

JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

R. RISSER

Les principes de la statistique mathématique

Journal de la société statistique de Paris, tome 78 (1937), p. 40-74

http://www.numdam.org/item?id=JSFS_1937__78__40_0

© Société de statistique de Paris, 1937, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

II

LES PRINCIPES DE LA STATISTIQUE MATHÉMATIQUE

TROISIÈME PARTIE

Retour sur les coefficients de covariation.

Courbes moyennes et utilisation de la théorie des différences ⁽¹⁾.

CHAPITRE I

Dans une étude antérieure, nous avons signalé tout d'abord l'indice de Fechner, puis le coefficient de dépendance, et enfin le coefficient de covariation, et montré à l'occasion du calcul de ces indices qu'il ne fallait pas s'en tenir aux variations entre éléments contigus, mais faire intervenir les variations entre éléments distants d'intervalles plus ou moins étendus.

Variations de faible et de grande amplitudes.

Les phénomènes statistiques que l'on est conduit à analyser et comparer peuvent subir des variations se rattachant à des phases diverses (mois, année, groupe d'années, décade, siècle, etc.); dans certains cas, on s'attache essentiellement à la recherche du coefficient de dépendance des variations annuelles, dans d'autres l'on se préoccupe plutôt des coefficients correspondant à des périodes décennales ou même à des périodes plus étendues.

Dans l'évolution des produits des récoltes d'un certain district d'un territoire, il y a lieu de distinguer d'une part une composante séculaire, représentée par une courbe continue et destinée à refléter l'influence des progrès de la technique agricole, et, d'autre part, une composante irrégulière et oscillante qui est une résultante des conditions météorologiques; l'examen des courbes des prix

(1) Les études qui font l'objet des deuxième et troisième parties de cette communication ont été établies en recourant à toute une série de documents français et étrangers; pour la traduction de ces derniers, j'ai été fort aidé dans ma tâche par certains de mes élèves et anciens élèves du Conservatoire des Arts et Métiers M^e Raut, membre diplômé de l'Institut des Actuaire français, MM. Dédé, ingénieur civil des Mines; Grosver, ancien élève de l'École polytechnique, et aussi par MM. Simionov, actuaire; G. Manuel, ancien élève de l'École polytechnique, et J. Dennery, ingénieur civil des Mines, et tiens à leur exprimer ici toute ma reconnaissance. Je suis heureux de renouveler ici mes remerciements à MM. Marais, ancien élève de l'École normale supérieure, agrégé de philosophie et membre agrégé de l'Institut des Actuaire français; Ulmer, statisticien à la Statistique générale de la France, ancien élève de l'École polytechnique; Fessard, directeur adjoint de laboratoire à l'École des Hautes Études, et Francis Netter, ancien élève de l'École polytechnique, actuaire à la Direction des Assurances sociales, qui nous ont apporté il y a quelques années déjà, à M. Traynard et à moi, leur concours si efficace lors de l'élaboration de notre ouvrage : *Les principes de la statistique mathématique.*

conduit à distinguer tout d'abord la composante séculaire, puis une courbe dite de conjoncture en forme de vague, enfin une composante dite saisonnière, et en dernier lieu une composante accidentelle.

On ne doit pas baser ses conclusions sur des valeurs des coefficients de covariation relatifs respectivement à des variations quinquennales ou décennales, car ce serait tenir comme négligeables les variations annuelles, considérées comme accidentelles, à moins que l'on puisse prouver que ces variations annuelles sont bien accidentelles.

Rappelons à ce propos que nous avons mis en évidence les valeurs des coefficients de dépendance, suivant que l'on opère par périodes décennales ou par périodes annuelles, en comparant les mariages et les naissances observés en Angleterre de 1851 à 1901 (Voir tableaux figurant, p. 258, 259 et 261, dans l'ouvrage : *Les principes de la statistique mathématique* de RISSER et TRAYNARD, et tirés de l'étude de MARCH : *La représentation graphique et la statistique comparative* — qui a fait l'objet d'une très intéressante communication à la Société de Statistique de Paris [novembre 1904 et janvier 1905]).

A côté du coefficient de Fechner qui ne saurait indiquer que la covariation différentielle, il y a lieu d'introduire avec March et d'autres statisticiens les coefficients de covariation r et j :

r (coefficient de covariation tendancielle)

$$r = \frac{\sum y_i' Y_i'}{\sqrt{(\sum y_i'^2) (\sum Y_i'^2)}}$$

où y_i' et Y_i' représentent les écarts respectifs des grandeurs y_i et Y_i , relatives aux phénomènes P_1 et P_2 par rapport à leurs moyennes :

$$\sum \frac{y_i}{n}, \sum \frac{Y_i}{n};$$

j (coefficient de covariation différentielle)

$$j = \frac{\sum (y_{i+1} - y_i) (Y_{i+1} - Y_i)}{\sqrt{\sum (y_{i+1} - y_i)^2 \cdot \sum (Y_{i+1} - Y_i)^2}}$$

qui rentrent dans le cadre de ceux présentés dans notre précédente étude (Voir chapitre de la corrélation).

A ce propos, March signale que l'on peut assez facilement réaliser des graphiques (y_i, Y_i) faisant apparaître (1) :

a) Coefficient de covariation tendancielle voisin de + 1, coefficient de covariation différentielle voisin de - 1;

b) Coefficient de covariation tendancielle voisin de + 1, coefficient de covariation différentielle voisin de 0;

c) Coefficients de covariation tendancielle et différentielle égaux tous deux à - 1;

d) Coefficients r et j , tous deux nuls ou voisins de 0.

La covariation tendancielle, caractéristique de la liaison de deux séries statistiques a suscité de nombreuses controverses, qui tiennent très probable-

(1) Différences et corrélation en statistique (*Journal de la Société de Statistique de Paris*, février 1928).

ment à ce que le problème que l'on se proposait de résoudre n'était pas toujours posé d'une manière précise; de plus, il suffira de remarquer que la liaison entre deux séries statistiques, soit à travers le temps, soit à travers l'espace, ou à travers ces deux facteurs, qui semblait justifiée par un coefficient de covariation d'une valeur élevée (c'est-à-dire voisine de 1), ne l'était nullement en réalité.

Aucun exemple n'est plus typique à cet égard que celui donné par M. Yule (2), et relatif à la relation entre la mortalité en Angleterre et la proportion des mariages au cours des années de la période 1866-1911, qui conduit à la valeur 0,9512 pour r . M. Yule fait d'ailleurs remarquer à cette occasion que si une formule appliquée à un cas particulier fournit un résultat qui — pour des raisons de bon sens — est jugé inacceptable et incorrect, il convient d'examiner les diverses hypothèses qui ont servi de base à l'établissement de la formule, et de discriminer toutes celles qui ne s'appliquent pas au cas étudié; grâce d'ailleurs à ce processus, il montre qu'il n'y a pas de lien de causalité entre les deux phénomènes précités, et cela malgré le coefficient 0,95 de covariation.

Alors même que l'on se conformerait à la méthode des variations concomitantes, comme le conseillait Edgeworth, lorsqu'il répondait au discours présidentiel précité de M. G. Yule, il faut avoir soin de recourir à la méthode des différences, sans oublier toutefois que celle-ci présente des difficultés particulières, et que son utilisation logique doit être confiée à des statisticiens de qualité.

