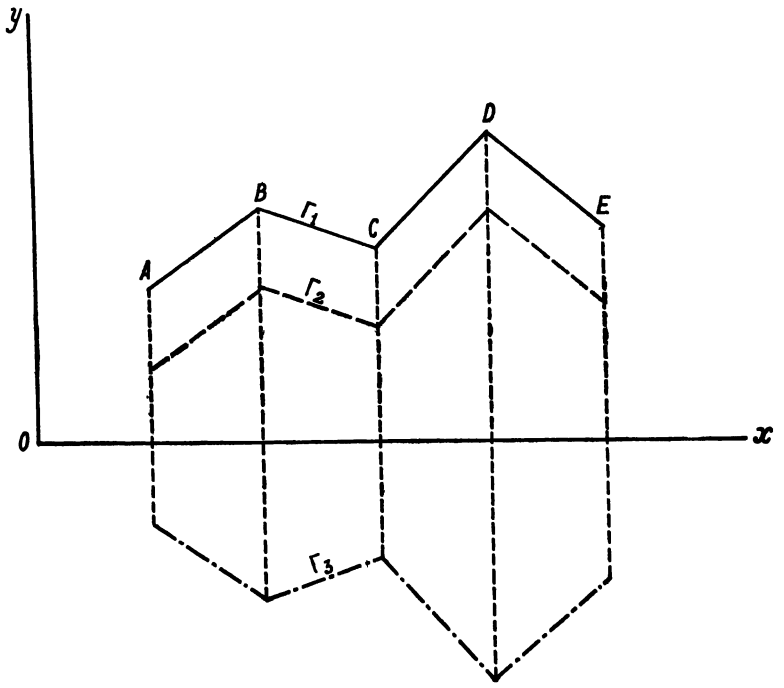
Nous allons aborder maintenant l'examen des procédés permettant aux statisticiens de fixer le mode de dépendance d'un phénomène P_1 par rapport à un autre phénomène P_2, en nous efforçant de mettre en lumière la chaîne logique qui a conduit à l'adoption des divers indices caractéristiques de la liaison de deux séries.

---

(1) *Grund begriffe und Grund problème der Korrelations théorie* (Leçons professées à l'Université de Christiana, chapitre I, p. 2).

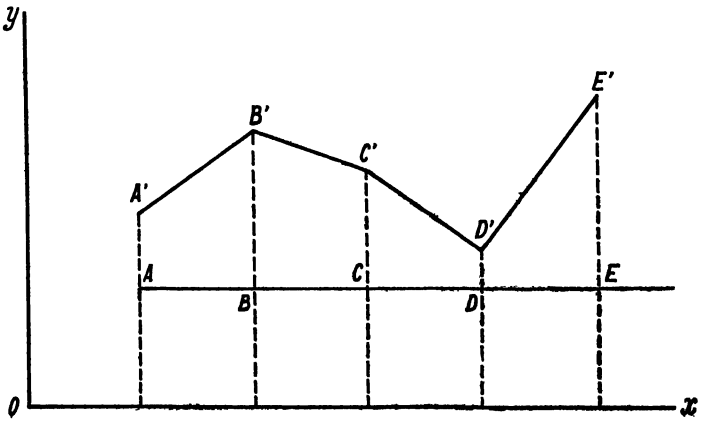
II. — *De la dépendance ou de l'indépendance apparente complètes entre deux séries de faits.*

Dépendance apparente absolue. — Considérons un phénomène P_1 dont la caractéristique statistique pour l'ensemble de la période $(x, x+1)$ est définie par l'ordonnée y_x, et établissons le graphique représentatif du phénomène au moyen de droites joignant les points $(y_x, y_{x+1}) \ldots\ldots, (y_{x+n-1})\ (y_{x+n})$; soit Γ_1, le graphique ainsi obtenu (A B C D E).



On peut concevoir un tel graphique déplacé parallèlement à *oy* d'une quantité k (positive ou négative), et donnant ainsi naissance au graphique Γ_2, puis au graphique Γ_3, obtenu en prenant le symétrique de Γ_2 par rapport à *ox* (Voir fig. 1).

Si donc les phénomènes à l'étude P_1 et P_2 sont représentés respectivement par les graphiques Γ_1 et Γ_2, ils sont en *dépendance apparente complète positive*,



et dans le cas où les séries de faits sont représentées par Γ_1 et Γ_3, la dépendance apparente est *complète*, mais *inverse* ou *négative*.

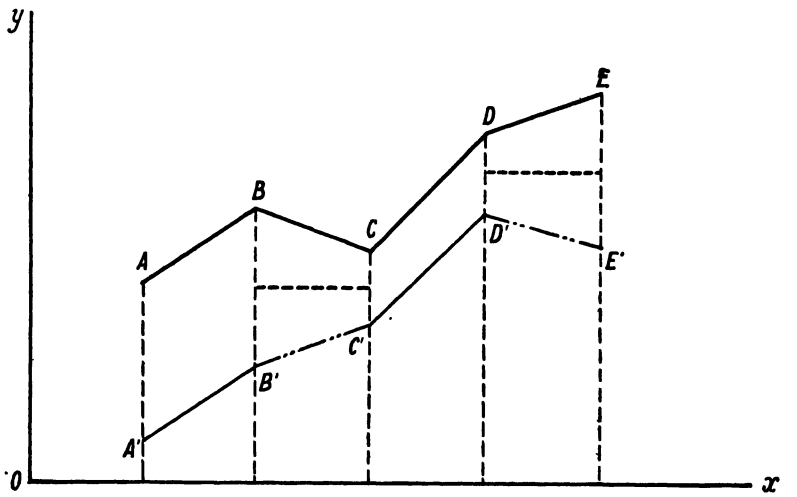
On peut aussi associer une translation k, parallèle à oy, une autre translation k' parallèle à ox afférente à p unités de période pour le graphique Γ_1 et l'on réalise ainsi un graphique Γ'_2; on pourra alors dire que le phénomène P'_2 défini par le graphique Γ'_2 est en concordance absolue positive avec P, mais avec décalage dans le temps.

Si l'on adjoignait en plus des translations k et k' précitées du graphique Γ_1 une symétrie par rapport à ox, on passerait du graphique Γ'_2 à Γ'_3 et le phénomène P'_3 qui serait défini graphiquement par Γ'_3 serait en *concordance apparente absolue négative* avec le phénomène P_1, mais avec décalage dans le temps (1).

Indépendance apparente complète. — Il existe — comme l'on peut s'en rendre compte facilement — des phénomènes P_1 et P_2 manifestant une indifférence apparente complète.

En effet, si le phénomène P_1 ne subit au cours du temps aucune fluctuation (c'est à dire si le graphique qui lui correspond est représenté par une droite parallèle à l'axe des x), alors que le phénomène P_2 offre des variations d'une période à la suivante, on peut dire que P_1 n'est nullement en liaison avec P_2, encore que P_1 est en état d'indépendance apparente complète par rapport à P_2 (Voir fig. 2).

Si maintenant, consacrant la dépendance absolue dans chacun des intervalles, l'on introduit l'*alternance*, *ce qui revient à faire suivre une dépendance absolue positive d'une dépendance absolue négative*, on réalise ainsi un nouveau type d'indépendance complète. Les phénomènes P_1 et P_2 étant alternative-



ment en accord, puis en désaccord, on doit les considérer comme rentrant dans un type d'*indépendance apparente complète* (Voir fig. 3).

Dépendance apparente partielle. — Les phénomènes naturels ne rentrent jamais ou presque jamais dans les cas concrets présentés ci-dessus; en les associant deux à deux, on peut évidemment, par comparaison des graphiques y afférents,

(1) L. MARCH. *Les représentations graphiques et la statistique comparative* (Journal de la Société de Statistique de Paris, novembre 1904 et janvier 1905).

entrevoir une dépendance plus ou moins marquée, mais l'on se trouve toujours dans l'impossibilité d'en fixer la *puissance par l'intermédiaire d'un chiffre.*

III. — *Indice de dépendance.*

Soient ${}_{(1)}y_x$ et ${}_{(2)}y_x$ les valeurs des fonctions représentatives des phénomènes 1 et 2 dans l'intervalle de temps, $(x,\ x+1)$, et ${}_{(1)}\Delta_x = {}_{(1)}y_x - {}_{(1)}y_{x+1}$, ${}_{(2)}\Delta_x = {}_{(2)}y_x - {}_{(2)}y_{x+1}$.

Au lieu de considérer les valeurs effectives des ${}_{(1)}\Delta$ et des ${}_{(2)}\Delta$, Fechner (1) n'envisage que leurs signes respectifs, et associe ensuite ces signes pour la période $(x,\ x+1)$, en donnant à l'association $[{}_{(1)}\Delta_x,\ {}_{(2)}\Delta_x]$ l'appellation de *concordance* ou de *discordance*, suivant que les deux différences qui la caractérisent ont le même signe, ou un signe contraire. L'association où l'une des quantités Δ est nulle correspond d'après Fechner à une *indifférence.*

Après avoir désigné par C le nombre des concordances, D celui des discordances et I celui des indifférences, Fechner fixe pour valeur de l'indice de dépendance des deux séries l'expression :

(1) $\quad i = \dfrac{C - D}{C + D + I} = \dfrac{C - D}{n}$, n étant le nombre d'intervalles de la période envisagée.

Le calcul de cet indice ne faisant intervenir que les *signes des différences successives* ${}_{(1)}\Delta$ et ${}_{(2)}\Delta$ *et le signe de leur association*, présente le très grand inconvénient de ne point tenir compte des valeurs de ces différences, car le mode de détermination de l'indice donne autant d'importance aux petites qu'aux grandes différences.

Coefficient de dépendance. — Pour faire disparaître cette cause d'erreur, il suffit de considérer l'expression .

(2) $$j_1 = \frac{\Sigma\ {}_{(1)}\Delta\ {}_{(2)}\Delta}{\Sigma\left|{}_{(1)}\Delta\ {}_{(2)}\Delta\right|}$$

qui a pour valeur $+1$ ou -1, suivant que la dépendance complète est positive ou négative, c'est-à dire lorsque tous les couples de valeurs associées sont de même sens ou de sens contraire.

Aussi de nombreux statisticiens ont — avec Cheysson — substitué aux ordonnées y_x le rapport $Y_x = \dfrac{y_x}{\Sigma \dfrac{y_x}{n}}$, et cela en vue de rendre les courbes comparables, puis, cette opération effectuée, ont formé les différences :

$$\Delta_x' = Y_x - Y_{x+1}$$

et enfin évalué :

(3) $$j_2 = \frac{\Sigma\ {}_{(1)}\Delta'_x \cdot {}_{(2)}\Delta'_{x+p}}{\Sigma\left|{}_{(1)}\Delta'_x \cdot {}_{(2)}\Delta'_{x+p}\right|}$$

pour avoir une idée de la dépendance entre le phénomène P_1 à l'époque x et le phénomène P_2 à l'époque $(x+p)$, pour toute la période $(x_0,\ x_0+n)$.

La coïncidence étant d'autant moins parfaite que les graphiques s'écartent davantage, on peut dire que la dépendance entre les deux phénomènes P_1 et P_2

(1) Fechner. Œuvres posthumes publiées par G. Lipp, sous le titre: *Kollektivmasslehre*, p. 305. Leipzig, 1897.

est d'autant moindre que l'écartement des graphiques envisagés est plus prononcé; dans ces conditions, l'on peut dire que la mesure de la dépendance de la liaison de P_1 avec P_2 est définie par le coefficient k.

$k = 1 - \lambda \Sigma ({}_{(1)}\Delta'_x - {}_{(2)}\Delta'_x)^2$, qui aura pour valeur 1, si pour tous les intervalles de la période envisagée ${}_{(1)}\Delta' = {}_{(2)}\Delta'$. On déterminera λ en tenant compte de ce que l'indépendance complète se trouve caractérisée par $\Sigma_{(1)}\Delta' \, {}_{(2)}\Delta' = o$, et l'on trouve ainsi que le coefficient de dépendance de k a pour valeur :

$$k = \frac{\Sigma_{(1)}\Delta' . {}_{(2)}\Delta'}{\frac{\Sigma_{(1)}\Delta'^2 + {}_{(2)}\Delta'^2}{2}} \tag{4}$$

Généralisation. — Coefficient de covariation.

L'extension de la dépendance absolue aux cas où chaque variation ${}_{(1)}\Delta$ de l'une des séries se trouve en rapport constant avec la variation ${}_{(2)}\Delta$ de l'autre série, conduit à cette remarque que l'on peut trouver pour le coefficient de dépendance de deux séries une valeur k' résultant de la substitution de la moyenne géométrique $\sqrt{\Sigma_{(1)}\Delta'^2 \, \Sigma_{(2)}\Delta'^2}$, à la moyenne arithmétique

$$\frac{\Sigma_{(1)}\Delta'^2 + \Sigma_{(2)}\Delta'^2}{2}.$$

Si m désigne la valeur moyenne des variations de la première série, m' celle des variations de la seconde série, $\left[\frac{{}_{(1)}\Delta'}{m} \cdot \frac{{}_{(2)}\Delta'}{m'}\right]$ représente avec son poids et son signe la valeur d'une concordance ou d'une discordance; il résulte de là que l'on peut appliquer cette conception à l'indice de Fechner et lui donner la valeur $k' = \frac{1}{n} \frac{\Sigma_{(1)}\Delta' \cdot {}_{(2)}\Delta'}{m \, m'}$, où n n'est autre que le nombre des intervalles, ou encore le nombre total des concordances, discordances et indifférences.

Si l'on veut que k' soit égal à 1 lorsque les ${}_{(1)}\Delta'$ et les ${}_{(2)}\Delta'$ sont tous égaux deux à deux et de même signe, il faut prendre pour k' la valeur

$$\frac{\Sigma_{(1)}\Delta' . {}_{(2)}\Delta'}{\sqrt{\Sigma_{(1)}\Delta'^2} \sqrt{\Sigma_{(2)}\Delta'^2}}.$$

Si l'on désigne par σ_1 et σ_2 les écarts moyens quadratiques relatifs respectivement à ${}_{(1)}\Delta'$ et à ${}_{(2)}\Delta'$, on remarque immédiatement que le coefficient k' peut encore s'écrire :

$$k' = \frac{\Sigma_{(1)}\Delta' . {}_{(2)}\Delta'}{\sigma_1 \, \sigma_2}. \tag{5}$$

Ce coefficient de dépendance ainsi défini par (5) a été désigné, à très juste titre, comme nous le montrerons plus loin, par J. P. Norton, coefficient de covariation; l'école française des statisticiens a consacré avec March cette dénomination, qui semble tout à fait justifiée.

IV — *Indice de dépendance basé sur l'ordre des éléments des séries.*

Avant de clore ce chapitre, signalons un indice de dépendance fondé sur la notion d'ordre des termes d'une série statistique.

Considérons par exemple les salaires moyens journaliers pour une catégorie

bien définie de travailleurs dans une série de départements en l'année A_0 et l'année A'_0, et cherchons à caractériser les séries correspondantes X et Y de salaires, afférentes respectivement à A_0 et A'_0; nous supposons que les salaires pour l'année A_0 ont été classés par ordre décroissant.

Ceci étant, si dans les départements 1, 2, 3,..... n, les salaires moyens journaliers pour les années A_0 et A'_0 ont été respectivement de :

$$S_1, S_2, S_3, \ldots\ldots S_n,$$
$$S'_1, S'_2, S'_3, \ldots\ldots S'_n,$$

on remarque que si les S_i (série x) forment une suite décroissante, les S_j sont tels que les départements 1, 2, 3..... n occupent respectivement au point de vue des salaires les places (α, β, γ,..... ν), (α..... ν) étant des nombres entiers de la série ($1, 2 \ldots\ldots n.$); ceci revient à dire que le département 1, qui occupait pour l'époque A_0 le premier rang dans l'échelle des salaires moyens journaliers, occupera pour l'époque A'_0 le α^{me} rang.

Dans le cas d'une dépendance directe bien accusée, les différences ($\alpha - 1$), ($\beta - 2$),..... ($\nu - n$) sont toutes nulles; plus les différences en question seront importantes, moins nette sera la dépendance.

La somme algébrique des différences étant toujours identiquement nulle, on est donc conduit à baser l'indice de dépendance sur une utilisation rationnelle des carrés des différences Δ.

En cas de *dépendance directe* marquée d'une manière précise, l'expression $\Sigma\Delta^2 = 0$; en cas de *dépendance inverse*, nettement accusée, l'on a :

$$\sum_{i=1}^{i=n} \Delta_i^2 = \sum_{h=1}^{h=n} [h - (n - h + 1)]^2 = n\,\frac{(n^2-1)}{3},$$

si n est le nombre des éléments de la série.

Si au contraire, les séries sont en *indépendance complète*, la somme des carrés des différences δ_i

$$\Sigma\,\delta_i^2 = \sum_{i=1}^{i=n} \frac{1}{n} \sum_{j=1}^{j=n} (i-j)^2 = \frac{n\,(n^2-1)}{6}$$

L'indice ρ, ainsi que le désigne Pearson, peut être défini par l'expression :

$$\rho = 1 - \frac{\Sigma\,\Delta^2}{\frac{n\,(n^2-1)}{6}}$$

qui, dans le cas d'*indépendance complète* des deux séries, prend la valeur zéro, car $\Sigma\Delta^2$ n'est autre que $\Sigma\delta_i^2$, et dans le cas de *dépendance complète directe a la valeur + 1*, puisque $\Sigma\Delta^2 = 0$; dans le cas de *dépendance complète inverse*, ρ est égal à — 1, car $\Sigma\Delta^2 = n\,\frac{(n^2-1)}{3}$.