Du coefficient j basé sur les différences premières, l'on est passé — comme nous le verrons plus loin — au produit de deux différences d'ordre n , puis l'on s'est préoccupé des séries à liaisons internes dont l'étude est plus complexe.

On se rend compte ainsi des étapes parcourues depuis l'apparition de l'indice de dépendance de Fechner, et de l'évolution des idées en matière de covariation et de corrélation; c'est cette évolution que nous allons tenter d'analyser à travers les travaux les plus marquants des statisticiens.

Classification des séries statistiques d'après Anderson.

Anderson fait observer qu'une série s'étendant sur une très longue période peut, ainsi que nous l'avons vu, être considérée comme la somme d'une série fondamentale continue, et d'une composante résiduelle, tantôt positive, tantôt négative, propre à certains éléments.

Graphiquement, une série séculaire est représentée de la manière suivante : la courbe A B définit la série continue fondamentale, et le tracé M N la composante résiduelle, discontinue que l'on peut aussi désigner par série épiséculaire.

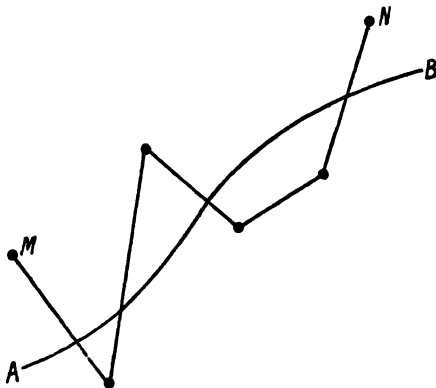
Suivant la nature du problème que nous aurons à résoudre, nous devons porter notre attention, tantôt sur l'étude de la série A B, tantôt sur celle de la série M N.

Si l'on est parvenu à isoler la composante résiduelle et s'il y a compensation entre la somme des éléments négatifs de la suite en question et la somme des éléments positifs, on est conduit avec Anderson à envisager trois espèces de suites MN.

(1) Voir discours présidentiel de M. G. Yule pour la session 1925-1926, prononcé à la Société royale de statistique, le 17 novembre 1925, à Londres.

La première correspond au cas où à un élément positif de la suite fait suite immédiatement un élément négatif plus souvent qu'un positif, et à un élément négatif un élément positif plus souvent qu'un négatif, le graphique de la série nous apparaît dentelé.

La deuxième espèce est relative au cas où à un élément positif fait suite un élément négatif aussi souvent qu'un positif et *vice versa*; le graphique a encore un aspect dentelé; mais toutefois moins prononcé que celui correspondant à la première espèce.



La troisième espèce est afférente au cas où à un élément positif fait suite le plus souvent un positif, à un négatif le plus souvent un négatif.

La distinction de ces trois espèces permet dans l'esprit de M. Anderson de mieux comprendre la classification des séries en quatre groupes :

Groupe I. — En dehors des séries sinusoidales à courtes périodes, séries étudiées par Yule, appartiennent au groupe I certains autres types de séries — dites du type « (Z) » — dont le caractère commun, qui par lui-même n'est pas suffisant, consiste en une corrélation négative entre les termes de la série se suivant immédiatement.

Groupe II. — A ce groupe appartiennent non seulement les séries oscillantes dont les éléments peuvent être définis comme des grandeurs empiriques que prend une variable avec une loi de distribution constante dans des épreuves indépendantes, mais encore les suites uniformes de Tschuprow; toutes ces séries rentrent dans le type (R).

Groupe III. — Dans ce groupe figurent tout d'abord des suites paraboliques d'ordre supérieur et des séries sinusoidales avec correspondance positive d'éléments voisins, puis d'autres types, dont la caractéristique commune, qui en elle-même n'est point suffisante, consiste en une corrélation positive des éléments voisins; ces séries appartiennent, d'après Anderson, au type (G).

Groupe IV. — A ce groupe sont rattachées des séries, dites du type (S), qui se composent comme des séries des trois premiers groupes lorsque l'on ne considère que les différences d'ordre relativement peu élevé, puis changent de groupe lorsque l'ordre des différences continue à croître.

Anderson s'est attaché tout particulièrement à l'étude des séries du groupe II, en faisant appel à la méthode des différences.

De l'utilisation du calcul des différences. — Les recherches de Student et d'Anderson.

Avant d'analyser les mémoires de Student et d'Anderson, nous allons procéder à l'étude succincte de séries représentées d'une manière approximative par des droites, alors que les séries statistiques ne le sont que par des figures polygonales plus ou moins complexes.

Considérons deux joueurs A et B lançant l'un un dé blanc et l'autre ensuite un dé rouge, à des intervalles de temps égaux (la seconde, par exemple).

A qui joue le premier, note le point α_1 ; B joue ensuite et note non le point β_1 , marqué par son dé, mais la somme $\alpha_1 + \beta_1$; A joue ensuite, note le point α_2 et marque le point $\alpha_1 + \alpha_2$, puis B joue, note le point β_2 et marque $\alpha_1 + \beta_1 + \alpha_2 + \beta_2$, et ainsi de suite.

Or on sait que les espérances mathématiques de α , et de $(\alpha + \beta)$ ont pour valeurs respectives (3,5) et 7; il résulte de là qu'aux droites $y_0 = 3,5 t$, $Y_0 = 7 t$ (où t est le temps), on peut faire correspondre les différences $(y - y_0, Y - Y_0)$, qui ne sont autres que des variables aléatoires dont la valeur probable est nulle, et dont la loi de probabilité varie avec t (1).

Aux temps 1, 2, 3,..... t , ces différences auront respectivement pour valeurs :

$$(\xi_1, \xi_2, \dots, \xi_t), (\tau_1, \tau_2, \dots, \tau_t).$$

Considérons maintenant les variables aléatoires :

$$u_i = \alpha_i - 3,5, \quad v_i = \beta_i - 3,5,$$

auxquelles l'on rattache

$$\xi_j = u_1 + u_2 + \dots + u_j, \quad \tau_j = u_1 + u_2 + \dots + u_j + v_1 + v_2 + \dots + v_j, \quad \text{(avec } j = 1, 2, 3, \dots, t);$$

Il est évident que les valeurs probables de ξ_t et τ_t sont toujours nulles quel que soit le temps t , alors que leurs écarts types varient comme \sqrt{t} .

On remarque que les variables ξ_t et ξ_{t+1} sont en liaison stochastique, et aussi que ξ_1 est liée à toutes les quantités ξ_t .

Si E est le symbole de l'espérance mathématique, l'on a évidemment :

$$E(\xi_1 \xi_t) = E(u_1^2) = \mu_2, \quad E(\xi_t^2) = \mu_2, \quad E(\xi_t^2) = t \mu_2,$$

et, par suite, le coefficient de corrélation entre ξ_1 et ξ_t a pour valeur :

$$\frac{\mu_2}{\sqrt{E(\xi_1^2) E(\xi_t^2)}} = \frac{1}{\sqrt{t}}.$$

A la série des ξ_t , on peut — grâce à l'introduction des différences premières —, substituer des variables indépendantes u_1, u_2, \dots, u_t constituant une série oscillante; on formerait de même la série de terme général.

$$\tau_t - \tau_{t-1} = u_t + v_t.$$

(1) Voir : la question d'égalité des joueurs à pile ou face (*Calcul des probabilités*, de BOREL); l'allure d'une longue série dans le jeu de pile ou face (*Le Hasard*, de BOREL); et aussi : DARMOIS, *Statistique mathématique*.

Compte tenu de ce que

$$E(\xi_i \xi_i) = t \mu_2, \quad E(\xi_i^2) = t \mu_2, \quad E(\eta_i^2) = 2 t \mu_2,$$

le coefficient de corrélation ρ' entre les termes de même rang ξ_i et η_i des séries ξ et η a pour valeur $\rho' = \frac{1}{\sqrt{2}}$; quant au coefficient de corrélation ρ'' caractérisant les termes ξ_{t_1} et η_{t_2} avec $t_2 > t_1$, il a pour valeur :

$$\rho'' = \frac{t_1 \mu_2}{\sqrt{t_1 \mu_2} (2 t_2 \mu_2)} = \frac{1}{\sqrt{2}} \sqrt{\frac{t_1}{t_2}}$$

et est d'autant plus faible que le rang t_2 s'éloigne plus du rang t_1 .

Cet exposé, qui met en évidence pour les courbes moyennes des deux séries des droites, fait apparaître un coefficient de corrélation constant entre les termes de même rang des séries ξ_i et η_i , et un coefficient de valeur $\sqrt{\frac{t_1}{2 t_2}}$ entre les termes ξ_{t_1} et η_{t_2} ; l'intervention des différences

$$\Delta \xi_i = u_i, \quad \Delta \eta_i = u_i + v_i,$$

qui nous conduit à un coefficient de corrélation constant $\frac{1}{\sqrt{2}}$ entre $\Delta \xi_{t_1}$ et $\Delta \eta_{t_1}$, et égal à zéro pour $\Delta \xi_{t_1}$ et $\Delta \eta_{t_2}$, met en lumière des séries oscillantes où, à travers chacune d'elles, n'existe plus aucune liaison interne.

Corrélation partielle relative aux différences premières.

Hooker avait en 1905 (1) préconisé une méthode en vue de la détermination de la corrélation des variations de la moyenne instantanée, en calculant la corrélation partielle relative aux différences premières.

Rappelons que March, dans la communication à laquelle il a été fait allusion ci-dessus (2), ayant constaté que la comparaison directe M_x du nombre des mariages célébrés en Angleterre au cours de l'année x de la période 1851-1901, avec les nombres des naissances N_{x+1}, \dots, N_{x+4} survenus dans les années $(x+1), \dots, (x+4)$ de la même période, ne permettait pas de mettre en lumière la répercussion du phénomène P_1 sur le phénomène P_2 , avait été conduit à substituer au tracé polygonal tant pour P_1 que pour P_2 une courbe moyenne, c'est-à-dire tracée de telle façon que la somme des écarts du graphique par rapport à la courbe interpolée fût nulle, et à calculer le coefficient de dépendance :

1° entre les ordonnées des courbes interpolées;

2° entre les écarts des mesures observées par rapport aux ordonnées des courbes moyennes.

Ce deuxième coefficient de dépendance peut en définitive être rapproché de celui considéré par Miss F. E. Cave en 1904 (3).

Alors que l'utilisation des nombres bruts fournis par le calcul du coefficient

(1) Voir : *Journal of the Royal Statistical Society*, vol. LXVIII (p. 396 et suiv.).

(2) La représentation graphique et la statistique comparative (Communication à la Société de Statistique de Paris, novembre 1904-janvier 1905).

(3) *Royal Society Proceedings* (vol. LXXIV, p. 407 et suiv.).

de liaison de deux séries conduisait à un chiffre ne pouvant nullement servir à l'interprétation de cette liaison, la méthode de Hooker a pu être appliquée avec succès à l'étude des problèmes économiques.

Le professeur Pearson ayant fait remarquer que la méthode susvisée n'était valable qu'en cas de connexion linéaire entre les variables et le temps, Student s'est efforcé de l'étendre grâce à un processus analytique simple (4).

$x_1, x_2, \dots, x_i, \dots, y_1, y_2, \dots, y_i, \dots$, étant des valeurs correspondantes des variables x et y , valeurs distribuées au hasard dans l'espace et dans le temps, il est facile de voir que la corrélation entre les $n^{\text{ièmes}}$ différences correspondantes est la même que celle relative aux couples (x, y) .

On suppose d'une part que x_i et y_i sont deux variables aléatoires, et d'autre part qu'il n'y a aucune corrélation entre x_i et x_j , entre y_i et y_j , entre x_i et y_j , mais corrélation entre les éléments de même rang x_i et y_i .

Formons les différences premières :

$$\Delta^1 x_i = x_{i+1} - x_i, \quad \Delta^1 y_i = y_{i+1} - y_i,$$

et rappelons que l'espérance mathématique des différences premières est nulle.

$$\begin{aligned} E(x_{i+1} - x_i) &= 0 & E(y_{i+1} - y_i) &= 0, \\ E(x_{i+1} - x_i)^2 &= E(x_{i+1}^2) + E(x_i^2) = 2 E(x^2), \\ E(x_{i+1} - x_i)(y_{i+1} - y_i) &= E(x_{i+1} y_{i+1}) + E(x_i y_i) = 2 E(x y); \end{aligned}$$

Il résulte de là que :

$$r_{\Delta^1 x \Delta^1 y} = \frac{2 E(x y)}{\sqrt{2 E(x^2) \cdot 2 E(y^2)}} = r_{xy}.$$

On trouverait de même que :

$$r_{\Delta^n x \Delta^n y} = r_{\Delta^{n-1} x \Delta^{n-1} y} = \dots = r_{xy}.$$

Considérons maintenant deux séries statistiques x_i, y_i définies respectivement par les développements

$$\begin{aligned} x_i &= X_i + b t_i + c t_i^2 + \dots + h t_i^n, \\ y_i &= Y_i + b' t_i^2 + c' t_i^2 \dots + h' t_i^n, \end{aligned}$$

et formons les différences $(\Delta^1 x_i, \Delta^1 y_i), \dots, (\Delta^{n+1} x_i, \Delta^{n+1} y_i)$; il est évident, eu égard à la forme même du développement de x_i et de y_i , que

$$\begin{aligned} \Delta^{n+1} x_i &= \Delta^{n+1} X_i + C_1, \\ \Delta^{n+1} y_i &= \Delta^{n+1} Y_i + C_2, \end{aligned}$$

d'où l'on déduit :

$$r_{\Delta^{n+1} x, \Delta^{n+1} y} = r_{\Delta^{n+1} X, \Delta^{n+1} Y} = r_{X, Y},$$

et aussi :

$$r_{\Delta^{n+2} x, \Delta^{n+2} y} = r_{\Delta^{n+1} x, \Delta^{n+1} y}, \text{ car } \Delta^{n+1} x \text{ et } \Delta^{n+1} y$$

sont maintenant des variables aléatoires indépendantes du temps.

En définitive, on est conduit à calculer les différences d'ordre 1, 2, $(n + 1)$ de x_i et y_i , puis le coefficient de corrélation entre les $\Delta^{n+1} x$ et les $\Delta^{n+1} y$;

(4) *The elimination of spurious correlation due to position in time or space* (BiométriKa, vol. X, 1914, p. 179-180).

on s'arrêtera à la différence d'ordre $(n + 1)$, lorsque les coefficients de corrélation correspondant aux Δ_{n+2} , Δ_{n+3} seront égaux au coefficient caractérisant les différences d'ordre $(n + 1)$, qui sera le coefficient cherché (1).