CHAPITRE II

LIAISON STOCHASTIQUE ET DÉPENDANCE FONCTIONNELLE ENTRE GRANDEURS VARIABLES

I. *Liaison stochastique.* — Avec les probabilistes, nous appellerons variable accidentelle du k^{me} ordre une grandeur qui, avec des probabilités données, peut prendre k valeurs différentes, et désignerons l'ensemble des valeurs possibles

et des probabilités correspondantes sous le nom de loi de répartition de la variable accidentelle. Si par exemple, l'on considère une urne renfermant autant de boules blanches que de boules de couleur, et que l'on fasse 20 tirages en remettant chaque fois la boule tirée, on voit que le nombre des boules blanches dans une série de 20 boules est une variable accidentelle du 21^e ordre, attendu que le nombre des boules extraites peut prendre les valeurs 0, 1, 2..... 20; quant aux probabilités respectives à la sortie de 0, 1, 2..... 20 blanches en 20 tirages, elles peuvent être facilement valuées. En ce qui concerne la loi de répartition de la variable accidentelle elle varie avec la composition de l'urne.

A la notion de variable accidentelle, il nous faut maintenant rattacher la conception de *liaison stochastique*, qui se distingue nettement de la notion de dépendance fonctionnelle. Si $y = f(x)$, f définissant la dépendance fonctionnelle, on voit qu'à une valeur de x correspond une valeur de y ; à une valeur de y se rattachent une ou plusieurs valeurs de x, mais le choix de la solution naturelle pour x est fixé par des considérations d'un certain ordre.

On se trouve en présence d'une *liaison stochastique* entre x et y, lorsque x étant fixé, y apparaît comme une variable accidentelle qui peut être affectée de valeurs diverses, à chacune desquelles se rattache une probabilité donnée.

Considérons, à titre d'exemple, X le point tiré avec un dé blanc, Y la somme des points tirés avec un dé blanc et un dé rouge; X et Y ont entre eux une liaison stochastique, car pour une valeur donnée de X, Y peut prendre avec des probabilités toutes égales à $\frac{1}{6}$, les six valeurs différentes qui seront : soit 4, 5, 6, 7, 8, 9, soit 6, 7, 8, 9, 10, 11, suivant que le numéro tiré avec le dé blanc était égal à 3 ou à 5.

Remarquons enfin que si X est une variable accidentelle et si $Y = f(X)$, Y est également une variable accidentelle, à la condition de ne pas fixer la valeur de X.

De la distinction des notions de liaison stochastique et de dépendance fonctionnelle, découle à première vue la différence entre les recherches de corrélation statistique et les recherches dans le domaine des sciences naturelles.

II. — *D'une indication sur les notations employées aujourd'hui par bon nombre de statisticiens et de probabilistes.*

Avec Keynes (A Treatise on probability), considérons un corps de prémisses h et diverses propositions a, b, c,..... et désignons par $\bar{a}$ la proposition non a (a étant par exemple afférente à l'extraction d'une boule blanche d'une urne renfermant des boules blanches, noires, rouges, et a relative à l'extraction d'une boule différente des blanches).

Le symbole $(a + b)$ désigne la proposition a ou la proposition b, et le symbole $a\,b$ l'affirmation simultanée des deux propositions a et b.

Si l'on considère les deux prémisses h et les deux propositions (a, b), on voit immédiatement apparaître les probabilités.

$$ab \mid h,\ (a + b) \mid h,\ \ a \mid bh,\ b \mid ah,$$

dont on conçoit de suite la signification. Ainsi $b \mid ah$ signifie la probabilité de la

proposition b, lorsque l'on sait qu'en sus des prémisses h la proposition a est vraie.

Aux deux théorèmes fondamentaux des probabilités correspondent dans les conditions précitées les deux identités :

$$(a + b) \mid h = a \mid h + b \mid h - ab \mid h,$$
$$ab \mid h = a \mid bh \,.\, b \mid h = b \mid ah \,.\, a \mid h;$$

si la condition $b \mid ah = b \mid h$, ou la condition équivalente $a \mid bh = a \mid h$ se trouve réalisée, on dit que les deux probabilités $a \mid h$, $b \mid h$ *sont indépendantes*, et dans ces conditions, l'on a :

$$ab \mid h = a \mid h \,.\, b \mid h$$

III. — *Système de masses figurant la loi de dépendance de deux variables.*

Nous avons défini ci dessus une grandeur aléatoire X, et pouvons dès maintenant considérer deux grandeurs aléatoires d'ordres respectifs k et l.

Ceci étant, à une valeur déterminée x_m de X, on rattache la probabilité $p_m = x_m \mid h$; si maintenant, l'on fixe la valeur x_m et si l'on fait apparaître la grandeur Y, on remarque que la probabilité pour que Y prenne une valeur y_n est $y_n \mid x_m h$, puisque la valeur de X est connue.

Si la grandeur Y est indépendante de X, on a :

$$y_n \mid x_m h = y_n \mid h, \text{ et aussi } x_m \mid y_n h = x_m \mid h;$$

en ce cas, les grandeurs X et Y sont indépendantes et leurs probabilités le sont aussi.

Dans le cas général, la loi de répartition des deux grandeurs est définie par :

$$p_{mn} = y_n \mid x_m h \,.\, x_m \mid h = x_m \mid y_n h \,.\, y_n \mid h = x_m y_n \mid h ;$$

elle conduit à un système de kl points de masses p_{mn} avec $\Sigma p_{mn} = 1$, et permet une représentation à trois dimensions P (x_m, y_n, p_{mn}).

IV. *Loi de répartition liée. — Moments liés.*

On peut dire également qu'à un système de deux variables aléatoires (X, Y) pouvant prendre les valeurs :

$$X_1, X_2, \ldots . X_k,$$
$$Y_1, Y_2, \ldots . Y_l,$$

se rattache l'ensemble des combinaisons (X_i, Y_j) avec les probabilités $p_{i\,j}$, telles que $\Sigma p_{i\,j} = 1$.

La probabilité d'une valeur Y_j s'obtient en divisant la masse $p_{i\,j}$ par $p_{i\mid}$ et s'écrit $p^{(i)}{}_{\mid j}$; $p_{i\mid}$ représente la somme des masses situées sur la droite $X = X_i$.

Supposons que nous fixions pour X la valeur X_i, Y demeure une variable aléatoire ; l'ensemble des valeurs que peut prendre la variable Y et l'ensemble des probabilités qui lui sont afférentes constituent la *loi de répartition déterminée de Y pour* $X = X_i$, *ou répartition des Y liée à* X_i.

On reconnaît d'ailleurs facilement que l'on a :

$$p_{i\mid} = \sum_{j=1}^{j=l} p_{i\mid j},\ p_{\mid j} = \sum_{i=1}^{i=k} p_{i\,j},\ 1 = \Sigma\, p_{i\mid} = \sum_{j} p_{\mid j} = \sum_{i}\sum_{j} p_{i\,j}$$

$$1 = \sum_{i} p^{(j)}_{i\mid} = \sum_{j} p^{(i)}_{\mid j}.$$

A l'ensemble des éléments de cette distribution spéciale ou répartition liée, on rattache son espérance mathématique, ses moments successifs ou moments liés, puis les écarts liés d'ordre k.

Ligne de régression. — Équation scédastique.

On définit ainsi $E^{(i)}Y = \Sigma p_{|j}^{(i)} Y_j$ l'espérance mathématique d'ordre 1 qui — en général — sera une fonction de l'abscisse X_i, $E^{(i)} Y = F(X_i)$; l'expression analytique de cette fonction sera l'équation de régression de Y par rapport à X, si l'on se conforme à la terminologie de Galton, puis de Pearson.

L'ensemble des points $G_i \{E^{(i)}Y, X_i\}$ afférents aux différentes valeurs de X_i, constitue la *ligne de régression de Y en X*, à laquelle on pourra associer — par un processus analogue à celui qui vient d'être exposé — la ligne de régression en Y.

A côté de l'espérance mathématique de Y, il y a lieu de faire apparaître l'écart moyen quadratique de la dispersion Y pour $X = X_i$, c'est à dire ce que l'on désigne avec Pearson par $\sigma^{(i)}(Y)$ ou l'*écart type lié*, et l'on a ainsi : $\sigma^{(i)}(Y) = \varphi(X_i)$ qui n'est autre chose que l'équation scédastique.

Si l'écart quadratique moyen de Y est constant pour toutes les valeurs de X, la liaison entre X et Y est dite *homoscédastique;* dans le contraire, elle est hétéroscédastique. On ferait de même intervenir l'écart $\sigma^{(j)}(X) = \psi(Y_j)$.

III. — *De l'indépendance stochastique et de la réciprocité.*

Si la loi de répartition déterminée de Y ne change pas, quelle que soit la valeur donnée à X, on dit que Y est *stochastiquement* indépendant de X, ce qui revient à dire que la loi de répartition déterminée de X reste la même quelle que soit la valeur de Y ou encore que l'indépendance des deux variables est alors réciproque.

Désignons par $p_{i|}$ la probabilité pour que X prenne l'une de ses k valeurs possibles X_i, par $p_{|j}$ la probabilité pour que Y prenne une de ses l valeurs possibles Y_j et enfin par p_{ij} la probabilité pour que X et Y prennent simultanément les valeurs respectives X_i et Y_j, et représentons par $p_{|j}^{(i)}$ la *probabilité déterminée* pour que Y prenne la valeur Y_j, lorsque X a reçu la valeur X_i, et par $p_{i|}^{(j)}$ la *probabilité déterminée* pour que X prenne la valeur X_i, lorsque Y a reçu la valeur Y_j.

La probabilité d'apparition simultanée de deux événements non indépendants étant égale au produit de la probabilité de l'un par la probabilité *déterminée* de l'autre, on a les deux relations :

$$p_{ij} = p_{i|}\, p_{|j}^{(i)} = p_{|j}\, p_{i|}^{(j)}$$

Or la variable Y est indépendante de X, lorsque la loi de répartition déterminée de Y reste la même pour toutes les valeurs de X; cela revient à dire que :

$$p_{|j}^{(i)} = p_{|j} \text{ pour } i = 1,2 \ldots\ldots k \text{ et } j = 1, 2 \ldots\ldots l$$

Inversement si $p_{|j}^{(i)} = p_{|j}$ pour toutes les valeurs possibles de (i, j), la loi de répartition déterminée de Y par rapport à X ne change pas avec X, et Y est indépendant de X.

En effet, si dans la relation $p_{i|}\, p_{|j}^{(i)} = p_{|j}\, p_{i|}^{(j)}$, nous faisons $p_{|j}^{(i)} = p_{|j}$, nous avons $p_{|i}\, p_{|j} = p_{|j}\, p_{i|}^{(j)}$ d'où $p_{i|}^{(j)} = p_{i|}$ pour toutes les valeurs de i et de j.

L'indépendance de Y par rapport à X entraîne celle de X par rapport à Y. En cas d'indépendance réciproque des variables on a : $p_{i|j} = p_{i|}\, p_{|j}$ pour tous les systèmes de valeurs de i et de j ; inversement si l'on a $p_{i|j} = p_{i|}\, p_{|j}$ pour tous les systèmes de valeurs de i et de j, on a : $p^{(i)}_{|j} = p_{|j}$ et les variations X et Y sont indépendantes.

A côté de cette notion, il est intéressant de signaler la notion dite par Pearson de *non corrélativité.*

Pearson dit que la variable Y est en corrélation avec X, si l'espérance mathématique déterminée de Y varie avec X ; si au contraire, l'espérance mathématique déterminée de Y reste constante pour toutes les valeurs de X, Y est sans corrélation avec X.

La non corrélativité de Y avec X s'exprime par le fait que la ligne de régression de Y par rapport à X représentative de $E^{(x)}(Y)$ est une parallèle à OX.

Rappelons qu'il ne faut pas confondre la notion de l'indépendance stochastique définie ci dessus et celle de la non corrélativité.

Toutes les fois que la variable Y est stochastiquement indépendante de X, elle ne peut pas être en corrélation avec X, au sens de la définition de Pearson ; par contre, si Y est sans corrélation avec X, il n'en résulte pas que Y est stochastiquement indépendant de X. En effet, l'espérance mathématique déterminée de Y étant constante quelle que soit la valeur de X_i, l'écart moyen quadratique déterminé peut, lui, varier avec X_i.

A l'*indépendance stochastique* qui entraîne la réciprocité, la non corrélativité de Y avec X n'a point comme conséquence la non corrélativité de X avec Y.

La ligne de régression de Y avec X peut être une droite parallèle à OX. sans que la ligne de régression de X par rapport à Y soit une parallèle à OY.

VI. — *Dépendance stochastique de deux véritables continues.*

Jusqu'ici, nous ne nous sommes préoccupé que de distributions discontinues $p_{i\,j}$; à ces distributions sur le plan des xy, substituons une distribution continue de densité $f(x, y)$.

La probabilité pour qu'un point M de la distribution soit intérieur à l'aire S est donnée par l'expression :

$$\iint_{(S)} f(x, y)\, dx\, dy.$$

Il est évident que la probabilité pour que le point M soit situé dans la bande $(x_i, x_i + dx)$ est :

$$dx \int f(x_i, y)\, dy = \varphi(x)\, dx,$$

et la probabilité pour qu'il appartienne à la région $(x_i, x_i + dx)$ $(y_j, y_j + dy)$ a pour valeur :

$$\frac{f(x_i, y_j)\, dx\, dy}{\varphi(x_i)\, dx} = \frac{f(x_i, y_j)\, dy}{\varphi(x_i)},$$

expression définissant la probabilité liée.

On peut aussi donner l'expression des moments liés :

$$E^{(y)}(y) = \frac{\int f(x_i, y)\, y\, dy}{\int f(x_i, y)\, dy} = F(x_i),$$

$$E^{(x)}(x) = \frac{\int f(x_i, y_j)\, x\, dx}{\int f(x_i, y_j)\, dx} = \Phi(y_j)$$

$$E^{(y)}(y^k) = \frac{\int f(x_i, y)\, y^k\, dy}{\int f(x_i, y)\, dy}.$$

Les équations $E^{i}(y) = F(x_i)$ et $E^{(j)}(x) = \Phi(y_j)$ *donnent les lignes de régression.*

VII. — *Application au cas de la loi de Laplace Gauss.*

La loi de probabilité étant représentée par l'équation :

$$f(r, y) = \frac{H}{2\pi} e^{-\frac{1}{2}(ax^2 + 2bxy + cy^2)}$$

on peut dire que la densité des masses reste constante tout le long des coniques $ax^2 + 2bxy + cy^2 = 2k$, que nous supposons être des ellipses susceptibles de couvrir tout le plan.

Il est évident que $\iint f(x, y)dxdy$ étendue à tout le plan a pour valeur 1; le calcul de cette intégrale double nous fournit facilement la valeur de H, en rapportant les coniques à leurs axes, et par suite, la densité de probabilité

$$f(x, y) = \frac{\sqrt{ac - b^2}}{2\pi} e^{-\frac{1}{2}(ax^2 + 2bxy + cy^2)}$$

Il s'en suit que la fonction $\varphi(x_i)$ est définie par :

$$\varphi(x_i) = \int_{-\infty}^{+\infty} f(x_i, y)\, dy = \sqrt{\frac{ac - b^2}{2\pi}}\, e^{-\frac{ac\, b^2}{2c} x_i^2} \int_{-\infty}^{+\infty} e^{-\frac{c}{2}\left(y + \frac{b}{c} x_i\right)^2} dy,$$

$$\varphi(x_i) = \frac{1}{\sqrt{2\pi}\, \sigma_x} \cdot e^{-\frac{x_i^2}{2\sigma_x^2}}, \quad \left(\text{avec } \sigma_x = \sqrt{\frac{c}{ac - b^2}}\right),$$

qui correspond à une distribution normale.

On trouverait de même :

$$\varphi(y_j) = \frac{1}{\sqrt{2\pi}\, \sigma_y} e^{-\frac{y_j^2}{2\sigma_y^2}}, \quad \left(\text{avec } \sigma_y = \sqrt{\frac{a}{ac - b^2}}\right).$$

La distribution de la variable y dans la bande $(x_i, x_i, + dx)$ est représentée par l'équation :

$$\frac{f(x_i, y)}{\varphi(x_i)} dy = \frac{\sqrt{c}}{\sqrt{2\pi}} e^{-\frac{c}{2}\left(y + \frac{b}{c} x_i\right)^2} dy ;$$

elle met en évidence une distribution normale caractérisée par l'écart quadratique $\frac{1}{\sqrt{c}}$, la moyenne étant donnée par $y + \frac{b}{c} x_i = 0$. Il résulte de là que l'on se trouve bien en présence d'une régression linéaire de y en x, à laquelle il y a lieu d'adjoindre une droite de régression de x en y. On peut aussi se rendre compte de la distribution des masses, en constatant que les droites de régression ne sont autres que les diamètres conjugués des directions oc et oy.