Cette délicate question des séries oscillantes a été reprise par Anderson dans deux mémoires (2), que nous allons analyser maintenant.

Nous procéderons ensuite à l'examen d'un mémoire dû à Karl Pearson et Ethel Elderton sur la méthode des différences (3), à une étude « sur les variations de l'équation personnelle et la corrélation entre des jugements successifs » d'Egon Person (4), à un examen des remarques tirées de l'important discours présidentiel de Yule pour la session 1925-1926, prononcé à la Société royale de Statistique de Londres (5).

Nous sommes ensuite conduit à faire appel à l'emploi de la fonction caractéristique pour mettre en évidence certains résultats des plus importants concernant les séries à liaison interne.

Examen des deux premiers mémoires d'Anderson sur l'emploi de la méthode des différences.

Anderson, qui avait depuis 1912, par une voie différente de celle de Student, retrouvé le résultat de ce dernier, s'est surtout inspiré dans ses recherches des méthodes chères à Tchebycheff, Markoff et Von Bortkiewicz.

Par série oscillatoire, Anderson, dans son premier mémoire, entend les séries de valeurs indépendantes d'une variable aléatoire x .

$$x_1, x_2, \dots, x_n,$$

où l'espérance mathématique de x_i est égale à :

$$E(x_i) = \text{constante},$$

et

$$E(x_i, x_j) = E(x_i) E(x_j).$$

Si nous ramenons ces variables à leur valeur probable, c'est-à-dire si nous prenons $E(x) = 0$, nous remarquons après avoir posé :

$$x_i - E(x_i) = \xi_i,$$

que

$$E(\xi_i) = 0, \quad E(\xi_i, \xi_j) = E(\xi_i) E(\xi_j) = 0;$$

ceci étant, l'écart quadratique de la série défini par l'expression $\sum_1^n \frac{[x_i - E(x)]^2}{n}$ a pour espérance mathématique :

$$\sigma_x^2 = E(x^2) - [E(x)]^2.$$

(1) Application de la statistique aux problèmes bancaires. (Note de E. G. Peake parue dans le *Banker Magazine*, juillet-août 1912; note de Hall et Mercer dans le *Journal of Agricultural Science*.)

(2) *Nochmals uber : The elimination of spurious correlation due to position in time or space* (*Biométrie*, vol. X, 1914).

et : *Ueber ein neues Verfahren bei Anwendung der « Variate difference » methode* (*Biométrie*, vol. XV, 1923).

(3) Voir : *Biométrie*, vol. XIV, p. 281-310.

(4) Voir : *Biométrie*, vol. XIV, p. 23-102. — Ce mémoire apporte une contribution importante à l'étude des liaisons internes d'une série de variables aléatoires.

(5) Mémoire intitulé : « Pourquoi nous décelons parfois des corrélations entre les séries, fonctions du temps. Étude sur l'échantillonnage et sur la valeur de ces séries. »

Considérons maintenant la moyenne arithmétique $M_x = \frac{\sum x_i}{n}$ et formons l'expression

$$E \left[\frac{1}{n} \sum_1^n (x_i - M_x)^2 \right];$$

compte tenu de ce que

$$x_i - M_x = \xi_i - M_\xi, \text{ avec } M_\xi = \frac{1}{n} \sum_1^n \xi_i,$$

on remarque, après un calcul simple, que

$$E \left[\sum_1^n (\xi_i - M_\xi)^2 \right] = (n - 1) E (\xi^2),$$

d'où l'on déduit :

$$\sigma_x^2 = E \left[\frac{\sum_1^n (x_i - M_x)^2}{n - 1} \right].$$

Procédons maintenant au calcul de l'écart quadratique afférent aux différences premières $\Delta^1 x$.

Comme l'on a :

$$\Delta^1 x_i = x_{i+1} - x_i = \xi_{i+1} - \xi_i,$$

et

$$E (\Delta^1 x) = E (\xi_{i+1} - \xi_i) = 0,$$

il s'ensuit que l'écart quadratique en question $\sigma_{\Delta^1 x}$ est défini par l'expression :

$$\begin{aligned} \sigma_{\Delta^1 x}^2 &= E \left[\frac{\sum_1^{n-1} (\xi_{i+1} - \xi_i)^2}{n - 1} \right] = \frac{1}{n - 1} \left\{ E \left[\sum_1^{n-1} (\xi_i^2) \right] - 2 E \left[\sum_1^{n-1} (\xi_{i+1} \xi_i) \right] \right. \\ &\quad \left. + E \left[\sum_2^n \xi_i^2 \right] \right\} = 2 E (\xi^2) \end{aligned}$$

d'où

$$\sigma_{\Delta^1 x}^2 = 2 \sigma_x^2.$$

On obtient par un processus analogue le carré de l'écart quadratique des différences d'ordre 2, 3,..... k, et l'on trouve ainsi les équations :

$$(1) \quad \sigma_x^2 = \frac{\sigma_{\Delta^2 x}^2}{2} = \frac{\sigma_{\Delta^3 x}^2}{6} = \dots = \frac{\sigma_{\Delta^k x}^2}{(2k)!} = E (\xi^2)$$

et les relations approchées :

$$(1a) \quad \frac{\sum_1^n (x_i - M_x)^2}{n - 1} = \frac{\sum_1^{n-1} (\Delta^1 x)^2}{2(n - 1)} = \frac{\sum_1^{n-2} (\Delta^2 x)^2}{6(n - 2)} = \dots = \frac{\sum_1^{n-k} (\Delta^k x)^2}{(2k)!} = \sigma_x^2.$$

Au point de vue des applications, il est intéressant de connaître les écarts quadratiques des variables aléatoires :

$$\frac{\sum_1^n [x_i - E(x)]^2}{n} \cdot \frac{\sum_1^n (x_i - M_x)^2}{n-1}, \frac{\sum_1^{n-1} (\Delta^1 x)^2}{2(n-1)}, \frac{\sum_1^{n-2} (\Delta^2 x)^2}{6(n-2)}, \dots \frac{\sum_1^{n-k} (\Delta^k x)^2}{(2k)! (k!)^2 (n-k)}$$

C'est ainsi qu'à la quantité

$$E \left[\frac{\sum_1^n \left\{ (x_i - E(x))^2 \right\}^2}{n} \right]$$

correspond l'écart quadratique

$$E \left[\frac{\sum_1^n \xi_i^2}{n} - \sigma_x^2 \right]^2 = E \left[\frac{\xi_1^2 + \xi_2^2 + \dots + \xi_n^2}{n} - \sigma_x^2 \right]^2$$

$$= E \left[\frac{\sum_1^n \xi_i^4 + 2 \sum \xi_i^2 \xi_j^2}{n^2} - 2 \sigma_x^2 \frac{\sum \xi_i^2}{n} + \sigma_x^4 \right] = \frac{1}{n} \left\{ E(\xi^4) - [E(\xi^2)]^2 \right\}.$$