Si maintenant l'on adopte comme unités de mesure σ_x et σ_y, et si l'on pose $\left(-\frac{b}{c} = r\right)$, on remarque que la loi de probabilité prend la forme

$$f(X, Y) = \frac{1}{2\pi\sqrt{1-r^2}} e^{-\frac{1}{2(1-r^2)}(X^2 - 2rXY + Y^2)}$$

dite de corrélation normale de Gauss.

Eu égard aux équations respectives des droites de régression, $y = rx_i$, $x = ry_j$, on voit de suite que la *condition d'indépendance répond à* $r = 0$; dans ce dernier cas ($r = 0$), la loi de probabilité prend la forme simple

$$f(X, y) = \frac{1}{2\pi} e^{-\frac{X^2+Y^2}{2}}$$

qui peut s'écrire :

$$\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{X^2}{2}}\right)\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{Y^2}{2}}\right)$$

Nous démontrerons ultérieurement que le coefficient r ne peut osciller qu'entre — 1 et + 1.

VIII. — *Des procédés destinés à la détermination et à la représentation de la liaison. — Leur examen critique.*

Nous avons précédemment signalé la différence très nette entre la liaison stochastique et la dépendance fonctionnelle; il nous faut maintenant procéder à la recherche et à l'examen des modes de détermination et de représentation de liaison stochastique en faisant apparaître leurs caractéristiques, en nous attachant tout spécialement aux modes correspondant à l'emploi de deux variables accidentelles.

Étant donné les deux variables accidentelles X et Y, pouvant prendre les valeurs :

$$X_1, X_2, \ldots\ldots X_k,$$
$$Y_1, Y_2, \ldots\ldots Y_l,$$

auxquelles se rattachent les probabilités $p_{i|}$ que X prenne la valeur X_i, $p_{|j}$ que Y prenne la valeur Y_j, $p_{i|j}$, que X et Y prennent simultanément les valeurs (X_i et Y_j), et les relations suivantes :

$$(1)\quad p_{i|} = \sum_{j=1}^{j=l} p_{i|j},\ p_{|j} = \sum_{i=1}^{i=k} p_{i|j},\ 1 = \sum_{i} p_{i|} = \sum_{j} p_{|j} = \sum_{i}\sum_{j} p_{i|j},\ 1 = \sum_{i} p_{i|}^{(j)} = \sum_{j} p_{|j}^{(i)}$$

On a d'ailleurs démontré que si $p_{i|j} = p_{i|} \cdot p_{|j}$ pour toutes les valeurs de i et de j, il y a *indépendance réciproque* entre les deux variables, et l'on a de plus fait remarquer que la condition d'indépendance se trouve encore mise en évidence par l'une ou l'autre des égalités

$$p_{|j}^{(i)} = p_{|j},\ p_{i|}^{(j)} = p_{i|}$$

Première mesure de la dépendance, ou carré moyen de contingence de Pearson.

S'il y a indépendance réciproque des variables, toutes les différences ($p_{i|j} - p_{i|} \cdot p_{j}$) sont nulles. Si l'une quelconque de ces différences ou plusieurs d'entres elles sont différentes de zéro, les variables ne sont plus indépendantes. On remarque que la liaison entre X et Y sera d'autant plus accusée que les différences susvisées seront plus grandes, et l'on voit que l'on peut, grâce à cette remarque, établir un critère de dépendance; c'est ce qu'a fait Pearson en introduisant le *man square contingency* (ou *carré moyen de contingence*), dans le cas de deux variables

$$\varphi^2 = \sum_i \sum_j \frac{(p_{i|j} - p_{i|} \cdot p_{|j})^2}{p_{i|} \cdot p_{|j}} \tag{2}$$

$\varphi = 0$, si les variables sont indépendantes, car $p_{i\,j} = p_{i|} \cdot p_{|j}$; inversement la quantité φ ne peut être nulle que si tous les carrés constitutifs de φ^2 sont nuls, et alors les variables sont indépendantes.

Cas de liaison fonctionnelle. — En l'occurrence, à une valeur X_i de X correspond une valeur unique (Y_i) de Y, et aux k couples (X_i, Y_i) sont affectées des masses $p_{i\,i} = p_{i|} = p_{|i}$, masses que nous pourrons encore désigner par $p_1, p_2, \ldots\ldots p_k$; quant aux combinaisons correspondant aux points X_i, Y_j), il ne s'y rattache que des masses nulles.

Si maintenant l'on forme avec Tschuprow le carré moyen de contingence afférent à l'association (X_i, Y_i), on a :

$$\frac{(p_i - p_{i|}\, p_{|i})^2}{p_{i|}^2} = (1 - p_i)^2$$

Pour toutes les associations (X_i, Y_j) avec ($j \neq i$), on a l'élément correspondant du carré moyen de contingence.

$$\sum_{i=1}^{i-1} p_i \,|\, p \,|_j + \sum_{j=i+1}^{k} p_i \,|\, p \,|_j = p_i(p_1 + p_2 + \ldots\ldots + p_{i-1} + p_{i+1} \ldots\ldots + p_k) = p_i(1 - p_i)\cdot$$

Comme pour une valeur de i caractérisant X_i, apparaît dans le carré moyen de contingence l'expression $(1 - p_i)^2 + p_i(1 - p_i) = 1 - p_i$, et puisqu'il faut envisager les k valeurs de X (X_1, X_2, $\ldots\ldots$ X_k), il s'ensuit que l'on est conduit à la somme :

$$\sum_{i=1}^{i=k} (1 - p_i) = k - \Sigma\, p_i = k - 1$$

Or k et l étant les ordres de X et de Y, on voit que l'on peut fournir un indice de la mesure de la dépendance en considérant par la suite l'expression $\dfrac{\varphi^2}{\sqrt{(k-1)\,(l-1)}}$, *qui sera nulle s'il y a indépendance absolue, et égale à 1 s'il y a liaison fonctionnelle.*

Nous venons de voir qu'au coefficient dit carré moyen de contingence, Tschuprow a substitué :

$$\tau^2 = \frac{\varphi^2}{\sqrt{k-1)\,(l-1)}},$$

qui est inférieur à l'unité, et qui n'atteint la valeur 1, qu'en cas de dépendance complète, comme l'a démontré G. Hössjer. De la relation :

$$\varphi^2 = \Sigma\Sigma \frac{p_{ij}^2}{p_{i|} \cdot p_{|j}} - 1,$$

l'on déduit la suivante :

$$\varphi^2 + 1 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{p_{ij}}{p_{i|}} \frac{p_{ij}}{p_{|j}}$$

S'il n'y a pas dépendance complète, on a au moins pour certaines valeurs de i et de j :

$$0 < p_{ij} < p_{|j}$$

d'où :

$$\varphi^2 + 1 < \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{p_{ij}}{p_{i|}} = \sum_{i=1}^{k} \frac{p_{i|}}{p_{i|}} = k$$

et, par suite :

$$\varphi < \sqrt{k-1}.$$

On obtiendrait de la même manière :

$$\varphi < \sqrt{l-1}, \text{ d'où } \varphi^2 < \sqrt{(k-1)(l-1)}, \text{ et enfin } \tau^2 < 1$$

M. Steffensen, dans ses conférences à l'Institut Henri-Poincaré en 1933, introduit la contingence moyenne et écrit :

$$\varphi^2 = \Sigma \Sigma (p_{|j}^{(i)} - p_{|j})(p_{i|}^{(j)} - p_{i|});$$

en effet, l'on remarque qu'il y a identité entre les deux expressions de φ^2, car la première a pour valeur :

$$\Sigma\Sigma \frac{p_{ij}^2}{p_{i|}\, p_{|j}} - 1$$

et la seconde :

$$\Sigma\Sigma (p_{|j}^{(i)}\, p_{i|}^{(j)} - p_{i|}\, p_{|j}^{(i)} - p_{|j}\, p_{i|}^{(j)} + p_{i|}\, p_{|j}).$$

Soit :

$$\Sigma\Sigma\, p_{|j}^{(i)}\, p_{i|}^{(j)} - \sum_{i=1}^{k} p_{i|} \sum_{j=1}^{l} p_{|j} - \sum_{i=1}^{k} p_{i|}^{(j)} + \Sigma\Sigma\, p_{i|}\, p_{|j},$$

ou, en vertu de relations classiques :

$$\Sigma\Sigma\, p_{|j}^{(i)}\, p_{i|}^{(j)} - 1 = \Sigma\Sigma \frac{p_{|j}^{(i)}\, p_{i|}^{(j)} - p_{i|}\, p_{|j}}{p_{i|}\, p_{|j}} - 1 = \Sigma\Sigma \frac{p_{ij}^2}{p_{i|}\, p_{|j}} - 1.$$

Il est utile de faire observer que l'on a supposé jusqu'ici k et l finis; dans l'hypothèse où ces nombres tendraient vers l'infini, on a pour τ^2 l'expression :

$$\tau^2 = \lim_{\text{pour} \begin{cases} k = \infty \\ l = \infty \end{cases}} \frac{1}{\sqrt{(k-1)(l-1)}} \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(p_{ij} - p_{i|}\, p_{|j})^2}{p_{i|}\, p_{|j}}$$

pour laquelle il faudrait montrer l'existence de la limite.

IX. — *D'une mesure nouvelle de la dépendance.*

M. Steffensen revenant en 1934 (Voir *Biometrika*, vol. XXVI, mai 1934 : Note sur certaines mesures de dépendance entre variables statistiques) sur une mesure de la dépendance qu'il avait préconisée en 1933 dans ses conférences à l'Institut Henri Poincaré, a donné une nouvelle mesure de la dépendance présentant certains avantages sur l'ancienne, et ne prêtant plus comme la précédente le flanc à une critique justifiée de Me Pollaczek.

ψ, mesure de la dépendance, est définie par la relation :

$$\psi^2 = \sum_{i,j} p_{i|j}\,\Phi_{ij}^2 \text{ avec } \Phi_{ij}^2 = \frac{(p_{ij} - p_{i|}p_{j|})^2}{p_{i|}\,(1-p_{i|})\,p_{|j}\,(1-p_{|j})} \qquad (1)\quad (2)$$

et jouit des propriétés suivantes :

I. ψ^2 est toujours comprise entre 0 et 1 ;

II. ψ^2 *s'annule si les variables sont complètement indépendantes, et seulement dans ce cas ;*

III. ψ^2 *prend la valeur 1 dans le cas de dépendance complète, et seulement dans ce cas.*

De l'examen de la formule (1), il sésulte que ψ^2 est une moyenne arithmétique et que la démonstration de la propriété I est par suite la conséquence de ce que $\Phi_{ij}^2 \leqslant 1$, pour toutes les valeurs de i et de j.

Nous considérerons deux cas, suivant que la probabilité $p_{i|j}$ est plus grande ou plus petite que $p_{i|} \cdot p_{|j}$.

Dans le premier cas, afférent à $p_{i|j} \geqslant p_{i|}p_{|j}$, nous écrirons Φ_{ij}^2 sous la forme :

$$\Phi_{ij}^2 = \frac{p_{i|j} - p_{i|}\,p_{|j}}{p_{i|}\,(1-p_{i|})} \cdot \frac{p_{i|j} - p_{i|}\,p_{|j}}{(p_{|j}\,(1-p_{|j})} ; \qquad (3)$$

Comme $p_{i|j} \leqslant p_{i|}$, il s'en suit que $p_{i|j} - p_{i|}\,p_{|j} \leqslant p_{i|}\,(1-p_{|j})$, et, par suite, que le premier facteur dans le deuxième membre de l'équation (3) ne peut dépasser l'unité. D'autre part $p_{i\,j}$ étant plus petite ou au plus égale à $p_{|j}$, on voit que l'expression $(p_{i|j} - p_{i|}\,p_{|j})$ est plus petite ou au plus égale à $p_{|j}$, et aussi que le second facteur du deuxième membre de (3) ne peut également pas dépasser l'unité.

Au cas $p_{i|j} \geqslant p_{i|}\,p_{|j}$, correspond donc l'inégalité $\Phi_{ij}^2 \leqslant 1$.

Examinons maintenant le deuxième cas, relatif à l'inégalité $p_{i|j} \leqslant p_{i|}\,p_{|j}$; nous écrivons alors Φ_{ij}^2 ainsi qu'il suit :

$$\Phi_{ij}^2 = \frac{p_{i|}\,p_{|j} - p_{i|j}}{p_{i|}\,p_{|j}} \times \frac{p_{i|}\,p_{|j} - p_{i|j}}{(1-p_{i|})\,(1-p_{|j})} . \qquad (4)$$

et nous remarquons que le premier facteur du deuxième membre ne peut dépasser l'unité.

Comme l'on a $\Sigma p_{i|} - \Sigma p_{i|j} = 1 - p_{|j}$, il en résulte $p_{i|} - p_{i|j} \leqslant 1 - p_{|j}$ et aussi, en recourant au même processus, $p_{|j} - p_{i|j} \leqslant 1 - p_{i|}$; du produit de ces deux dernières inégalités, on déduit :

$$(p_{i|} - p_{i|j})\,(p_{|j} - p_{i|j}) = p_{i|}\,p_{|j} - p_{i|j}\,(p_{i|} + p_{|j} - p_{i|j}) < (1-p_{i|})\,(1-p_{|j}). \qquad (5).$$

On remarque aussi par addition des deux inégalités que :

$$p_{i|} + p_{|j} - 2\,p_{i|j} \leqslant 2 - p_{i|} - p_{|j}, \text{ et par suite } p_{i|} + p_{|j} - p_{i|j} \leqslant 1$$

Dans ces conditions de l'inégalité (5)l on déduit :

$$p_{i|}\,p_{|j} - p_{i|j}\,(p_{i|} + p_{|j} - p_{i|j}) < p_{i|}\,p_{|j} - p_{i|j} < (1 - p_{i|})\,(1 - p_{|j})$$

et l'on remarque enfin que le second facteur du deuxième membre de (4) est inférieur à l'unité, et par suite, que $\Phi_{ij}^2 \leqslant 1$, quel que soit le cas envisagé.

Démonstration de la propriété II. — On constate immédiatement que ψ^2 s'annule si $p_{i'j} = p_{i|}\,.\,p_{|j}$, pour toutes les valeurs de i et de j, c'est-à-dire si les variables sont indépendantes.

Réciproquement, si ψ^2 s'annule, la relation $p_{i\,j} = p_{i|}\,p_{|j}$ a lieu pour toutes les valeurs de i et de j. La démonstration de la réciproque est basée sur ce que si $\psi^2 = 0$, aucune des probabilités $p_{i'j}$ ne peut s'annuler. Supposons pour un instant que $p_{r|s} = 0$, et montrons qu'une telle hypothèse conduit à une contradiction.

Nous avons, en effet :

$\sum_{ij} (p_{i|j} - p_{i|}\,p_{|j}) = 0$, et comme le terme $(p_{r|s} - p_{r|}\,p_{|s})$ qui se réduit à $(-p_{r|}\,p_{|s})$ est négatif, il doit donc y avoir au moins un terme positif dans la somme Σ; supposons que ce soit $(p_{m'n} - p_{m|}\,p_{|n})$. Dans ce cas, Φ_{mn}^2 ne s'annule pas, et ψ^2 contient le terme positif $(p_{m|n}\,\Phi_{mn}^2)$ et est donc positive. Or nous avons supposé ψ^2 nulle; nous sommes ainsi amenés à une contradiction, et pour la lever, il faut qu'aucune des probabilités $p_{i|j}$ ne puisse s'annuler si $\psi^2 = 0$, et par suite que toute expression Φ_v' s'annule si $\psi^2 = 0$, c'est-à dire que l'on a $p_{i|j} = p_{j|}\,p_{|j}$, pour toutes les valeurs de i et de j.

Démonstration de la propriété III. — Nous allons tout d'abord montrer que si la dépendance est complète $\psi^2 = 1$.

En effet, dans ce cas, l'on a :

$$p_{i'j} = p_{i|} = p_{|i}, \text{ et } p_{i|j} = 0 \text{ pour } i \neq j,$$

et la relation (1) s'écrit :

$$\sum_{i} p_{i|i}\,\frac{(p_{i|i} - p_{i|i}^2)^2}{p_{i|i}\,(1 - p_{i|i})\;p_{i|i}\,(1 - p_{i|i})} = \sum_{i} p_{i|i} = 1$$

Réciproquement, supposons que $\psi^2 = 1$, ou que $\sum_{ij} p_{i\,j}\,\Phi_v^2 = 1$ (1)'.

Comme $\sum_{ij} p_{i|j} = 1$, et que $\Phi_v^2 \leqslant 1$, la relation (1)' ne peut être vérifiée que si $\Phi_v^2 = 1$, pour toutes les valeurs de i et de j pour lesquelles $p_{i|j}$ ne s'annule pas. Or si $p_{i|j} > 0$, nous ne pouvons avoir en même temps $p_{i'j} \leqslant p_{i}\,p_{|j}$ et $\Phi_v^2 = 1$, car le premier facteur dans la relation (4) est inférieur à l'unité.

Nous avons donc $p_{i|j} > p_{i|}\,p_{|j}$, et comme chacun des facteurs dans la relation (3) doit être égal à l'unité, si $\Phi_v' = 1$, nous en déduisons que $p_{i|j} = p_{i|} = p_{|j}$, c'est à dire que y_j arrive nécessairement si x_i arrive.

Il résulte de là que la condition $\psi^2 = 1$ entraîne la dépendance complète entre les variables.

Dans le domaine de l'application statistique, on substitue aux probabilités les fréquences relatives correspondantes :

Si l'on désigne comme primitivement par $n_{i|j}$ la fréquence absolue de la combinaison (x_i, y_j), on voit que $n_{i|} = \sum_{j} n_{i|j}$, $n_{|j} = \sum_{i} n_{i|j}$ et que $N = \sum_{i} n_{i|} = \sum n_{|j}$ est le nombre total des observations.

La grandeur Ψ, expression approchée de ψ, se déduit de la relation :

$$\Psi^2 = \sum_i \sum_j \frac{n_{i|j}}{N} \frac{(N n_{i|j} - n_{i|}\, n_{|j})^2}{n_{i|}\,(N - n_{i|})\, n_{|j}\,(N - n_{|})_j} \tag{5}$$

Steffensen, après avoir remarqué que le calcul de Ψ^2 est relativement laborieux, a été amené à établir une autre mesure ω de la dépendance introduisant des calculs plus simples qui est définie par la relation :

$$\omega = \frac{\Sigma \,|\, p_{i|j} - p_{i|}\, p_{|j} \,|}{\bar{\Sigma}\,(p_{i|j} - p^2_{i|j}) + \underline{\Sigma}\, p_{i|}\, p_{|j}}, \tag{6}$$

où Σ représente une sommation double par rapport à tous les i et j, $\bar{\Sigma}$ une sommation par rapport à tous les i et j pour lesquelles $p_{i|j} > p_{i}\ p_{|j}$ et $\underline{\Sigma}$ une sommation afférente aux indices i et j pour lesquels $p_{i|j} \leqslant p_{i|}\, p_{|j}$.

L'expression de ω jouit des mêmes propriétés fondamentales que ψ.

X. — *Valeur particulière du carré moyen de contingences de Pearson, dans le cas où X et Y ne peuvent prendre chacune que deux valeurs.*

En l'occurrence, l'on compte quatre probabilités $p_{i|j}$:

$$p_{1|1},\ p_{1|2},\ p_{2|1},\ p_{2|2},$$

telles que l'on a :

$$p_{1|1} + p_{1|2} = p_{1|},\ p_{2|1} + p_{2|2} = p_{2|},\ p_{1|} + p_{2|} = 1,$$
$$p_{1|1} + p_{2|1} = p_{|1},\ p_{1|2} + p_{2|2} = p_{|2},\ p_{|1} + p_{|2} = 1,$$

et l'on constate que :

$$\varphi^2 = \sum_i \sum_j \frac{(p_{i|j} - p_{i|}\, p_{|j})^2}{p_{i|}\, p_{|j}} \text{ a pour valeur } \frac{\delta^2}{p_{1|}\, p_{2|}\, p_{|1}\, p_{|2}}$$
$$\text{avec } \delta = p_{1|1} - p_{1|}\, p_{|1}$$

Les indices basés sur l'introduction de δ et de φ^2 ne font intervenir que les probabilités des valeurs possibles de X et de Y, ainsi que les combinaisons de ces probabilités, et *non les valeurs mêmes de X et de Y*. En raison de cette propriété particulière, cette catégorie d'indices convient bien à l'examen des cas où les variables étudiées ne peuvent, que sous certaines restrictions, être considérées comme des variables accidentelles liées d'une manière stochastique (1).

XI. — *Moments liés et coefficient de corrélation.*

Grâce à l'idée d'espérance mathématique et de moments se rattachant à l'étude d'une série statistique simple, introduite dans les séries à deux variables, on peut faire apparaître facilement les indices caractéristiques de la loi de dépendance entre X et Y.

Désignons par $m_{f|g}$ l'espérance mathématique de $x^f y^g$, ou *moment d'ordre* $(f + g)$.

---

(1) Voir Tschuprow : Chapitre IV : *Das a priorische Agbhängigkeitsgesetz. Grunbegriffe und Grundprobleme der Korrelations Théorie*, p. 42.

$$m_{f|g} = \mathrm{E}x^f y^g = \sum_i \sum_j p_{i|j}\, x_i^f\, y_j^g.$$

Les quantités $m_{1|0} = \sum p_{i|}\, x_i = x_0$, $m_{0|1} = \sum p_{|j}\, y_j = y_0$ ne sont autres que les coordonnées du centre de gravité de la distribution représentative des masses dans le plan des xy,

Si maintenant l'on prend les moments des divers ordres, non plus par rapport à une origine quelconque, mais par rapport au point (x_0, y_0), on évalue ce que l'on appelle les *écarts moyens.*

$$\mu_{f|g} = \mathrm{E}\,(x - m_{1|0})^f (y - m_{0|1})^g;$$

c'est ainsi qu'à l'ordre 2, correspondent les quantités $\mu_{2|0}$, $\mu_{1|1}$, $\mu_{0\,2}$, dont la première et la troisième sont respectivement les carrés des écarts moyens quadratiques σ_x et σ_y.

Au moment $\mu_{f|g}$ d'ordre $(f + g)$, on rattache le moment d'ordre 0 défini par le rapport :

$$r_{f|g} = \frac{\mu_{f|g}}{\mu_{2|0}^{\frac{f}{2}}\, \mu_{0|2}^{\frac{g}{2}}}$$

Le coefficient particulier $r_{1|1} = \dfrac{\mu_{1|1}}{\sigma_x\, \sigma_y}$ est appelé *coefficient de corrélation.* Il est évident qu'à une loi de répartition bien définie, se rattachent toute une série de moments $\mu_{f\ g}$ et de coefficients $r_{f\ g}$; ils caractérisent la *loi de répartition réduite.*

Limites de variation du coefficient de corrélation.

Eu égard aux relations :

$$\mathrm{E}\left(\frac{x - m_{1|0}}{\sigma_x}\right)^2 = \mathrm{E}\left(\frac{y - m_{0\,1}}{\sigma_y}\right)^2 = 1$$

$$\mathrm{E}\left[\frac{(x - m_{1|0})\,(y - m_{0|1})}{\sigma_x\, \sigma_y}\right] = \frac{\mu_{1\,1}}{\sigma_x\, \sigma_y} = r_{1\,1},$$

il résulte que :

$$\mathrm{E}\left(\frac{x - m_{1|0}}{\sigma_x} - r_{1|1}\,\frac{y - m_{0|1}}{\sigma_y}\right)^2 = 1 - r_{1|1}^2$$

d'où l'on déduit $r_{1|1}^2 \leqslant 1$.

XII. — *Dispersions non déterminées et déterminées. Conditions d'indépendance des variables.*

A l'espérance mathématique « déterminée » $m_{|1}^{(i)}$ de Y pour $\mathrm{X} = \mathrm{X}_i$, associons la *dispersion déterminée* $\mu_{|2}^{(i)}$, ainsi que la dispersion déterminée moyenne $\sum_i p_{i|}\, \mu_{|2}^{(i)}$, et la *dispersion non déterminée* $\mu_{0|2}$.

Aux paramètres $m_{|1}^{(i)} = \sum_j p_{|j}^{(i)}\, y_j$, $\mu_{|2}^{(i)} = \sum_j p_{|j}^{(i)}\, [y_j - m_{|1}^{(i)}]^2$, nous adjoignons les deux autres $m_{1|}^{(j)}$ et $\mu_{2|}^{(j)}$, résultant de la substitution des x_i et aux y_j, et nous remarquons que

$$\sum_i p_{i|}\, m_{|1}^{(i)} = \sum_i \sum_j p_{i|} \cdot p_{|j}^{(i)}\, y_j = \sum\sum p_{i|j}\, y_j = m_{0|1},$$

quantité qui n'est autre que :

$$\sum_j p_{|j}\, y_j .$$

Faisant état de la valeur de $\mu^{(i)}_{|2}$, on détermine facilement celle de la dispersion déterminée moyenne $\sum_i p_{i|}\, \mu^{(i)}_{|2}$; en effet, l'on a :

$$\sum_i p_{i|}\, \mu^{(i)}_{|2} = \sum_i \sum_j p_{i|}\, p^{(i)}_{|j}\, [y_j - m^{(i)}_{|1}]^2 = \sum_i \sum_j p_{i|j}\, [(y_j - m_{0|1}) - (m^{(i)}_{|1} - m_{0|1})]^2$$

Cette somme double se calcule immédiatement d'une manière rapide, en rappelant que :

$$\sum_i \sum_j p_{i|j}\, (y_j - m_{0|1})\, (m_{0|1} - m^{(i)}_{|1}) = \sum_i p_{i|}\, [m^{(i)}_{|1} - m_{0|1}]$$

il en résulte que :

$$\sum_i p_{i|}\, \mu^{(i)}_{|2} = \sum_i \sum_j p_{i|j}\, (y_j - m_{0|1})^2 - \sum_i p_{i|}\, [m^{(i)}_{|1} - m_{0|1}]^2,$$

relation qui peut encore s'écrire :

$$\sum_i p_{i|}\, \mu^{(i)}_{|2} = \mu_{0|2} - \sum_i p_{i|}\, [m^{(i)}_{|1} - m_{0|1}]^2.$$

On trouverait de même :

$$\sum_j p_{|j}\, \mu^{(j)}_{2|} = \mu_{2|0} - \sum_j p_{|j}\, [m^{(j)}_{|1} - m_{1|0}]^2.$$

Des valeurs de $m_{f\,g}$, $m_{f\,o}$, $m_{o\,g}$ et du produit

$$m_{f\,0}\, m_{0\,g} = \sum_i \sum_j p_{i|}\, p_{j|}\, x_i^f\, y_j^g .$$

On déduit la relation

$$(5) \qquad m_{f|g} - m_{f|0}\, m_{0|g} = \sum_i \sum_j (p_{i|j} - p_{i|}\, p_{|j})\, x_i^f y_j^g,$$

dont l'examen conduit à une remarque importante; en effet, si le premier membre est nul, il s'en suit que la somme double du second membre est aussi nulle, et cela pour des valeurs de x et de y positives, ce qui exige que $p_{i|j} - p_{i|} \cdot p_{|j} = 0$, c'est à dire que les variables soient indépendantes.

. Si les variables sont indépendantes, on a, pour des valeurs quelconques de h :

$$m^{(j)}_{h|} = m_{h|0} \qquad \text{pour } j = (1, 2, \ldots\ldots\ldots l),$$
$$m^{(i)}_{|h} = m_{0|h} \qquad \text{pour } i = (1, 2, \ldots\ldots\ldots k).$$

XIII. — *Équations de régression de Y par rapport à X (et de X en Y) du type parabolique et du type linéaire.*

Le calcul de l'espérance mathématique de Y par rapport à X, nous a conduit à l'équation de régression de Y par rapport à X $\quad m^{(i)}_{|1} = \mathrm{F}\,(x_i)$.

Si la fonction F est du type parabolique de degré f, on a l'expression :

$$(6) \qquad m^{(i)}_{|1} = a_{|0} + a_{\,1}\, x_i + a_{|2}\, x_i^2 + \ldots\ldots + a_{|f}\, x_i^f,$$

dont les coefficients peuvent être calculés en fonction des paramètres m, μ et r.

En effet, il suffit de multiplier les deux membres de l'équation (6) par $p_{i|}\, x_i^h$,

et de sommer par rapport à i en laissant h constant, et l'on obtient ainsi l'équation suivante :

$$(7) \qquad m_{h|1} = a_{|0}\, m_{h|0} + a_{|1}\, m_{h+1|0} + \ldots\ldots\ldots + a_{|f}\, m_{h+f|0};$$

Si dans (7) l'on fait $h = (0, 1, 2, \ldots\ldots f)$, on se trouve en présence d'un système de $(f + 1)$ équations linéaires à $(f + 1)$ inconnues.

Si l'équation de Y en X est linéaire, le calcul de $a_{|0}$ et de $a_{|1}$ effectué en partant des équations

$$(7') \qquad \begin{cases} m_{0|1} = a_{|0} + a_{|1}\, m_{1|0} \\ m_{1|1} = a_{|0}\, m_{1|0} + a_{|1}\, m_{2|0}, \end{cases}$$

montre que l'équation de régression peut s'écrire :

$$m_{|1}^{(i)} - m_{0|1} = \frac{m_{1|1} - m_{1|0}\, m_{0|1}}{m_{2|0} - (m_{1|0})^2}\,(x_i - m_{1|0}).$$

L'élimination de $a_{|0}$ et de $a_{|1}$ entre les équations (7') et l'équation (7)''.

$$(7'') \qquad m_{h|1} = a_{|0}\, m_{h|0} + a_{|1}\, m_{h+1|0}$$

fournit la condition suivante valable quelle que soit la valeur positive et entière de h.

$$(8) \qquad \frac{m_{h\,1} - m_{h|0} \cdot m_{0\,1}}{m_{h+1|0} - m_{h|0} \cdot m_{1|0}} = \frac{m_{1\,1} - m_{1|0}\, m_{0|1}}{m_{2|0} - (m_{1|0})^2}$$

Si l'on rapporte les coordonnées au point $(m_{1\,0}, m_{0\,1})$, on peut donner à l'équation de régression de Y en X la forme :

$$(9) \quad m_{|1}^{(i)} - m_{0\,1} = b_{\,0} + b_{|1}\,(x_i - m_{1|0}) + \ldots\ldots + b_{|f}\,(x_i - m_{1|0})^f;$$

en opérant sur (9) comme on l'a fait sur (6), c'est-à-dire en multipliant par $p_i\ (x_i - m_{1|0})^h$ et sommant, l'on trouve l'équation :

$$(10) \qquad \mu_{h\,1} = b_{\,0}\,\mu_{h\,0} + b_{\,1}\,\mu_{h+1\,0} + \ldots\ldots\ldots + b_{|f}\,\mu_{h+f\,0}$$

Si la régression est rectiligne, on constate — grâce un calcul simple — que :

$$b_{\,0} = 0 \text{ et } b_{|1} = \frac{\mu_{1|1}}{\mu_{2|0}}$$

et l'équation de la droite de régression s'écrit :

$$m_{|1}^{(i)} - m_{0|1} = \frac{\mu_{1|1}}{\mu_{2|0}}\,(x_i - m_{1|0})$$

Le coefficient :

$$b_{|1} = \frac{\mu_{1|1}}{\mu_{2|0}} = \frac{r_{1|1}\,\sigma_x\,\sigma_y}{\sigma_x^2} = r_{1|1}\,\frac{\sigma_y}{\sigma_x}$$

est dénommé à tort coefficient de régression, attendu que tous les coefficients de la formule de régression parabolique peuvent recevoir cette détermination ; peut-être faut-il attribuer cette terminologie à ce que les biologistes qui se sont occupés d'hérédité ont été amenés par l'intermédiaire de ce coefficient à caractériser en première approximation le phénomène de régression.

Dans le cas d'emploi des paramètres μ au lieu des paramètres m, on exprime

la condition pour que la régression de Y par rapport à X soit rectiligne, au moyen de la relation :

$$\frac{\mu_{h|1}}{\mu_{h+1|0}} = \frac{\mu_{1|1}}{\mu_{2|0}}, \text{ avec } h = (2, 3.....).$$

Si la régression de X par rapport à Y est linéaire, on trouve $m_{1|}^{(j)} - m_{1|0} = b_{1|}\,(y_j - m_{1\,0}) = \frac{\mu_{1|1}}{\mu_{0\,2}}\,(y_j - m_{0|1})$, avec la relation de condition

$$\frac{\mu_{1|h}}{\mu_{0|h+1}} = \frac{\mu_{1|1}}{\mu_{0\,2}}, \text{ avec } h = (2, 3.....).$$

Si les deux lignes de régression (de Y en X et de X en Y) sont rectilignes, l'on a :

$$b_{|1} = \frac{\sigma_y}{\sigma_x} r_{1\,1},\ b_{1|} = \frac{\sigma_x}{\sigma_y} r_{1|1}, \text{ et par suite}$$

$$b_{\,1}\,.\,b_1 = r_{1|1}^2 = \frac{\mu_{1|1}^2}{\mu_{2\,0}\,\mu_{0|2}};$$

Le coefficient de corrélation est la moyenne géométrique des coefficients de régression $b_{\,1}$ *et* b_1. Si dans l'équation de régression linéaire de Y par rapport à X, on introduit comme unités de mesure σ_x et σ_y, c'est-à-dire *si l'on adopte le système de coordonnées dites normales*, et si l'on pose $\frac{m_{1|}^{(i)} - m_{0\,1}}{\sigma_y} = \mathfrak{M}_{1|}^{(i)}$, $\frac{x_i - m_{1|0}}{\sigma_x} = \mathfrak{X}_i$, on voit que cette équation se met sous la forme réduite $\mathfrak{M}_{1|}^{(i)} = r_{1\,1}\,\mathfrak{X}_i$;

quant à la condition que doivent remplir les paramètres r pour que la régression soit linéaire, elle se réduit à : $r_{h\,1} - r_{1\,1}\,r_{h+1\,0}$, pour $h = (2, 3, 4).....$).

Avec ce même système de coordonnées, on trouve que la régression parabolique d'ordre f est caractérisée par l'équation $\mathfrak{M}_{1|}^{(i)} = c_{|0} + c_{\,1}\,\mathfrak{X}_i + c_{\,2}\mathfrak{X}_i^2 + ... + c_{|f}\,\mathfrak{X}_i^f$.