Pour obtenir l'écart quadratique de $\sum_1^n \frac{(x_i - M_x)^2}{n}$, il faut calculer :

$$E \left[\frac{\sum_1^n (x_i - M_x)^2}{n-1} - \sigma_x^2 \right]^2;$$

or, si l'on tient compte de ce que

$$E(\sum \xi_i \xi_j) = 0 \text{ et } E\left\{ \sum \xi_i^2 \sum \xi_j \xi_k \right\} = 0,$$

on trouve pour le carré de l'écart quadratique :

$$\frac{E(\xi^4) - [E(\xi^2)]^2}{n} + \frac{2[E(\xi^2)]^2}{n(n-1)}.$$

Quant au carré de l'écart quadratique de $\sum_1^{n-k} \frac{(\Delta^k x)^2}{(2k)! (k!)^2 (n-k)}$, il est défini par

l'expression assez complexe suivante :

$$\frac{1}{A_0^2 (n-K)^2} \left[A_0^2 (n-2k) [E(\xi^4) - \{E(\xi^2)\}^2] + 4[E(\xi^2)]^2 [A_1^2 (n-2k+1) + A_2^2 (n-2k-2) + \dots + A_k^2 (n-k)] \right. \\ \left. + 2B_0^2 [E(\xi^4) - \{E(\xi^2)\}^2] + 8[E(\xi^2)]^2 [B_1^2 + B_2^2 + \dots + B_{k-1}^2] \right]$$

dans laquelle les coefficients A_i^2 ont pour valeur :

$$A_i^2 = \left[\frac{(2k)!}{(k-i)!(k+i)!} \right]^2, \text{ avec } i = 0, 1, \dots, k.$$

$$B_2^2 = (b_0 b_1)^2 + (b_0 b_1 + b_1 b_{+1})^2 + \dots + (b_0 b_1 + b_1 b_{+1} + \dots + b_{k-1-i} b_{k-1})^2, \dots$$

$$B_{k-1}^2 = (b_0 b_{k-1})^2 \text{ avec } i = 0, 1, 2, \dots, (k-1).$$

En première approximation, ces écarts varient comme $\frac{1}{\sqrt{n}}$.

Dans le cas où la variable aléatoire suit la loi de Gauss, il suffit de substituer $3\sigma_x^4$ à $E(\xi^4)$, et l'on constate que les carrés des écarts quadratiques relatifs aux quantités :

$$\frac{\sum_1^n [x_i - E(x)]^2}{n}, \frac{\sum_1^n (x_i - M_x)^2}{n-1}, \frac{\sum_1^{n-1} (\Delta^1 x)^2}{2(n-1)}, \frac{\sum_1^{n-2} (\Delta^2 x)^2}{6(n-2)}, \frac{\sum_1^{n-3} (\Delta^3 x)^2}{20(n-3)},$$

sont représentés respectivement par :

$$\frac{2\sigma_x^4}{n}, \frac{2\sigma_x^4}{n-1}, \frac{(3n-4)}{(n-1)^2}\sigma_x^4, \frac{(35n-88)}{9(n-2)^2}\sigma_x^4, \frac{(231n-843)}{50(n-3)^2}\sigma_x^4.$$

De l'examen de deux séries (x_i, y_i) et calcul de $\sum_1^n \frac{x_i y_i}{n}$.

Passons maintenant au cas de deux séries oscillantes (x_1, x_2, \dots, x_n) (y_1, y_2, \dots, y_n) , et remarquons que

$$E\{[x_i - E(x)][y_j - E(y)]\} = 0$$

toutes les fois que $i \neq j$, si l'on admet que la corrélation ne subsiste qu'entre les termes de même rang.

Si maintenant nous posons :

$$\begin{cases} x_i - E(x) = \xi_i \\ y_i - E(y) = \psi_i \end{cases} \quad \hat{p}_{xy} = E \left[\frac{\sum_1^n \xi_i \psi_i}{n} \right] = E(\xi_i \psi_i),$$

et si nous désignons respectivement par M_x et M_y les valeurs moyennes de x et y , nous trouvons :

$$E \left\{ \frac{\sum_1^n (x_i - M_x)(y_i - M_y)}{n-1} \right\} = E(\xi_i \psi_i) = p_{xy},$$

$$p_{\Delta^1 x \Delta^1 y} = E \left\{ \frac{\sum_1^{n-1} \Delta^1 x_i \Delta^1 y_i}{n-1} \right\} = 2 p_{xy}, \dots \dots \dots \text{et } p_{\Delta^k x \Delta^k y} = E \left\{ \frac{\sum_1^{n-k} \Delta^k x_i \Delta^k y_i}{n-k} \right\} = \frac{(2k)!}{(k!)^2} p_{xy};$$

Nous pouvons ainsi écrire deux systèmes de relations dont le premier est exact, et le second n'est qu'approché.

$$(2) \quad p_{xy} = \frac{p_{\Delta^1 x \Delta^1 y}}{2} = \frac{p_{\Delta^2 x \Delta^2 y}}{6} = \dots = \frac{p_{\Delta^k x \Delta^k y}}{(2k)!} = E(\xi_i \psi_i)$$

$$(2 a) \quad \frac{\sum_1^n [x_i - E(x)][y_i - E(y)]}{n-1} = \frac{\sum_1^{n-1} \Delta^1 x_i \Delta^1 y_i}{2(n-1)} = \dots = \frac{\sum_1^{n-k} \Delta^k x_i \Delta^k y_i}{(2k)! (n-k)} = p_{xy}$$

Avant de passer à la relation de liaison entre $r_{\Delta^k x \Delta^k y}$ et $p_{\Delta^k x \Delta^k y}$, il est nécessaire de faire une remarque sur le coefficient de corrélation défini ordinairement par la formule

$$R_0 = \frac{\sum_1^n (x_i - M_x)(y_i - M_y)}{\sqrt{\sum_1^n (x_i - M_x)^2 \sum_1^n (y_i - M_y)^2}};$$

Cette expression de R_0 doit-elle être considérée comme une valeur approchée de :

$$E \left\{ \frac{\sum_1^n (x_i - M_x)(y_i - M_y)}{\sqrt{\sum_1^n (x_i - M_x)^2 \sum_1^n (y_i - M_y)^2}} \right\}, \text{ ou de } \frac{E \left[\sum_1^n (x_i - M_x)^2 (y_i - M_y)^2 \right]}{\sqrt{E \left[\sum_1^n (x_i - M_x)^2 \right] E \left[\sum_1^n (y_i - M_y)^2 \right]}}$$

Ces formules, qui ne peuvent être identifiées, sont en première approximation équivalentes; en raison de ce que la seconde est d'un maniement plus facile, et qu'elle est aussi plus conforme aux méthodes de calcul de l'école anglaise, nous l'adopterons pour la définition de r_{xy} ,

$$r_{xy} = \frac{p_{xy}}{\sigma_x \sigma_y}$$

Ceci étant, on voit que l'expression de $r_{\Delta^k x \Delta^k y}$ s'écrit alors :

$$\begin{aligned} r_{\Delta^k x \Delta^k y} &= \frac{E \left(\sum_1^{n-k} \Delta^k x_i \Delta^k y_i \right)}{\sqrt{E \left[\sum_1^{n-k} (\Delta^k x_i)^2 \right] E \left[\sum_1^{n-k} (\Delta^k y_i)^2 \right]}} = \frac{p_{\Delta^k x \Delta^k y}}{\sqrt{\sigma_{\Delta^k x}^2 \sigma_{\Delta^k y}^2}} \\ &= \frac{\frac{(2k)!}{(k!)^2} p_{xy}}{\sqrt{\frac{(2k)!}{(k!)^2} \sigma_x^2 \frac{(2k)!}{(k!)^2} \sigma_y^2}} = r_{xy} \end{aligned}$$

Dans la pratique, aux $r_{\Delta^k x \Delta^k y}$, il y a lieu de substituer les valeurs approchées R_h , qui se calculent au moyen des éléments d'observation

$$R_h = \frac{\sum_1^{n-h} \Delta^h x_i \Delta^h y_i}{\sqrt{\sum_1^{n-h} (\Delta^h x_i)^2 \sum_1^{n-h} (\Delta^h y_i)^2}}, \text{ avec } h = (1, 2, \dots, k),$$

et aussi savoir de combien ces R_h diffèrent de leur valeur vraie.