Compte tenu des relations :

$$\sum_i p_{i|}\,\mathfrak{X}_i^h = \frac{\mu_{h|0}}{\sigma_x^h} = r_{h|0},\ \sum_i p_{i|}\,\mathfrak{M}_{1|}^{(i)}\,\mathfrak{X}_i^h = \frac{\mu_{h\,1}}{\sigma_x^h\,\sigma_y} = r_{h|1}$$

$$r_{0|0} = 1, r_{1|0} = r_{0|1} = 0, r_{2|0} = r_{0|2} = 1,$$

on établit facilement les équations suivantes :

$$0 = c_{|0} + c_{|2} + \ldots + r_{f|0}\,c_{|f}$$
$$r_{1|1} = c_{|1} + r_{3|0}\,c_{|2} + ... + r_{f+1|0}\,c_{|f}$$
$$r_{2|1} = c_{|0} + r_{3|0}\,c_{|1} + r_{4|0}\,c_{|2} + ... + r_{f+2\,0}\,c_{|f}$$

Avec $f = 2$, on est ramené à un système simple pour le calcul des coefficients c_i, et à l'équation de régression :

$$\mathfrak{M}_{1|}^{(i)}\,\frac{r_{2|1} - r_{3|0}\,r_{1|1}}{r_{4|0} - r_{3|0}^2 - 1} + \left[r_{1|1} - \frac{r_{3|0}(r_{2|1} - r_{3|0}\,r_{1|1})}{r_{4|0} - r_{3|0}^2 - 1}\right]\mathfrak{X}_i + \frac{r_{2\,1} - r_{3|0}\,r_{1|1}}{r_{4|0} - r_{3|0}^2 - 1}\,\mathfrak{X}_i^2\,.$$

Alors que $r_{1|1}$ *serait nul, la ligne de régression resterait du type parabolique.*

Condition de non corrélation de Y avec X.

Lorsque l'espérance mathématique de Y avec X reste la même quel que soit X_i, l'équation de régression correspondante se réduit à :

$$m'^{(i)}_{|1} - m_{0\,1} = 0 \text{ ou } \mathfrak{M}^{(i)}_{|1} = 0$$

Il faut donc que les paramètres m remplissent les conditions : $m_{h|1} = m_{h|0} m_{0|1}$, ou encore que l'on ait $\mu_{h\,1} = 0$ ou $r_{h|1} = 0$ pour h : 1, 2, 3.....

Si la non corrélation de Y en X entraîne nécessairement la condition $r_{1|1} = 0$, *inversement lorsque le coefficient de corrélation est nul, on ne peut pas en conclure que les variables sont sans corrélation entre elles: Ce n'est que dans le cas où la régression est linéaire que l'on peut conclure à la non corrélation ; la condition* $r_{1|1} = 0$ *entraîne avec elle* $m'^{(i)}_1 = m_{0\,1}$.

De l'équation $\mu_{h\,1} = 0$ pour $h = 1, 2, 3, \ldots$, on ne peut pas en déduire $\mu_{1\,h} = 0$ pour $h = 2, 3, \ldots$; *il résulte de là que lorsque la variable Y est sans corrélation avec X, on ne peut pas en conclure que X est sans corrélation avec Y, alors que si les variables sont indépendantes l'une de l'autre, elles sont également sans corrélation entre elles.*

Les droites de régression et la méthode des moindres carrés.

La régression de Y par rapport à Y étant une parabole de degré f, on peut rechercher une droite telle que la somme des carrés des écarts des espérances mathématiques calculées en partant de la droite par rapport aux valeurs exactes soit minimum.

Les coefficients de la droite $M'^{(i)}_{|1} = A_0 + A_1 x_i$, devront être tels que l'expression $\Sigma p_i \left[m'^{(i)}_1 - A_0 - A_1 x_i\right]^2$ soit minimum; le calcul conduit pour $A_{|0}$ et A_1 à des valeurs qui sont les mêmes que celles afférentes au cas où la régression est linéaire.

Il résulte de cet exposé que l'équation

$$m'^{(i)}_{|1} - m_{0\,1} = \frac{\mu_{1\,1}}{\mu_{2|0}} (x_i - m_1)$$

représente la ligne de régression de Y par rapport à X, si cette régression est rectiligne, et s'en rapproche le plus dans le cas où la ligne de régression n'est pas une droite.

XIV. — *Dispersion déterminée moyenne et rapport de corrélation.*

De la connaissance de l'équation de régression déterminée de Y en X, on déduit la *valeur probable* de Y afférente à une valeur donnée X_i. En cas de liaison stochastique, on sait que Y conserve le caractère d'une variable accidentelle, et ses valeurs oscillent lorsque $X = X_i$ *autour de l'espérance mathématique déterminée.* Si l'on fait abstraction de l'équation de régression, on remarque que la variable Y est caractérisée par la dispersion non déterminée $\mu_{0|2}$; dans l'hypothèse contraire, on évalue l'espérance mathématique $m'^{(i)}_1$ relative à la valeur X_i de X, et l'on détermine l'amplitude des fluctuations de Y au moyen de la dispersion déterminée $\mu^{(i)}_{|2}$.

Si nous nous reportons à la relation

$$(11). \quad \sum_i p_{i|}\, \mu^{(i)}_{|2} = \sum_i \sum_j p_{i|}\, p^{(i)}_{|j} \left[y_i - m'^{(i)}_{|1}\right] = \mu_{0|2} - \sum_i p_{i|} \left[m^{(i)}_{|1} - m_{0|1}\right]^2,$$

nous constatons que $\Sigma_i\, p_{i|}\, \mu^{(i)}_{|2}$ est inférieure à $\mu_{0|2}$, sauf dans le cas où $(m^{(i)}_{|1} - m_{0,1}) = 0$, pour $i = (1, 2\,3, \ldots . k)$, c'est-à-dire si la variable Y est sans corrélation avec X.

Au carré moyen de contingence, fournissant une indication chiffrée de la liaison stochastique, grâce à l'introduction des probabilités $p_{i|}$, p_i , $p_{|j}$, il y a lieu, pour caractériser cette liaison, de rechercher d'autres indices qui tiennent compte non seulement des probabilités des diverses valeurs possibles des variables, mais de ces valeurs mêmes.

Parmi ces derniers, il faut attacher une importance particulière au *coefficient de corrélation et au rapport de corrélation.*

Par rapport de corrélation $\eta_{y|x}$, Pearson entend une expression définie par la relation (*) :

$$\eta^2_{y|x} = 1 - \frac{1}{\mu_{0\,2}} \sum_i p_{i|}\, \mu^{(i)}_{|2}; \tag{12}$$

eu égard à la valeur $\Sigma_i\, p_i\, \mu^{(i)}_2$ donnée ci-dessus, il s'en suit que le carré du rapport de corrélation s'écrit alors :

$$\eta^2_{y|x} = \frac{1}{\mu_{0|2}} \sum_i p_{i|}\, [m^{(i)}_{|1} - m_{0|1}]^2. \tag{12'}$$

Dans le cas où l'on a recours aux coordonnées normales, on peut mettre la valeur de $\eta^2_{y|x}$ sous la forme suivante : $\eta^2_{y\,x} = \Sigma_i\, p_{i|}\, (\mathfrak{M}^{(i)}_{|1})^2$.

S'il y a liaison stochastique de Y avec X, les dispersions déterminées de Y restent différentes de zéro. En cas de dépendance fonctionnelle de Y par rapport à X, les dispersions déterminées sont nulles, et le rapport de corrélation est égal à 1 ; inversement, si $\eta_{y\,x} = 1$, cela revient à dire que la dispersion déterminée moyenne de Y est nulle, ce qui ne peut se produire que si toutes les dispersions déterminées s'annulent, ce qui ne peut se produire que si la *dépendance est fonctionnelle.* Si maintenant l'on revient à la relation (11), on voit que $\Sigma_i\, p_{i|}\, [m^{(i)}_{|1} - m_{0\,1}]^2$ est inférieur à $\mu_{0|2}$, puisque la dispersion déterminée moyenne $\Sigma_i\, p_{i|}\, \mu^{(i)}_{|2}$ est positive; il s'en suit que $\eta^2_{y\,x}$ *ne peut qu'osciller entre 0 et 1.*

Le rapport de corrélation n'est égal à zéro que si toutes les quantités $m^{(i)}_{|1}$ sont égales entre elles, c'est à dire si Y est sans corrélation avec X.

Calcul du rapport de corrélation. — Pour ce calcul, on peut partir aussi bien de :

$$\eta^2_{y|x} = 1 - \frac{1}{\mu_{0|2}} \sum_i p_{i|}\, \mu^{(i)}_{|2}, \text{ que de } \eta^2_{y|x} = \sum_i p_{i|}\, (\mathfrak{M}^{(i)}_{|1})^2$$

1° Cas où l'équation de régression de Y en X est du type parabolique de degré 2.

Cette équation en coordonnées normales s'écrit ainsi qu'il suit :

$$\mathfrak{M}^{(i)}_{|1} = r_{1|1}\, \mathfrak{X}_i + c_{|2}\, [\mathfrak{X}^2_i - r_{3|0}\, \mathfrak{X}_i - 1]$$

(*) K. Pearson. *On thee general theory ofshew corrélation and non lincar regression* (*Drapers Company research Memoirs*, Biométrie, séries II, 1905, p. 10).

avec :

$$c_{1|2} = \frac{r_{2\,1} - r_{3|0}\, r_{1|1}}{r_{4|0} - r_{3|0}^2 - 1}$$

et le carré du rapport de corrélation, compte tenu de ce que :

$$\Sigma\, p_{i|}\, \mathfrak{X}_i^2 = 1,\ \Sigma\, p_{i|}\, \mathfrak{X}_i = 0,\ \Sigma\, p_{i|}\, \mathfrak{X}_i^3 = r_{3|0},\ \Sigma\, p_{i|}\, \mathfrak{X}_i^4 = r_{4|0},$$

a pour valeur :

$$\eta_{y|x}^2 = r_{1|1}^2 + \frac{(r_{2\,1} - r_{3\,0}\, r_{1|1})^2}{r_{4|0} - r_{3|0}^2 - 1}$$

2° La régression de Y par rapport à X est linéaire.

On a, dans ce cas :

$$\mathfrak{M}_{1}^{(i)} = r_{1|1}\, \mathfrak{X}_i,\ \text{et}\ \eta_{y|x}^2 = \Sigma\, p_{i|}\, r_{1|1}^2\, \mathfrak{X}_i^2\, r_{1|1}^2;$$

Le rapport de corrélation et le coefficient de corrélation sont alors égaux en valeur absolue.

Si la régression n'est pas linéaire, on a $r_{1\,1}^2 < \eta_{y\,x}^2$; en effet, l'équation de la *droite ajustée à la vraie ligne de régression de Y en X* pouvant être écrite comme il suit :

$$\text{(13)} \qquad \mathrm{M}_1^{(i)} = m_{0|1} + \frac{\mu_{1|1}}{\mu_{2\,0}}(x_i - m_{1|0}),$$

revenons à la dispersion

$$\sum_i p_{i|}\, [m_{|1}^{(i)} - \mathrm{M}_{|1}^{(i)}]^2,$$

dont l'expression a pour valeur :

$$\mu_{0|2} - \sum_i p_{i|}\, \mu_{|2}^{(i)} - \frac{\mu_{1|1}^2}{\mu_{2\,0}};$$

il résulte de la relation :

$$\eta_{y|x}^2 - r_{1|1}^2 = \frac{1}{\mu_{0\,2}} \Sigma\, p_{i|}\, [m_{|1}^{(i)} - \mathrm{M}_{|1}^{(i)}]^2,$$

et l'on constate que la différence

$$(\eta_{y|x}^2 - r_{1|1}^2)$$

ne peut être négative, et qu'elle n'est nulle que si les grandeurs $m_{|1}^{(i)}$ coïncident avec les quantités correspondantes $\mathrm{M}_1^{(i)}$, ce qui ne peut avoir lieu que si la *droite ajustée (13) se confond avec la ligne de régression exacte.*

Lorsque la régression est rectiligne, il y a lieu d'interpréter parallèlement les valeurs prises par le coefficient de corrélation et le rapport de corrélation; si le coefficient de corrélation est égal à 1, Y est liée à X par une équation linéaire, et si ce même coefficient est nul, la variable Y est sans corrélation avec X.

Si la régression n'est pas rectiligne, le coefficient de corrélation reste toujours INFÉRIEUR *au rapport de corrélation;* à un rapport de corrélation égal à 1, correspond un coefficient $r_{1|1} < 1$; il résulte de là qu'en cas de dépendance fonctionnelle non linéaire entre Y et X, le coefficient de corrélation est inférieur à 1, et s'en écarte plus ou moins suivant la nature de la liaison fonctionnelle.

Si l'on n'a point fourni à l'avance au chercheur une indication sur la linéarité

de la régression, celui-ci ne peut pas conclure à la non-corrélation des variables si le coefficient de corrélation est nul; de même si ce coefficient est inférieur à 1, on ne doit nullement en déduire qu'il n'existe point de dépendance fonctionnelle, car une dépendance non linéaire peut fort bien exister.

XV. — *De quelques remarques importantes sur le coefficient de corrélation.*

L'Institut International de Statistique ayant constaté que certains statisticiens employaient sans précautions préalables le coefficient de corrélation, a jugé utile de mettre à l'étude l'emploi de ce coefficient et a chargé M. Frechet, professeur à la Faculté des Sciences de Paris, d'établir un rapport sur cette délicate question.

C'est son rapport publié à l'occasion de la XXIIe session de l'Institut International de Statistique à Londres en 1934, qui fait suite à sa note parue dans la revue du même groupement scientifique, que nous allons analyser ici et qui résume les avis exprimés au sujet de l'usage du coefficient de corrélation par les membres de la Commission nommés à cet effet.

Rappelons tout d'abord que de nombreux statisticiens emploient pour repérer le coefficient de dépendance entre deux variables statistiques (X, Y), le coefficient

$$r = \frac{\sum_{i,j} n_{ij} (x_i - a)(y_j - b)}{\sqrt{\sum n_{i.} (x_i - a)^2}\ \sqrt{\sum n_{.j} (y_j - b)^2}}$$

où n_{ij} est le nombre de fois que le couple (X, Y) prend le couple de valeurs (x_i, y_j), où $n_i = \sum_j n_{ij}$, $n_{.j} = \sum_i n_{ij}$, et où a et b sont les valeurs moyennes de x_i et y_j, et remarquons que l'on a toujours $|r| \leqslant 1$, et que l'égalité $|r| = 1$ est la condition nécessaire et suffisante pour que X et Y soient liées par une relation linéaire.

Il y a lieu d'observer que le coefficient r peut être aussi voisin de zéro que l'on veut, alors même que X et Y sont liées par une relation biunivoque, comme l'a fait remarquer d'une part si judicieusement M. de Mises, et, d'autre part, M. Rietz, dans son traité : *Handbook of Mathematical Statistics*, paru en 1927.

Les fausses interprétations de ce coefficient observées jusqu'ici auraient été fort probablement évitées en modifiant l'appellation de r, et en le désignant avec M. Frechet *coefficient de linéarité.*

Conservant les notations de cet auteur, désignons par b_i l'ordonnée du centre de gravité de la file i :

$$b_i = \frac{\sum_j n_{ij} y_j}{\sum_j n_{ij}}$$

Nous voyons de suite que :

$$\sum_{i,j} n_{ij} (x_i - a)(y_j - b) = \sum_i (x_i - a)(b_i - b)\, n_{i.}$$

Ceci étant, reportons nous à la valeur du rapport de corrélation de Pearson,

qui est définie par l'expression $\eta_{y|x}$ (ou η pour simplifier l'écriture), que l'on peut écrire :

$$\eta = \sqrt{\frac{\Sigma n_{i|} (b_i - b)^2}{\Sigma n_{|j} (y_j - b)^2}},$$

et rapprochons-la de la valeur ρ de l'expression suivante :

$$\rho = \frac{\Sigma (x_i - a) n_{i|} (b_j - b)}{\sqrt{\Sigma n_{i|} (x_i - a)^2} \sqrt{\Sigma n_{i|} (b_i - b)^2}}$$

On voit ainsi que r s'écrit :

$$r = \frac{\Sigma_{ij} n_{ij} (x_i - a) (y_j - b)}{\sqrt{\Sigma_i n_{i|} (x_i - a)^2} \sqrt{\Sigma_i n_{i|} (b_i - b)^2}} \frac{\sqrt{\Sigma_i n_{i|} (b_i - b)^2}}{\sqrt{\Sigma_j n_{|j} (y_j - b)^2}}$$

et que l'on aboutit à la formule très simple (1) $r = \rho\eta$ (Voir *C. R. Académie des Sciences*, t. 197, M. Fréchet).

Il y a lieu de remarquer que ρ peut être considéré *comme le coefficient de corrélation de la distribution moyenne*, c'est à-dire de la distribution (x_i, b_i) que l'on réaliserait en remplaçant les y_j relatifs à un même x_i par leur moyenne b_i répétée $n_{i|}$ fois.