Nous donnons à titre indicatif le carré de l'écart draquatique caractérisant R_h , dans l'hypothèse où les variables x et y suivent l'une et l'autre la loi de Gauss.

$$\begin{aligned} \sigma_{R_h}^2 &= \left(\frac{1 - r_{xy}^2}{n - k} \right) \left[(n - k) + 2(n - k - 1) \left(\frac{k}{k + 1} \right)^2 \right. \\ &\left. + 2(n - k - 2) \left\{ \frac{k(k - 1)}{(k + 1)(k + 2)} \right\}^2 + \dots + 2(n - 2k) \left\{ \frac{k!}{(k + 1)(k + 2) \dots 2k} \right\}^2 \right] \end{aligned}$$

Les séries du groupe II et la méthode des différences.

Anderson fait remarquer que la méthode des différences est utilisée pour mettre en lumière les séries épiséculaires du type II, et que son emploi dépend tout d'abord, comme le dit K. Pearson, des résultats de recherches statistiques concrètes. Si ces recherches permettent de faire ressortir la présence plus ou moins fréquente de séries du type II, la méthode sera pleinement justifiée.

Les deux suites statistiques

$$(x_1, x_2, \dots, x_n), (y_1, y_2, \dots, y_n),$$

telles que pour $k < n$, l'on ait le système de relations :

$$(3) \quad \frac{E \sum_{i=1}^{n-k-j} x_i y_{i+j}}{n-k-j} = \frac{E \sum_{i=1}^{n-k-j} x_{i+1} y_{i+j+1}}{n-k-j} \\ = \dots = \frac{E \sum_{i=1}^{n-k-j} x_{i+k} y_{i+j+k}}{n-k-j} = \frac{E \sum_{i=1}^{n-j} x_i y_{i+j}}{n-j}$$

sont celles auxquelles Anderson a consacré ses deux mémoires publiés respectivement dans *Biométriqa*, vol. X et XV.

Si l'on pose :

$$(4) \quad r_j = \frac{E \sum_{i=1}^{n-j} x_i y_{i+j}}{n-j} : \sqrt{\frac{E \sum_{i=1}^n x_i^2}{n} \cdot \frac{E \sum_{i=1}^n y_i^2}{n}} = \frac{E \sum_{i=1}^{n-j} x_i y_{i+j}}{(n-j) \sigma_x \sigma_y}$$

(où x et y représentent respectivement pour chacune des séries un quelconque des écarts par rapport à la valeur moyenne), on est conduit naturellement à exprimer les éléments figuratifs de $r_{\Delta^k x, \Delta^k y}$, en fonction des r_j , compte tenu du système de relations (3).

Il nous faut maintenant procéder au calcul de $\frac{E \sum_{i=1}^{n-k-j} \Delta^k x_i \Delta^k y_{i+j}}{n-k-j}$; pour cela, considérons l'une des valeurs de i , et rappelons que $\Delta^k x_i$ peut s'écrire symboliquement $(1 - \varepsilon)^k x_i$ avec $\varepsilon^s x_i \stackrel{\Omega}{=} x_{i+s}$.

Dans ces conditions, $\Delta^k x_i \cdot \Delta^k y_{i+j}$, n'est autre chose qu'une somme de termes analogues au suivant

$$(-1)^{\alpha+\beta} x_{i+\alpha} y_{i+j-\beta} C_k^\alpha C_k^\beta,$$

où C_k^α représente le nombre des combinaisons de k objets α à α .

Si nous désignons la différence des indices attachés respectivement à y et x par $k + j - p$:

$$j + \beta - \alpha = k + j - p,$$

nous remarquons que l'on peut satisfaire à cette relation en prenant :

$$\beta = k - p + \lambda, \alpha = \lambda, \text{ avec } \lambda = (0, 1, 2, \dots, p);$$

à chacun de ces systèmes de β et de α correspond un produit :

$$(-1)^{\alpha+\beta} x_{i+\alpha} y_{i+j+\beta} C_k^\alpha C_k^\beta = (-1)^{k-p} (-1)^{2\lambda} x_{i+\lambda} y_{i+\lambda+(k+j-p)} C_k^\lambda C_k^\beta$$

La somme des produits $C_k^\alpha C_k^\beta$ afférents aux diverses associations de α et de β n'est autre :

$$C_k^{k-p} + C_k^1 C_k^{k-p+1} + C_k^2 C_k^{k-p+2} + \dots + C_k^p C_k^k,$$

ou encore — comme on le vérifie facilement — le coefficient de z^p dans le produit de $(1+z)^k (1+z)^k$, soit dans le développement de $(1+z)^{2p}$, soit C_{2p}^p .

Remarquant, d'une part, qu'à une valeur fixe de i , et à une différence des indices de y et x correspond le coefficient $(-1)^{k+p} C_{2k}^p$, on voit d'autre part,

en vertu de (4), puisque i peut prendre les valeurs $(1, 2, \dots, (n - k - j))$, que l'on a :

$$(5) \quad \frac{E \sum_{i=1}^{n-k-j} \Delta^k x_i \Delta^k y_{i+j}}{n-k-j} = \sigma_x \sigma_y (+1)^k \left[r_{k+j} - 2k r_{k+j-1} + \frac{2k(2k-1)}{1 \cdot 2} r_{k+j-2} + \dots + r_{j-k} \right];$$

comme la parenthèse figurant dans le second membre de (5) n'est autre que $\Delta^{2k} r_{k+j}$, de la suite $r_{k+j}, r_{k+j-1}, \dots, r_{j-k}$, il s'ensuit que l'on obtient définitivement :

$$(5)' \quad \frac{E \sum_{i=1}^{n-k-j} \Delta^k x_i \Delta^k y_{i+j}}{n-k-j} = (-1)^k \sigma_x \sigma_y \Delta^{2k} r_{k+j}.$$

relation d'où l'on déduit :

$$(5)'' \quad \frac{E \sum_{i=1}^{n-k} \Delta^k x_i \Delta^k y_i}{n-k} = (-1)^k \sigma_x \sigma_y \Delta^{2k} r_k.$$