Ce coefficient est ρ donc insensible à la plus ou moins grande dispersion autour de la courbe moyenne; quant au coefficient η, il n'est pas affecté par la déformation de la courbe des moyennes, résultant d'une modification des x_i.

On voit ainsi apparaître dans la valeur de $r = \rho\eta$ un nombre de ρ qui est en définitive un facteur étranger à la dépendance de (X, Y) et vient en quelque sorte fausser sa mesure.

Ceci étant posons :

$$y_j = b_i + \lambda_{ij}, \quad \sigma^2 = \Sigma n_{i|} (b_i - b)^2,$$

et

$$\mu^2 = \Sigma_i \Sigma_j n_{ij} (\lambda_{ij})^2,$$

et remarquons que l'expression de μ^2 peut s'écrire :

$$\mu^2 = \Sigma\Sigma n_{ij} [(y_j - b) - (b_i - b)]^2 = \Sigma\Sigma n_{ij} (y_j - b)^2 - 2\Sigma\Sigma n_{ij} (y_j - b) (b_i - b) + \Sigma\Sigma n_{ij} (b_i - b)^2$$

ou encore :

$$\mu^2 = \Sigma n_{|j} (y_j - b)^2 - 2\Sigma\Sigma n_{ij} (y_j - b) (b_i - b) + \Sigma n_{i|} (b_i - b)^2,$$

Comme l'on a :

$$\Sigma\Sigma n_{ij} (y_j - b) (b_i - b) = \Sigma (b_i - b) \Sigma n_{ij} (y_j - b) = \Sigma_i n_{i|} (b_i - b)^2,$$

il en résulte que :

$$\mu^2 = \Sigma_j n_{|j} (y_j - b)^2 - \Sigma_i n_{i|} (b_i - b)^2,$$

et l'on trouve en définitive :

$$\mu^2 + \sigma^2 = \sum_j n_{ij} (\bar{y}_j - b)^2.$$

De cette dernière relation, l'on déduit l'expression du rapport de corrélation de y par rapport à x :

$$\eta = \frac{\sigma}{\sqrt{\mu^2 + \sigma^2}} = \frac{1}{\sqrt{1 + \frac{\mu^2}{\sigma^2}}},$$

et l'on constate que la grandeur de η est plus petite ou au plus égale à 1; on remarque de plus que pour deux distributions moyennes identiques, celle pour laquelle la valeur de Y relative à $X = X_i$ est la mieux définie est celle pour laquelle les écarts de la distribution réelle avec la distribution moyenne sont les plus faibles.

En définitive, on est conduit naturellement à prendre pour une même distribution moyenne la quantité η comme donnant une mesure de l'étroitesse de la dépendance fonctionnelle de X et de Y.

De la relation $r = \varrho\eta$, on déduit qu'à une valeur de ϱ fixe, l'on fait correspondre une valeur de η proportionnelle à r, et, par suite, que pour une même distribution moyenne, la dépendance fonctionnelle est d'autant plus accusée que r est plus grand, alors que la courbe moyenne n'est pas une droite.

Le maximum de r pour une même distribution moyenne correspond à celui de η ; la valeur de ce maximum qui dépend de la forme de la ligne moyenne est au plus égale à $|\varrho|$, et n'est égale à l'unité que si la ligne moyenne est une droite.

A la variation d'une distribution se rattache celle de son coefficient de corrélation, qui dépend d'une part de la forme de la ligne moyenne, et d'autre part de l'étroitesse de la dépendance fonctionnelle.

Jusqu'ici, nous avons comparé deux distributions *ayant la même courbe moyenne* ; examinons maintenant le cas de deux distributions dont les courbes moyennes sont à peu près les mêmes.

A chacune de ces deux distributions, on fait correspondre à la première la relation $r = \varrho\eta$, et à la seconde la relation $r' = \varrho'\eta'$ (ϱ et ϱ' étant très voisins), et l'on supposera de plus que ϱ et ϱ' diffèrent peu de 1.

Il est évident que si la valeur du rapport de corrélation η afférente à la première distribution est petite, et celle de η' voisine de 1, il s'en suit que la valeur de r est voisine de zéro, et celle de r' voisine de 1. Au cas où la différence des valeurs de r et de r' n'est pas aussi marquée, il est évidemment très peu prudent de déduire des conclusions de l'examen des valeurs des coefficients de corrélation; il suffit pour s'en rendre compte de prendre pour première distribution, une distribution correspondant *exactement* à une courbe tendant vers une droite, par exemple :

$Y = X \,.\, |X|^{\varepsilon}$, où ε est un nombre positif, et où X varie de —1 à +1.

Pour une telle courbe, l'on a :

$$\eta = 1, r = \rho = \frac{\sum_i x_i^2 |x_i|^{\varepsilon}}{\sqrt{(\sum_i x_i^2) \sum_i x_i^2 |x_i|^{2\varepsilon}}}$$

et l'on voit que r tend vers 1 lorsque ε tend vers zéro.

Ceci étant, considérons une seconde distribution où la ligne des moyennes est rectiligne, mais où à chaque abscisse x_i correspondent des valeurs distinctes de Y, ou $(b_i + \lambda_{ij})$. On aura ici $\rho' = \pm 1$ (par exemple : $\rho' = 1$) et le coefficient de corrélation r' sera égal à η', d'où il résulte :

$$r' = \eta' = \frac{1}{\sqrt{1 + \frac{\mu^2}{\sigma^2}}},$$

Si donc l'on laisse fixe la distribution moyenne et si l'on fait tendre μ vers zéro, r' tendra vers l'unité, et il s'en suit que pour toute valeur de $\varepsilon < \omega$, l'on aura une valeur de μ et même une infinité de valeurs de $\mu \neq 0$, telles que $r < r' < 1$. En définitive, l'on se trouve en présence de deux distributions, la première afférente à une relation fonctionnelle parfaite et l'autre à une relation fonctionnelle imparfaite, et telles que le coefficient de corrélation de la première est inférieur à celui de la seconde; il découle de là que dans la comparaison de deux distributions, il ne sera guère possible comme le dit M. Fréchet, « de tirer des conclusions à l'étroitesse plus petite ou plus grande des dépendances fonctionnelles correspondantes que si les lignes moyennes sont peu différentes de forme, et si en outre les coefficients de corrélation sont très différents ».

Remarques :

(1) Nous avons montré — au cours de notre étude — que si les variables X et Y suivent la loi de Laplace-Gauss, les lignes de régression sont des droites; en ce cas, l'on a $|r| = \eta$. Dans l'hypothèse où X et Y n'obéissent qu'approximativement à la loi de Laplace Gauss, il est évident que le statisticien doit recourir à l'examen du rapport de corrélation η.

(2) On sait que pour la loi de distribution

$$f(x, y) = \sqrt{\frac{ac - b^2}{2\pi}}\, e^{-\frac{(ax^2 + bxy + cy^2)}{2}},$$

les moments du second ordre $\mu_{2|0}$, $\mu_{0\,2}$ et $\mu_{1|1}$ et le coefficient de corrélation $r_{1|1}$ ont respectivement pour valeur :

$$\mu_{2|0} = \frac{c}{ac - b^2},\ \mu_{0|2} = \frac{a}{ac - b^2},\ \mu_{1|1} = -\frac{b}{ac - b^2} \text{ et } r_{1|1} = \frac{\mu_{1|1}}{\sqrt{\mu_{2|0}\,\mu_{0|2}}}.$$

On en déduit :

$$1 - r_{1|1}^2 = \frac{\mu_{2|0}\,\mu_{0|2} - \mu_{1|1}^2}{\mu_{2|0}\,\mu_{0|2}},$$

et l'on remarque que la dispersion liée de y pour une valeur donnée de x, est égale à $\frac{1}{\sqrt{c}}$ ou $(\lambda \sqrt{\mu_{0|2}})$, le facteur λ étant — comme l'on s'en rend facilement compte — égal à $(\sqrt{1 - r^2})$. Il s'en suit que les valeurs de la dispersion liée sont pour

$$r^2 = \quad 0{,}96, \quad 0{,}64, \quad 0{,}50, \quad 0{,}04, \quad 0{,}01,$$
$$0{,}2\sqrt{\mu_{0|2}},\ 0{,}6\sqrt{\mu_{0|2}},\ \frac{1}{\sqrt{2}}\sqrt{\mu_{0|2}},\ \sqrt{0{,}96 \times \mu_{0|2}},\ \sqrt{0{,}99 \times \mu_{0|2}}$$

CHAPITRE III

LA LOI DE LAPLACE-GAUSS — DÉVELOPPEMENTS PLUS APPROCHÉS GÉNÉRALISATION — INTRODUCTION DE LA CORRÉLATION TOTALE ET DES COEFFICIENTS DE CORRÉLATION PARTIELLE.

Alors que pour les séries statistiques simples, nous avons été conduit à utiliser en vue d'une représentation satisfaisante, soit des formes de fonctions différant de la forme classique de Laplace Gauss, soit des développements en série fondés sur l'utilisation de la fonction normale $e^{-h^2x^2}$ et de ses dérivées successives, nous allons maintenant rechercher — après avoir remarqué que la représentation de Laplace-Gauss ne fournissait pas toujours pour les séries statistiques doubles des images correctes — si des développements plus approchés pourraient nous conduire à des résultats meilleurs. Nous donnerons ensuite quelques indications sur la généralisation et l'utilisation dans un espace à n dimensions de la surface de Laplace Gauss en faisant apparaître les coefficients de corrélation totale et de corrélation partielle.

1) *La fonction caractéristique et le schéma de Bernouilli à deux variables.*

Considérons une urne renfermant trois espèces de boules, des boules blanches, des boules noires et des boules d'autres couleurs, auxquelles on rattache les probabilités respectives p_1, p_2, p_3. On a fait N extractions après avoir remis dans l'urne après chaque tirage la boule tirée, et soient m_1, m_2, m_3 les nombres des boules blanches, noires et d'autres couleurs qui sont apparues. La probabilité d'une telle éventualité a pour valeur :

$$\frac{N!}{m_1!\, m_2!\, m_3!}\, p_1^{m_1}\ p_2^{m_2}\ p_3^{m_3} = P.$$

Si l'on n'envisage que les boules noires et blanches sorties (en nombre m_1 pour les blanches et en nombre m_2 pour les noires), on voit — en remarquant que les variables m_1, m_2 ne sont pas indépendantes — que la fonction caractéristique qui leur est attachée est définie par l'expression :

$$\Sigma\, P\, e^{u_1 m_1 + u_2 m_2} = (p_1\, e^{u_1} + p_2\, e^{u_2} + p_3)^N,$$

comme l'on s'en rend compte en revenant à la somme $\Sigma p_{i,j} e^{u_i x_i + v_j y_j}$; qui pour une seule épreuve donne $p_1 e^{u_1} + p_2 e^{u_2} + p_3$ puisque l'on ne considère que les boules blanches et noires.

Si maintenant l'on compte les variables aléatoires, à partir de leurs valeurs probables, on doit pour une seule épreuve introduire $e^{-(u_1 p_1 + u_2 p_2)}$; il s'en suit que pour N épreuves, la fonction caractéristique est représentée par :

$$\varphi\,(u_1, u_2) = e^{-N(u_1 p_1 + u_2 p_2)}\,(p_1\, e^{u_1} + p_2\, e^{u_2} + p_3)^N$$

Le logarithme de la fonction caractéristique s'écrit alors :

$$\begin{aligned}\operatorname{Log} \varphi\,(u_1, u_2) &= -N\,(u_1 p_1 + u_2 p_2) + N \log\,(p_1\, e^{u_1} + p_2\, e^{u_2} + p_3) \\ &= -N\,(u_1 p_2 + u_2 p_2) + N \log\,[1 + p_1\,(e^{u_1} - 1) + p_2\,(e^{u_2} - 1)]\end{aligned}$$

et, en posant :

$$l_1 = m_1 - N\, p_1,\ l_2 = m_2 - N\, p_2,\ 1 - p_1 = q_1,\ 1 - p_2 = q_2,$$

il est alors représenté par l'expression suivante :

$$\log \varphi = \psi(u_1, u_2) = N\left[\frac{p_1 q_1 u_1^2}{2} + \frac{p_2 q_2 u_2^2}{2} - p_1 p_2 u_1 u_2\right| + \cdots$$

Si l'on rapporte les écarts l_1 et l_2 aux grandeurs

$$\sqrt{N p_1 q_1},\ \sqrt{N p_2 q_2}$$

l'on est conduit à substituer à u_1 et u_2 les quantités respectives

$$\frac{u_1}{\sqrt{N p_1 q_1}},\ \frac{u_2}{\sqrt{N p_2 q_2}};$$

il en résulte que le logarithme de φ *avec la loi réduite* (1) a pour valeur :

$$\overline{\psi}(u_1, u_2) = \frac{1}{2}(u_1^2 + u_2^2 + 2 r u_1 u_2) + \frac{1}{\sqrt{N}} A_3(u_1 u_2) + \frac{1}{N} A_4(u_1, u_2) + \cdots$$

Au cas où le nombre des tirages est grand, on peut en première approximation prendre pour fonction caractéristique de la probabilité :

$$\varphi = e^{\frac{1}{2}(u^2 + u_2^2 + 2 r u_1 u_2)},$$

avec :

$$r = \sqrt{\frac{p_1 p_2}{(1 - p_1)(1 - p_2)}}.$$

Quant à la première approximation de la loi de probabilité, elle sera définie par l'expression :

$$f(x_1, x_2) = \int_{-\infty}^{+\infty}\int e^{\frac{1}{2}(u_1^2 + u_2^2 + 2 r u_1 u_2)} e^{u_1 x_1 + u_2 x_2} du_1\, du_2 = \frac{1}{2\pi\sqrt{1 - r^2}} e^{-\frac{x_1^2 + x_2^2 - 2 r x_1 x_2}{1 - r^2}}$$

Remarque. — On peut donner au résultat trouvé une forme symétrique des plus remarquables. Il suffit en effet de revenir à la forme

$$\psi(u_1, u_2) = \frac{N}{2}[p_1 u_1^2 + p_2 u_2^2 - (p_1 u_1 + p_2 u_2)^2] + \cdots$$

et de rapporter les écarts de l_1 et l_2 à une unité commune $\sqrt{N}$; dans ces conditions, le logarithme de la fonction caractéristique s'écrit :

$$\frac{1}{2}[p_1 u_1^2 + p_2 u_2^2 - (p_1 u_1 + p_2 u_2)^2] + \frac{(A_1(u_1, u_2)}{N} + \cdots$$

Ceci étant, ne conservons pour ψ que le premier élément (forme quadratique en u_1 et u_2, que nous considérerons comme l'équation tangentielle d'une conique, et remarquons que l'on est amené — si l'on veut passer aux coordonnées carté-

(1) On dit qu'une loi de probabilité $f(x)$ est *réduite* lorsque l'origine des abscisses est la valeur probable E (x) et que l'écart moyen quadratique est pris pour unité de mesure, c'est-à-dire lorsque l'on a $m_1 = \int_{-\infty}^{+\infty} x f(x)\, dx = 0$, et $\mu_2 = \int_{-\infty}^{+\infty} x^2 f(x)\, dx = 1$; on étend ces notions à des lois de probabilités afférentes à un espace à n dimensions.

siennes — à introduire les équations suivantes déduites de la théorie des enveloppes :

$$(1)\quad p_1[u_1 - (p_1 u_1 + p_2 u_2)] = x_1, \quad p_2[u_2 - (p_1 u_1 + p_2 u_2)] = x_2,$$
$$(2)\quad p_1 u_1^2 + p_2 u_2^2 - (p_1 u_1 + p_2 u_2)^2 = u_1 x_1 + u_1 x_2,$$

d'où l'on déduit facilement :

$$(2)'\quad p_1 u_1^2 + p_2 u_2^2 - (p_1 u_1 + p_2 u_2)^2 = \frac{(x_1 + x_2)^2}{p_3} + \frac{x_1^2}{p_1} + \frac{x_2^2}{p_2},$$

Si maintenant, nous faisons intervenir l'écart $l_3 = m_3 - Np_3$, et la variable $x_3 = \frac{l_3}{\sqrt{N}}$, on voit, en posant : $x_1 = \frac{l_1}{\sqrt{N}}, x_2 = \frac{l_2}{\sqrt{N}}$, que l'on a :

$$x_1 + x_2 + x_3 = \frac{l_1 + l_2 + l_3}{\sqrt{N}} = \frac{m_1 + m_2 + m_3 - N(p_1 + p_2 + p_3)}{\sqrt{N}} = 0$$

et que le second membre de (2)' s'écrit :

$$\frac{x_1^2}{p_1} + \frac{x_2^2}{p_2} + \frac{x_3^2}{p_3}$$

Or, à la fonction de dépendance ou loi de probabilité :

$$f(x_1 y) = \frac{\sqrt{ac - b^2}}{2\pi} e^{-\frac{ax^2 + 2bxy + cy^2}{2}}$$

correspond la fonction caractéristique :

$$\varphi(u_1 v) = e^{\frac{K(u, v)}{2}},$$

avec :

$$K(u, v) = \frac{cu^2 + av^2 - 2buv}{ac - b^2};$$

il résulte de là qu'à la fonction :

$$K(u_1, u_2) = p_1 u_1^2 + p_2 u_2^2 - (p_1 u_1 + p_2 u_2)^2$$

l'on rattache la loi de probabilité :

$$\frac{1}{2\pi\sqrt{p_1 p_2 p_3}} e^{-\frac{1}{2}\left[x_1^2\left(\frac{1}{p_1} + \frac{1}{p_3}\right) + x_2^2\left(\frac{1}{p_2} + \frac{1}{p_3}\right) + \frac{2x_1 x_2}{p_3}\right]} = \frac{1}{2\pi\sqrt{p_1 p_2 p_3}} e^{-\frac{1}{2}\left(\frac{x_1^2}{p_1} + \frac{x_2^2}{p_2} + \frac{x_3^2}{p_3}\right)}$$

2) *Développement d'Edgeworth.*

Si dans le développement de la fonction caractéristique ramenée à sa forme réduite, nous ne faisons intervenir que l'expression $e^{\frac{1}{2}(u^2 + 2ruv + v^2)}$, nous aboutissons à la loi de probabilité normale; dans le cas invoqué ci-dessus du tirage de N boules d'une urne renfermant des boules blanches et noires, ainsi que des boules d'autres couleurs, où l'on ne s'en tient pas à la première approximation, la fonction caractéristique a pour valeur :

$$e^{\frac{1}{2}(u^2 + 2ruv + v^2)} + \frac{A_3(u, v)}{\sqrt{N}} + \frac{A_4(u, v)}{N} + \ldots\ldots$$

qui peut encore s'écrire :

$$\varphi(u, v) = \varphi_0(u, v)\,[1 + \alpha_0(u, v) + \ldots\ldots] \qquad \text{avec } \varphi_0 = e^{\frac{1}{2}(u^2 + 2ruv + v^2)}.$$

Revenons maintenant à la définition de la fonction caractéristique :

$$\varphi(u, v) = \int\int f(x, y)\, e^{ux + vy}\, dx\, dy,$$

où l'intégration est étendue à toute la région du plan où $f(x, y)$ est définie (où à tout le plan, puisqu'il suffit de remplacer $f(x, y)$ par zéro dans certains domaines, et admettons que l'on substitue à $f(x, y)$ la fonction

$$f_0 = \frac{1}{2\pi\sqrt{1 - r^2}}\, e^{-\frac{2}{1 - r^2}(x^2 + y^2 - 2rxy)}$$

représentative de la corrélation normale.

Ceci étant, remarquons que :

$$\int_{-\infty}^{+\infty}\int \frac{\partial f_0}{\partial x}\, e^{ux + vy}\, dx\, dy = \int_{-\infty}^{+\infty}\int \left[\frac{\partial}{\partial x}\left(f_0\, e^{ux + vy}\right) - f_0\, u e^{ux + vy}\right] dx\, dy,$$

expression qui se réduit à :

$$-u \int_{-\infty}^{+\infty}\int f_0\, e^{ux + vy}\, dx\, dy = -u\,\varphi_0(u, v)$$

en raison de ce que f_0 s'annule aux limites extrêmes du champ comme $e^{-\rho^2}$ (avec $\rho = 0M$, distance de l'origine au point M).

On trouve par un procédé analogue que :

$$\int_{-\infty}^{+\infty}\int \frac{\partial^{m+n} f_0}{\partial x^m\, \partial y^n}\, e^{ux + vy}\, dx\, dy = (-1)^{m+n} u^m v^n \varphi_0(u, v),$$

et il s'en suit qu'à un terme $u^m v^n \varphi_0$ correspond le terme $\dfrac{\partial^{m+n} f_0}{\partial x^m\, \partial y^n}$ pour f.

Au développement :

$$\varphi(u, v) = \varphi_0\,[1 + (\lambda_{3,0}\, u^3 + \lambda_{1,2}\, u\, v^2 + \lambda_{2,1}\, u^2\, v + \lambda_{0,3}\, v^3) + (\lambda_{4,0}\, u^4 + \ldots)],$$

se rattache la fonction de probabilité :

$$f(x, y) = f_0(x, y) - \left(\lambda_{3,0}\frac{\partial^3 f_0}{\partial x^3} + \lambda_{2,1}\frac{\partial^3 f_0}{\partial x^2\, \partial y} + \lambda_{1,2}\frac{\partial^3 f_0}{\partial x\, \partial y^2} + \lambda_{0,3}\frac{\partial^3 f_0}{\partial y^3}\right) + \ldots,$$

et après réductions :

$$f(x, y) = f_0\,[1 + \lambda_{3,0}\,[(x - ry)^3 - 3(x - ry)] + \lambda_{2,1}\,[(x - ry)^2 (y - rx) + 3rx - y(1 + 2r^2) + \ldots\ldots]$$

Eu égard à la définition de la fonction caractéristique, on sait que l'on a :

$$\frac{d^{g+h}\,\varphi(u, v)}{du^g\, dv^h} = m_{g,h}$$

(pour $u = v = 0$),

et comme :

$$\varphi(u, v) = e^{\frac{1}{2}(u^2 + v^2 + 2ruv)} [1 + \lambda_{3,0} u^3 + \lambda_{2,1} u^2 v + \lambda_{1,2} u v^2 + \lambda_{0,3} v^3 + \ldots]$$

peut encore s'écrire :

$$\varphi(u, v) = 1 + \frac{u^2 + v^2 + 2ruv}{2} + \lambda_{3,0} u^3 + \lambda_{2,1} u^2 v + \lambda_{1,2} u v^2 + \lambda_{0,1} v^3 + \ldots,$$

On remarque immédiatement que :

$$\frac{\partial^3 \varphi}{\partial u^3} = 3!\,\lambda_{3,0} = m_{3|0}, \frac{\partial^3 \varphi}{\partial u^2 \partial v} = 2!\,1!\,\lambda_{2|1} = m_{2|1}, \frac{\partial^3 \varphi}{\partial u \partial v^2} = 1!\,2!\,\lambda_{1,2} = m_{1,2}$$

$$\frac{\partial^3 \varphi}{\partial v^3} = 3!\,\lambda_{0,3} = m_{0|3}.$$

Or, à l'aide des données expérimentales, l'on peut déterminer les valeurs empiriques de $m_{3,0}$, $m_{2\,1}$, $m_{1,2}$, $m_{0,3}$....., c'est-à-dire :

$$\Sigma \frac{x'^3}{N}, \Sigma \frac{x'^2 y'}{N}, \Sigma \frac{x' y'^2}{N}, \Sigma \frac{y'^3}{N} \ldots;$$

on peut donc calculer $\lambda_{3,0}$,..... $\lambda_{0,3}$....., et, par suite, obtenir le développement de $f(x, y)$.

3) *Les polynomes d'Hermite et leur utilisation dans le développement d'une fonction de probabilité.*

Nous avons montré antérieurement que l'on a utilisé les dérivées successives de la fonction $f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, pour définir la fonction de fréquence d'une série statistique en faisant appel au développement suivant :

$$f_0(x) = \alpha_0 f_0 + \alpha_1 f_0' + \alpha_2 f''(o) + \ldots\ldots + \alpha_n f_0^{(n)} + \ldots\ldots;$$

cette méthode a été transposée dans l'espace en introduisant les systèmes de polynomes d'Hermite $U_{m,n}$, $V_{m,n}$, par M. Guldberg (*Application des polynomes d'Hermite à un problème de statistique*, Congrès international des Mathématiciens de 1920, Strasbourg).

Pour cela, on fait apparaître à côté de la fonction $\varphi_0 = ax^2 + 2\,bxy + cy^2$, que l'on suppose constamment positive, son contrevariant quadratique $\psi_0 = cx^2 - 2\,bxy + ay^2$ et l'invariant $\delta = (ac - b^2)$, et l'on définit le système des polynomes $U_{m,n}$ par l'équation :

$$e^{-\frac{1}{2}\varphi_0(x + h, y + k)} = e^{-\frac{1}{2}\varphi_0(x, y)} \Sigma \frac{h^m k^n}{m!\,n!} U_{m,n},$$

d'où l'on déduit :

$$e^{-[h(ax + by) + k(bx + cy)]} = e^{\frac{1}{2}\varphi_0(h, k)} \Sigma \frac{h^m k^n}{m!\,n!} U_{m,n}.$$

Le second système de polynomes $V_{m,n}$ se déduit du précédent par la substitution :

$$h = \frac{\partial \psi_0(h_1, k_1)}{\partial h_1}, k = \frac{\partial \psi_0 h_1, k_1)}{\partial k_1}$$

et par l'équation

$$e^{-\frac{1}{2}\varphi_0\left(x+\frac{\delta\psi_0}{\delta h_1},\ y+\frac{\delta\psi_0}{\delta k_1}\right)} = e^{-\frac{1}{2}\varphi_0(x,y)} \cdot \Sigma \frac{h_1^m k_1^n}{m!\,n!} V_{m,n}$$

Posant $(ax + by = \xi,\ bx + cy = \eta)$, on calcule $U_{0,0}\,U_{1,0}\,U_{0,1}\,U_{2,0}\,U_{1,1}\,U_{0,2}\ldots$ et l'on passe des $U_{m,n}$ aux $V_{m,n}$ en substituant aux quantités

$$\xi,\ \eta,\ a,\ b,\ c,$$

les quantités

$$x,\ y,\ \frac{c}{\delta},\ -\frac{b}{\delta},\ \frac{a}{\delta}.$$

Les U satisfont aux équations :

$$\begin{cases} U_{m+1,n} - \xi\, U_{m,n} + a\, m\, U_{m-1,n} + b\, n\, U_{m,n-1} = 0, \\ U_{m,n+1} - \eta\, U_{m,n} + b\, m\, U_{m-1,n} + c\, n\, U_{m,n-1} = 0. \end{cases}$$

et l'on a de plus les relations classiques :

$$\overline{\int\int}_{+\infty}^{+\infty} e^{-\frac{1}{2}\varphi_0(x,y)}\, U_{m,n}\, V_{p,q}\, dx\, dy = 0 \quad \text{si } m \neq n \text{ et } p \neq q,$$

$$\overline{\int\int}_{+\infty}^{+\infty} e^{-\frac{1}{2}\varphi_0(x,y)}\, U_{m,n}\, V_{m,n}\, dx\, dy = m!\, n!\, (4\delta)^{m+n} \sqrt{\frac{\pi^2}{\delta}}.$$

Soit $f(x, y)$ la fonction de dépendance que l'on écrit sous la forme d'une série

$$f(x, y) = \Sigma\, A\, m, n\, e^{-1/2\,\varphi_0(x,y)}\, U_{m,n},$$

où les $A_{m,n}$ qui sont des constantes sont déterminées au moyen de la relation

$$A_{m,n} = \frac{\sqrt{\delta}}{m!\, n!\, (4\delta)^{m+n}\, \pi} \overline{\int\int}_{-\infty}^{+\infty} f(x, y)\, V_{m,n}\, dx\, dy.$$

Si la fonction $f(x, y)$ est continue et a un nombre limité de maxima et de minima, et si de plus à l'infini pour $x = \infty$ et $y = \infty$, elle est infiniment petite au moins du troisième ordre, la série $\Sigma A_{m,n} e^{-\frac{1}{2}\varphi_0} U_{m,n}$ est uniformément convergente et représente la fonction $f(x, y)$ comme l'a montré M. Muller-Lebedeff (Voir *Journal de Crelle*, vol. 66).

4) *La relation de récurrence relative à la loi de Bernoulli à deux variables.*

Nous avons étudié primitivement le schéma de Bernoulli à deux variables, en calculant tout d'abord la probabilité $f(x, y)$ qu'au cours de k extractions d'une urne renfermant des boules blanches, rouges et noires en proportions respectives $p, q\,(1—p—q)$, l'on constate la sortie de x boules blanches et y boules rouges, dans l'hypothèse où l'on remet la boule extraite après chaque tirage :

$$(1) \qquad f(x, y) = \frac{k!}{x!\, y!\, (k - x — y)!}\, p^x\, q^y\, (1 - p — q)^{k-x-y}.$$

M. Guldberg, poursuivant ses intéressantes recherches sur les critères permettant de reconnaître si une série statistique donnée F (x) peut être définie par l'une des fonctions suivantes de fréquence : fonction binomiale, fonction de Poisson, fonction de Pascal et fonction hypergéométrique a apporté une contribution nouvelle à ce problème (1).

Il remarque que l'équation (1) satisfait à l'équation aux différences finies :

$$(2)\quad f(x+1, y+1) = \frac{pq}{(1-p-q)^2} \frac{(k-x-y)(k-x-y-1)}{(y+1)(x+1)} f(x, y),$$

et que les moments

$$m_{f|g} = \mathrm{E}(x^f y^g) = \sum_i \sum_j x_i^f y_j^g f(x_i, y_j),$$

se calculent facilement en introduisant la fonction caractéristique :

$$\varphi(u, v) = \mathrm{E}(e^{ux+vy}) = \Sigma\Sigma f(x, y) e^{ux+vy},$$

et en utilisant la relation classique

$$m_{f|g} = \frac{\partial^{f+g}}{\partial u^f \partial v^g} [\varphi(u, v)], \text{ avec } u = v = 0.$$

Comme $\varphi(u, v)$ s'écrit :

$$\varphi(u, v) = \Sigma\Sigma \frac{k!}{x!\, y!\, (k-x-y)!} (pe^u)^x (qe^v)^y r^{k-y-x}$$

ou encore :

$$\varphi(u, v) = (pe^u + qe^v + r)^k, \text{ avec } r = 1-p-q,$$

il s'ensuit :

$$(3)\qquad m_{1\,0} = kp,\ m_{0|1} = kq,\ m_{1|1} = k(k-1)pq.$$

Si de ces deux dernières relations on tire k, p, q en fonction de $m_{1|0}$, $m_{0|1}$ et de $m_{1|1}$, on remarque que l'on peut remplacer l'équation (2) par l'équation

$$(4)\quad \left\{ \frac{[m_{0|1} m_{1|0} - m_{0|1}\Delta - m_{1|0}\Delta]^2}{m_{1|0} m_{0|1}} (x+1)(y+1) \frac{f(x+1, y+1)}{f(x, y)} + [2 m_{1,0} m_{0|1} - \Delta][x+y]\Delta - (x+y)^2 \Delta^2] \right\} = m_{1|0} m_{0|1} m_{1|1}.$$

avec :

$$\Delta = m_{1|0} m_{0|1} - m_{1|1}.$$

Désignant le premier membre de l'équation (4) par $\psi(x, y)$ on voit de suite que la fonction

$$(4')\qquad \alpha(x, y) \equiv \frac{\psi(x, y)}{m_{1|0} m_{0|1} m_{1|1}}$$

est égale à 1 pour toutes les valeurs de x et de y.

Une série à double entrée étant donnée, l'on calcule les trois moments $m_{1|0}$, $m_{0|1}$, $m_{1|1}$, et l'on forme l'expression $\alpha(x, y)$; si cette dernière expression — pour les différentes valeurs du système (x, y) — est voisine de 1, on peut dire que la série expérimentale à l'étude peut être définie par la loi de Bernouilli à deux variables.

---

(1) Voir note aux Comptes rendus de l'Académie des Sciences de M. Guldberg (Séance du 29 mai 1933).

Rappelons enfin que si la série à double entrée est exactement du type de Bernouilli, on sait que les moments successifs doivent satisfaire à certains critères; c'est ainsi que les moments du second ordre sont liés aux moments du premier ordre par les relations :

$$m_{2|0}\, m_{0|1} - m_{1|0}\, m_{0|1} = m_{1|1}\, m_{1|0};$$
$$m_{0|2}\, m_{1,0} - m_{0|1}\, m_{1|0} = m_{1|1}\, m_{0|1}.$$

Quant au coefficient de corrélation, il est égal à :

$$\frac{m_{1|1} - m_{1|0}\, m_{0,1}}{\sqrt{m_{2|0} - m_{1|0}^2}\sqrt{m_{0|2} - m_{0|1}^2}}, \text{ ou } \sqrt{\frac{pq}{(1-p)(1-q)}}$$

si l'on tient compte des relations (3) et des valeurs de $m_{2|0}$ et $m_{0|2}$.

Un processus analytique analogue à celui que nous venons d'exposer peut être appliqué à l'étude des lois de Poisson, de Pascal, et de la loi hypergéométrique à deux variables.