Si maintenant l'on remplace dans (4) y_{i+j} par x_{i+j} , r_j se change en r'_j :

$$(6) \quad r'_j = \frac{E \sum_{i=1}^{n-j} x_i x_{i+j}}{n-j} : \frac{E \sum_{i=1}^n x_i^2}{n}$$

et les équations (5)' et (5)'' deviennent :

$$(7) \quad \frac{E \sum_{i=1}^{n-k-j} \Delta^k x_i \Delta^k x_{i+j}}{n-k-j} = \sigma_x^2 (-1)^k \Delta^{2k} r'_{k+j},$$

$$(7a) \quad \frac{E \sum_{i=1}^{n-k} \Delta^k x_i^2}{n-k} = (-1)^k \sigma_x^2 \Delta^{2k} r'_k,$$

en remarquant de plus que pour une suite donnée purement algébrique, l'on a :

$$r'_j = r'_{-j}, \text{ et que : } r'_0 = 1.$$

Pour les séries oscillantes au sens d'Anderson, on a :

$$\frac{E \sum_{i=1}^{n-k-1} \Delta^{k+1} x_i \Delta^{k+1} y_i}{(2k+2)! [(k+1)!]^2 (n-k-1)} = \frac{E \sum_{i=1}^{n-k} \Delta^k x_i \Delta^k y_i}{(2k)! (k!)^2 (n-k)},$$

relation qui en vertu de (5)' et de la formule relative à $\Delta^{2k} r_{k+j}$, peut s'écrire :

$$(7)' \quad 0 = \frac{(-1)^k \sigma_x \sigma_y \Delta^{2k} r_k}{(2k+1)! (k+1)! k!} \left[-\frac{1}{2} \left(\frac{\Delta^{2k} r_{k+1} + \Delta^{2k} r_{k-1}}{\Delta^{2k} r_k} \right) - \frac{k}{k+1} \right];$$

ceci étant, si l'on désigne par ${}_k r_1$ l'expression :

$$\frac{\Delta^{2k} r_{k+1}}{\Delta^{2k} r_k} = \frac{(-1)^k \sigma_x \sigma_y \Delta^{2k} r_{k+1}}{(-1)^k \sigma_x \sigma_y \Delta^{2k} r_k} = \frac{E \sum_{i=1}^{n-k-1} \Delta^k x_i \Delta^k y_{i+1}}{n-k-1} : \frac{E \sum_{i=1}^{n-k} \Delta^k x_i \Delta^k y_i}{n-k}$$

et par $({}_k r_{-1})$, le rapport :

$$\frac{E \sum_{i=1}^{n-k-1} \Delta^k x_{i+1} \Delta^k y_i}{n-k-1} : \frac{E \sum_{i=1}^{n-k} \Delta^k x_i \Delta^k y_i}{n-k},$$

on constate de suite que l'équation (7)' s'écrit alors :

$$(8) \quad 0 = \frac{(-1)^k \alpha_x \alpha_y \Delta^{2k} r_k}{(2k+1)!} \left[-\frac{{}_k r_1 + {}_k r_{-1}}{2} - \frac{k}{k+1} \right],$$

ce qui revient à dire que la condition

$$\frac{{}_k r_1 + {}_k r_{-1}}{2} = -\frac{k}{k+1}$$

doit être réalisée, pour que les suites x et y soient, d'après Anderson, des séries oscillantes.

En substituant aux y_n les x_n , et en posant

$$(9) \quad \frac{E \sum_{i=1}^{n-k-1} \Delta^k x_i \Delta^k y_{i+1}}{n-k-1} : \frac{E \sum_{i=1}^{n-k} (\Delta^k x_i)^2}{n-k} = {}_k r'_1,$$

on trouve immédiatement que la relation

$$\frac{E \sum_{i=1}^{n-k-1} (\Delta^{k+1} x_i)^2}{(2k+2)! [(k+1)!]^2 (n-k-1)} = \frac{E \sum_{i=1}^{n-k} (\Delta^k x_i)^2}{(2k!) (k!)^2 (n-k)}$$

n'est réalisée que si l'on a :

$${}_k r'_1 = -\frac{k}{k+1} = \frac{{}_k r_1 + {}_k r_{-1}}{2}.$$

En définitive, si l'on se trouve en présence de séries oscillantes du type Anderson, on voit que l'on peut passer de

$$\mathcal{E} = \frac{E \sum_{i=1}^{n-k} \Delta^k x_i \Delta^k y_i}{n-k} \quad \text{à} \quad \frac{E \sum_{i=1}^{n-k-1} \Delta^{k+1} x_i \Delta^{k+1} y_i}{n-k-1}$$

et par l'intermédiaire des ${}_k r_1$, ${}_k r_{-1}$ et de \mathcal{E} , aux espérances mathématiques de

$$\frac{\sum_{i=1}^{n-k-1} \Delta^k x_i \Delta^k y_{i+1}}{n-k-1}, \quad \frac{\sum_{i=1}^{n-k-1} \Delta^k x_{i+1} \Delta^k y_i}{n-k-1},$$

et l'on saisit ainsi le mécanisme d'utilisation des ${}_k r'_1$.

Une généralisation immédiate nous conduit immédiatement à l'étude de l'expression (10),

$$(10) \quad \left[\frac{E \sum_{i=1}^{n-k-m} \Delta^{k+m} x_i \Delta^{k+m} y_i}{(2k+2m)! [(k+m)!]^2 (n-k-m)} - \frac{E \sum_{i=1}^{n-k} \Delta^k x_i \Delta^k y_i}{(2k!) (k!)^2 (n-k)} \right],$$

ou encore de l'expression équivalente :

$$(-1)^k \sigma_x \sigma_y \left[\frac{(-1)^m \Delta^{2k+2m} r_{k+m}}{(2k+2m)!} - \frac{\Delta^{2k} r_k}{(k!)^2} \right].$$

Nous n'entrerons pas ici dans le détail des calculs que suggère l'étude de l'expression (10), mais nous ferons remarquer finalement que si l'on fait intervenir les rapports :

$${}_k r_j = \frac{E \sum_{i=1}^{n-k-j} \Delta^k x_i \Delta^k y_{i+j}}{n-k-j} : \frac{E \sum_{i=1}^{n-k} \Delta^k x_i \Delta^k y_i}{n-k},$$

$${}_k r'_j = \frac{E \sum_{i=1}^{n-k-j} \Delta^k x_i \Delta^k x_{i+j}}{n-k-j} : \frac{E \sum_{i=1}^{n-k} (\Delta^k x_i)^2}{n-k},$$

on constante, en prenant :

$$(11) \quad \frac{{}_k r_j + {}_k r_{-j}}{2} = {}_k r'_j, \text{ avec } (12) \quad {}_k r'_j = (-1)^j \frac{(k!)^2}{(k+j)!(k-j)!}, \text{ avec } j = (1, 2, \dots, m),$$

que les différences (10) se réduisent à zéro.

La substitution de x_i à y_i conduit au résultat remarquable suivant :

$$\left(4 - \frac{2}{k+h} \right) \frac{E \sum_{i=1}^{n-k-h+1} (\Delta^{k+h-1} x_i)^2}{n-k-h=1} = \frac{E \sum_{i=1}^{n-k-h} (\Delta^{k+h} x_i)^2}{n-k-h}$$

avec $h = (1, 2, 3, \dots, m)$, toutes les fois que les ${}_k r'_1$ ont les valeurs définies ci-dessus.

Revenons maintenant à l'expression :

$$\frac{E \sum_{i=1}^{n-k-1} \Delta^{k+1} x_i \Delta^{k+1} y_i}{(2k+2)! [(k+1)!]^2 (n-k-1)} - \frac{E \sum_{i=1}^{n-k} \Delta^k x_i \Delta^k y_i}{(2k)! (k!)^2 (n-k)}, \text{ égale à } \left\{ \sigma_x \sigma_y \Delta^{2k} r_k \left(-\frac{{}_k r_1 + {}_k r_{-1}}{2} - \frac{k}{k+1} \right) (-1)^k \right\},$$

et rappelons qu'elle est nulle si les séries sont bien des séries oscillantes.