5) *Corrélation normale dans le cas de plusieurs variables.*

Considérons tout d'abord le cas où la fonction de probabilité est définie par une forme quadratique à 3 variables, et remarquons que si le discriminant de cette forme n'est pas nul et si les racines de l'équation en S sont toutes trois positives, l'on peut — après un changement de variables — mettre la loi de probabilité sous la forme :

$$f(x_1, x_2, x_3) = \mathrm{Ce}^{-\frac{1}{2}\left(\frac{x_1^2}{\sigma_1^2} + \frac{x_2^2}{\sigma_2^2} + \frac{x_3^2}{\sigma_3^2}\right)}$$

où x_1, x_2, x_3 sont indépendantes; quant à la constante C, on en calcule la valeur au moyen de la relation :

$$\int_{-\infty}^{+\infty}\int\int f(x_1, x_2, x_3)\, dx_1\, dx_2\, dx_3 = 1,$$

qui, dans le cas actuel, nous donne :

$$\mathrm{C} = \frac{1}{(2\pi)^{\frac{3}{2}}\,\sigma_1\,\sigma_2\,\sigma_3} = \frac{\sqrt{\mathrm{D}}}{(2\pi)^{\frac{3}{2}}}$$

car le discriminant de la forme

$$\Sigma \frac{x_i^2}{\sigma_i^2} \text{ a pour valeur } \mathrm{D} = \frac{1}{\sigma_1^2\,\sigma_2^2\,\sigma_3^2}$$

Ceci étant, considérons la forme quadratique :

$$\mathrm{H}(x_1, x_2, x_3) = a_{11}\, x_1^2 + a_{22}\, x_2^2 + a_{33}\, x_3^2 + 2\, a_{2,3}\, x_2\, x_3 + 2\, a_{3,1}\, x_3\, x_1 + 2\, a_{1,2}\, x_1\, x_2,$$

son discriminant D :

$$\mathrm{D} = \begin{vmatrix} a_{11}, & a_{12}, & a_{13} \\ a_{21}, & a_{22}, & a_{23} \\ a_{31}, & a_{32}, & a_{33} \end{vmatrix}, \text{ et la loi de dépendance } \varphi = \mathrm{Ce}^{-\frac{\mathrm{H}}{2}}.$$

Dans un espace à trois dimensions, la surface H = 1 devient après un changement approprié d'axes :

$$\frac{X_1^2}{\sigma_1^2} + \frac{X_2^2}{\sigma_2^2} + \frac{X_3^2}{\sigma_3^2} = 1,$$

et par suite, du fait que dans ladite transformation le discriminant est un invariant, on remarque que :

$$f = \frac{\sqrt{D}}{(2\pi)^2}\, e^{-\frac{1}{2}\left[\Sigma\, a_{ii}\, x_i^2 + 2\,\Sigma\, a_{ij}\, x_i\, x_j\right]};$$

quant à la fonction caractéristique, elle n'est autre que l'espérance mathématique de

$$e^{u_1 x_1 + u_2 x_2 + u_3 x_3},$$

soit :

$$\varphi(u_1, u_2, u_3) = \frac{+\infty}{\iiint}_{-\infty} f(x_1, x_2, x_3)\, e^{u_1 x_1 + u_2 x_2 + u_3 x_3}\, dx_1\, dx_2\, dx_3 = e^{\frac{1}{2} K(u_1, u_2, u_3)}$$

avec :

$$K(u_1, u_2, u_3) = \sigma_1^2 u_1^2 + \sigma_2^2 u_2^2 + \sigma_3^2 u_3^2$$

comme l'on s'en rend compte facilement en utilisant le passage des coordonnées cartésiennes aux coordonnées tangentielles, car si H = 1 est l'équation ponctuelle de la quadrique, K = 1 est celle de son équation tangentielle, et K est dite la forme adjointe de H.

Pour cela, au système x_1, x_2, x_3, substituons un nouveau système trirectangulaire X_1, X_2, X_3 formé par les axes de la quadrique; dans ce cas, l'intégrale triple ci-dessus prend la forme :

$$\iiint e^{-\frac{1}{2}\left(\frac{X_1^2}{\sigma_1^2} + \frac{X_2^2}{\sigma_2^2} + \frac{X_3^2}{\sigma_3^2}\right)} e^{U_1 X_1 + U_2 X_2 + U_3 X_3}\, dX_1\, dX_2\, dX_3 = e^{\frac{1}{2}(\sigma_1^2 U_1^2 + \sigma_2^2 U_2^2 + \sigma_3^2 U_3^2)}$$

On pourra donc, par une transformation inverse, passer de la forme adjointe en u_1, u_2, u_3, à la forme quadratique en x_1, x_2, x_3.

Soit en effet :

$$\mathcal{K} = \alpha_{11} u_1^2 + \alpha_{22} u_2^2 + \alpha_{33} u_3^2 + 2\,\alpha_{23} u_2 u_3 + 2\,\alpha_{31} u_3 u_1 + 2\,\alpha_{12} u_1 u_2,$$

où les α_{ij} représentent les moments du second ordre de la distribution :

$$\mu_{2,0,0},\ \mu_{0,2,0},\ \mu_{0,0,2},\ \mu_{0\,1,1},\ \mu_{1,0,1},\ \mu_{1,1,0}$$

que nous désignerons par :

$$\sigma_1^2,\ \sigma_2^2,\ \sigma_3^2,\ \sigma_2\,\sigma_3\, r_{2,3},\ \sigma_3\,\sigma_1\, r_{3,1},\ \sigma_1\,\sigma_2\, r_{1,2}.$$

Si maintenant nous prenons comme unités de longueur dans le système d'axes x_1, x_2, x_3, les grandeurs σ_1, σ_2, σ_3, on voit que la forme réduite $\overline{\mathcal{K}}$ s'écrit :

$$\overline{\mathcal{K}} = u_1^2 + u_2^2 + u_3^2 + 2\, r_{2,3}\, u_2 u_3 + 2\, r_{3,1}\, u_3 u_1 + 2\, r_{1,2}\, u_1 u_2.$$

La théorie des enveloppes nous fournit le système des équations :

$$\frac{\partial \overline{\mathcal{K}}}{\partial u_1} = \frac{\partial \overline{\mathcal{K}}}{\partial u_2} = \frac{\partial \overline{\mathcal{K}}}{\partial u_3} = 1$$

$$x_1 \qquad x_2 \qquad x_3$$

et, par suite :

$$\overline{\mathcal{K}}(u_1, u_2, u_3) = \Sigma\, u_i\, x_i.$$

Après avoir remarqué que le discriminant Δ de la forme réduite $\overline{\mathcal{K}}$ est lié au déterminant D de la forme $\overline{\mathcal{H}}$ par la relation $D\,\Delta = 1$, avec :

$$\Delta = (1 - r_{2,3}^2)(1 - r_{1,3}^2) - (r_{1,2} - r_{2,3}\, r_{1,3})^2 \text{ et } r_{ij} = r_{ji},$$

l'on trouve :

$$\overline{\mathcal{H}} = \frac{1}{\Delta}[(1 - r_{23}^2)\, x_1^2 + (1 - r_{31}^2)\, x_2^2 + (1 - r_{12}^2)\, x_3^2 + 2\,(r_{23} - r_{12}\, r_{13})\, x_2\, x_3$$
$$+ 2\,(r_{31} - r_{23}\, r_{21})\, x_1\, x_3 + 2\,(r_{12} - r_{31}\, r_{32})\, x_1\, x_2]$$

et

$$f(x_1, x_2, x_3) = \frac{1}{\sqrt{\Delta}\,(2\pi)^{3/2}}\, e^{-\frac{\mathcal{K}}{2}}$$

si l'on mesure toujours x_1, x_2, x_3 avec les unités respectives σ_1, σ_2, σ_3.

6) *Plans de régression. — Coefficients de corrélation partielle.*

Si l'on fait usage de la méthode classique pour la décomposition de la forme quadratique $\overline{\mathcal{H}}$, on trouve que $f dx_1$, dx_2, dx_3 peut s'écrire :

$$\left(\frac{1}{\sqrt{2\pi}}\sqrt{\frac{1 - r_{23}^2}{\Delta}}\, e^{-\frac{(1 - r_{23}^2)}{2\Delta} x_1^2} dx_1\right)\left(\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{1 - r_{23}^2}}\, e^{-\frac{1}{2(1 - r_{23}^2)}(x_2^2 + x_3^2 - 2\, r_{23} x_2 x_3)} dx_2 dx_3\right)$$

ou encore :

$$(A\, d\, X_1)\,(B\, dx_2\, dx_3),$$

avec :

$$X_1 = x_1 - \frac{r_{12} - r_{13}\, r_{23}}{1 - r_{23}^2}\, x_2 - \frac{r_{31} - r_{32}\, r_{12}}{1 - r_{23}^2}\, x_3;$$

Il s'ensuit que si l'on prend le rapport des masses situées dans le parallélipipède élémentaire limité par $[x_1, x_1 + dx_1]$ $[x_2, x_2 + dx_2]$ $[x_3, x_3 + dx_3]$, aux masses comprises dans le parallélipipède de base $(x_2, x_2 + dx_2)$ $(x_3, x_3 + dx_3)$ ayant pour hauteur le champ de variation de x_1, on obtient alors la probabilité liée de x_1 en (x_2, x_3), soit :

$$\frac{1}{\sqrt{2\pi}}\sqrt{\frac{1 - r_{23}^2}{}}\, e^{-\frac{1 - r_{23}^2}{2\Delta} X_1}\, dX.$$

Il résulte de l'expression ci dessus de la probabilité liée que la distribution de x, qui est du type normal, a une dispersion caractérisée par $\sqrt{\frac{\Delta}{1 - r_{23}^2}}$; *de plus, la dispersion étant constante dans toutes les files parallèles à* ox_1, ON VOIT QU'IL Y A HOMOSCÉDASTICITÉ.

On remarque enfin qu'à $X_1 = 0$ correspond le plan

$$x_1 = \frac{r_{12} - r_{13}\, r_{23}}{1 - r_{23}^2}\, x_2 + \frac{r_{31} - r_{32}\, r_{12}}{1 - r_{23}^2}\, x_3,$$

qui n'est autre que le lieu des centres de gravité des files parallèles à Ox_1.

A la loi de distribution normale $f(x_1, x_2, x_3)$ on rattache donc trois plans de régression : de x_1 *en* x_2 et x_3, de x_2 en x_3 et x_1, de x_3 en x_1 et x_2, dont les équations sont définies respectivement par $\frac{\partial \overline{\mathcal{H}}}{\partial x_i} = 0$, avec $(i = 1, 2, 3)$ et représentent les plans diamétraux conjugués des directions ox_1, ox_2, ox_3 dans la quadrique $\overline{\mathcal{H}} = 0$. Ces équations peuvent s'écrire :

$$\left.\begin{array}{ll} \text{Plan } P_1 & x_1 = b_{12,3}\, x_2 + b_{13,2}\, x_3 \\ \text{Plan } P_2 & x_2 = b_{21,3}\, x_1 + b_{23,1}\, x_3 \\ \text{Plan } P_3 & x_3 = b_{31,2}\, x_1 + b_{32,1}\, x_2 \end{array}\right\} \quad \text{avec} \quad \left.\begin{array}{l} b_{12,3} = \frac{r_{12} - r_{13}\, r_{23}}{1 - r_{23}^2},\ b_{13,2} = \frac{r_{13} - r_{21}\, r_{23}}{1 - r_{23}^2} \\ b_{21,3} = \frac{r_{21} - r_{13}\, r_{23}}{1 - r_{13}^2},\ b_{23,1} = \frac{r_{23} - r_{12}\, r_{13}}{1 - r_{13}^2} \\ b_{31,2} = \frac{r_{31} - r_{23}\, r_{21}}{1 - r_{12}^2},\ b_{32,1} = \frac{r_{32} - r_{13}\, r_{21}}{1 - r_{12}^2} \end{array}\right.$$

Considérons l'ensemble des masses situées entre les plans $x_3 = c$, et $x_3 + dx_3$, ainsi que le plan P_1, lien des centres des files parallèles à ox_1 ; il est évident que le lieu des centres de gravité des portions de files parallèles à ox_1, situées à l'intérieur de la tranche $(x_3, x_3 + dx_3)$ n'est autre que la droite D_1 d'intersection du plan P_1 avec le plan $x_3 = c$, et le lieu des centres de gravité des portions de files parallèles à ox_2 situées dans la tranche susvisée est la droite D_2 intersection de P_2 avec $x_3 = c$.

D_1 et D_2 se rencontrent en un point G qui est le centre de gravité de la tranche; *de plus ces droites représentent les lignes de dépendance de* x_2 *par rapport à* x_1, *et de* x_1 *par rapport à* x_2.

Or les équations de ces lignes (lignes de régression) peuvent s'écrire pour une valeur constante de x_3.

$$x_1 = b_{12,3}\, x_2 + \gamma_1$$
$$x_2 = b_{21,3}\, x_1 + \gamma_2.$$

Il s'ensuit que le COEFFICIENT DE CORRÉLATION *caractérisant la dépendance de* $(x_1\ x_2)$ *a pour valeur :*

$$r_{12,3} = \sqrt{b_{12,3}\, b_{21,3}} = \frac{\sqrt{(r_{12} - r_{13}\, r_{23})\,(r_{12} - r_{13}\, r_{23})}}{\sqrt{1 - r_{23}^2}\quad \sqrt{1 - r_{13}^2}}$$

et l'on démontre que le coefficient de corrélation est représenté par :

$$\frac{+\ (r_{12} - r_{13}\, r_{23})}{+\sqrt{(1 - r_{23}^2)\,(1 - r_{13}^2)}}$$

Coefficient de corrélation totale.

On considère le parallélipipède de base $(x_2, x_2 + dx_2)$ $(x_3, x_3 + dx_3)$ de hauteur indéfinie parallèle à ox_1, et l'on suppose que l'on concentre toutes les masses contenues à l'intérieur dudit volume sur la base (dx_2, dx_3); on a ainsi

une répartition des masses dans le plan (x_2, x_3), *à laquelle on rattache le coefficient* r_{23}, *dit* COEFFICIENT DE CORRÉLATION TOTALE.

Si maintenant on ne fait intervenir que la portion de files parallèles à Ox_1, et appartenant au plan $x_3 = c^{te}$ (ou plus exactement les files comprises entre x_3 et $x_3 + dx_3$), on voit alors apparaître à la limite un coefficient de corrélation, caractérisant la distribution dans le plan $x_3 = c^{te}$, qui n'est autre que $r_{12,\,3}$, que l'on désigne par coefficient de CORRÉLATION PARTIELLE.

Il y a lieu de faire ici une remarque fort curieuse au sujet de ces coefficients de corrélation totale et de corrélation partielle basée sur l'emploi de la trigonométrie sphérique; on démontre en effet qu'*aux coefficients de corrélation totale représentant les cosinus des faces d'un certain trièdre, l'on rattache les coefficients de corrélation partielle qui sont les cosinus des angles dièdres.*

Remarque. — Le lecteur désireux d'étudier les caractéristiques de la surface normale de dispersion dans un espace à n dimensions, et celles des plans de régression qui lui correspondent, devra se reporter aux travaux de Karl Pearson; il devra également, en ce qui concerne l'examen des coefficients de corrélation partielle — lire les mémoires d'Edgeworth, et le bel exposé de Yule basé sur l'élargissement de la notion du coefficient de corrélation dans le cas de deux variables qui repose sur une application fort judicieuse de la méthode des moindres carrés.

Comme le fait remarquer si justement Karl Pearson, « il ne faut pas plus faire un fétiche de la méthode des moindres carrés que de la distribution normale » (*Notes on the History of Correlation*, *Biometrika*, vol. 13, 1920), car, pour justifier l'emploi de la méthode des moindres carrés à l'adoption d'une ligne ou d'un plan à un ensemble de points, nous devons admettre que les groupes suivent une distribution normale.

Rappelons que si ${}_1X_i, {}_2X_i, \ldots\ldots {}_NX_i$, sont les N mesures d'une grandeur X_i, et les écarts ${}_{(j)}X_i$ par rapport à sa moyenne M_i sont faibles, l'on est conduit en effet à résoudre n systèmes analogues au suivant, composé de N équations du premier degré.

$$ {}_{(j)}x_1 = b_{12,34\ldots\ldots n}\,{}_{(j)}x_2 + b_{13,24\ldots\ldots n}\,{}_{(j)}x_3 + \ldots\ldots + b_{1n,\,23\ldots\ldots n-1}\,{}_{(j)}x_n $$

avec $(j = 1, 2, \ldots$ N) d'où l'on déduit en ayant recours à une méthode de multiplicateurs (et en particulier à la méthode Cauchy) les coefficients :

$$ (b_{1k,23\ldots\ldots\overline{k-1}\,\overline{k+1}\ldots n}). $$

Il est évident que ce mode de calcul est moins synthétique et moins rapide que celui de Yule, et aussi que celui de Karl Pearson qui est basée sur le calcul préalable des r_{ij}, puis sur celui des déterminants mineurs R_{st} déduits du déterminant classique R

$$ R = \begin{vmatrix} 1, & r_{12}, & r_{13}\ldots\ldots & r_{1n} \\ r_{21}, & 1, & r_{23}\ldots\ldots & r_{2n} \\ r_{n1}, & r_{n,2}, & r_{n,3}\ldots\ldots & 1 \end{vmatrix} \quad (\text{avec } r_{st} = r_{ts}), $$

par la suppression dans ce dernier de la $s^{ème}$ ligne et de la $t^{ème}$ colonne; on trouve ainsi que les équations des plans de régression s'écrivent ainsi qu'il suit :

$$\frac{x_\lambda \, R_{\lambda\lambda}}{\sigma_\lambda} = - \sum \frac{R_{\lambda j}}{\sigma_j} x_j, \qquad \begin{matrix} \text{avec } j = (1,2 \ldots\ldots \lambda - 1, \lambda + 1, \ldots\ldots n) \\ \text{et } \lambda = (1,2, \ldots\ldots n), \end{matrix}$$

et où σ_j est l'écart quadratique relatif aux écarts ${}_{(1)}x_j \ldots\ldots {}_{(N)}x_j$.

Toutefois, nous avons pu nous rendre compte que le calcul des coefficients b, effectué au moyen de la méthode Cauchy est assez rapide et permet surtout de réaliser de nombreuses vérifications au cours des opérations.

R. Risser.