Si, dans l'expression précédente, on remplace k par $(k+m)$, on forme ainsi une différence δ_1 , qui, eu égard à la relation ${}_k r_j = \frac{\Delta^{2k} r_{k+j}}{\Delta^{2k} r_k}$ et à une transformation simple, peut être représentée sous deux formes différentes (12) et (12)'.

$$(12) \quad \delta_1 = H \left[-\frac{\Delta^{2k+2m} r_{k+m+1} + \Delta^{2k+2m} r_{k+m-1}}{2 \Delta^{2k+2m} r_{k+m}} - \frac{k+m}{k+m+1} \right].$$

$$(12)' \quad \delta_1 = H \left[-\frac{\Delta^{2m} {}_k r_{m+1} + \Delta^{2m} {}_k r_{m-1}}{2 \Delta^{2m} {}_k r_m} - \frac{k+m}{k+m+1} \right],$$

où

$$H = (-1)^{k+m} \frac{\sigma_x \sigma_y \Delta^{2k+2m} r_{k+m}}{(2k+2m+1)! (k+m+1)! (k+m)!}$$

$$\delta_1 = \frac{E \sum_{i=1}^{n-k-m-1} \Delta^{k+m+1} x_i \Delta^{k+m+1} y_i}{(2k+2m+2)! [(k+m+1)!]^2 (n-k-m-1)} - \frac{E \sum_{i=1}^{n-k-m} \Delta^{k+m} x_i \Delta^{k+m} y_i}{(2k+2m)! [(k+m)!]^2 (n-k-m)}.$$

En remplaçant, dans l'expression de la différence δ_1 (y_i par x_i), on forme ainsi une expression δ_2 et l'on est conduit naturellement aux deux formes classiques de δ_2 .

$$(13) \quad \delta_2 = \frac{(-1)^{k+m} \sigma, 2 \Delta^{2k+2m} r'_{k+m}}{(2k+2m+1)!} \left[-\frac{\Delta^{2k+2m} r'_{k+m+1}}{\Delta^{2k+2m} r'_{k+m}} - \frac{k+m}{k+m+1} \right].$$

$$(13)' \quad \delta_2 = \frac{(-1)^{k+m} \sigma, 2 \Delta^{2k+2m} r'_{k+m}}{(2k+2m+1)!} \left[-\frac{\Delta^{2m} r'_{m+1}}{\Delta^{2m} r'_m} - \frac{k+m}{k+m+1} \right].$$

Critère relatif aux séries oscillantes du type Anderson.

Compte tenu de la relation (7 a), on constate que le signe de δ_2 est défini par celui de la quantité D.

$$(14) \quad D = \frac{-\Delta^{2k+2m} r'_{k+m+1}}{\Delta^{2k+2m} r'_{k+m}} - \frac{k+m}{k+m+1} = -\frac{\Delta^{2m} r'_{m+1}}{\Delta^{2m} r'_m} - \frac{k+m}{k+m+1}.$$

Si $D = 0$, la déviation standard de la $(k+m+1)^{i\text{ème}}$ différence se comporte par rapport à la déviation standard relative à la $(k+m)^{i\text{ème}}$ différence, d'une manière analogue à celle qui se manifeste dans une série oscillante.

Si $D > 0$, la série croit plus vite que dans le cas d'une série du type II, et, par suite, la série peut être considérée comme étant du type I pur, ou assimilée à une série du type IV avec une composante du type I.

Si $D < 0$, la composante continue n'est pas complètement éliminée, et la suite épicyclaire est caractérisée par un élément ou série du type III.

L'expression de D peut fournir un critère utile pour l'emploi de la méthode des différences, en admettant qu'elle puisse être calculée au moyen des données empiriques, et à la condition :

- 1° De conserver la même limite pour des différences k_1, k_2, \dots ;
- 2° De ne pas avoir une erreur quadratique trop forte.

Au lieu de D, l'on peut — en faisant abstraction du facteur $\frac{1}{k+m+1}$ —, étudier soit l'expression :

$$(15a) \quad {}_0D_{k+m+1} = -\frac{(\Delta^{2k+2m} r'_{k+m+1})(k+m+1)}{\Delta^{2k+2m} r'_{k+m}} - (k+m),$$

soit encore l'expression :

$$(15b) \quad {}_kD_{m+1} = -\frac{(k+m+1) \Delta^{2m} r'_{m+1}}{\Delta^{2m} r'_m} - (k+m).$$

A titre indicatif, nous allons déterminer ${}_0D_3$ et ${}_kD_3$:

$${}_0D_3 = \frac{-3 \Delta^4 r'_3}{\Delta^4 r'_2} - 2 = \frac{-3 (r'_3 - 4 r'_2 + 6 r'_1 - 4 r'_0 + r'_{-1})}{r'_2 - 4 r'_1 + 6 r'_0 - 4 r'_{-1} + r'_{-2}} - 2,$$

Et comme :

$$r'_0 = 1; r'_{-1} = r'_1; r'_{-2} = r'_2,$$

on trouve :

$${}_0D_3 = \frac{-5r'_1 + 8r'_2 - 3r'_3}{6 - 8r'_1 + 2r'_2};$$

$${}_kD_3 = \frac{-(k+3)\Delta^4_k r'_3}{\Delta^4_k r'_2} - (k+2) = \frac{-(k+3)[-{}_k r'_3 - 4{}_k r'_2 + 6{}_k r'_1 - 4{}_k r'_0 + {}_k r'_{-1}]}{{}_k r'_2 - 4{}_k r'_1 + 6{}_k r'_0 - 4{}_k r'_{-1} + {}_k r'_{-2}} - (k+2),$$

et du fait que

$${}_k r'_{-2} = {}_k r'_2; {}_k r'_{-1} = {}_k r'_1; {}_k r'_0 = 1,$$

l'on obtient

$${}_kD_3 = \frac{-2k - (5-k){}_k r'_1 + (2k+8){}_k r'_2 - (k+3){}_k r'_3}{2{}_k r'_2 - 8{}_k r'_1 + 6}$$

Nous donnons ci-dessous les valeurs de

$${}_0D_1, {}_0D_2, {}_0D_4, {}_0D_5; {}_kD_1, {}_kD_2, {}_kD_4, {}_kD_5.$$

$$(16a) \left\{ \begin{array}{l} {}_0D_1 = -r'_1 \quad {}_0D_4 = \frac{-7r'_1 + 14r'_2 - 9r'_3 + 2r'_4}{10 - 15r'_1 + 6r'_2 - r'_3} \\ {}_0D_2 = \frac{-r'_1 + r'_2}{1 - r'_1} \quad {}_0D_5 = \frac{-42r'_1 + 96r'_2 - 81r'_3 + 32r'_4 - 5r'_5}{70 - 112r'_1 + 56r'_2 - 16r'_3 + 2r'_4} \end{array} \right\}$$

$$(16b) \left\{ \begin{array}{l} {}_kD_1 = - (k+1) {}_k r'_1 - k \\ {}_kD_2 = \frac{-2{}_k r'_1 + (k+2) {}_k r'_2 - k}{2 - 2{}_k r'_1} \end{array} \right.$$

$${}_kD_4 = \frac{(4k-14){}_k r'_1 + (4k+28){}_k r'_2 - (4+18){}_k r'_3 + (k+4){}_k r'_4 - 5k}{20 - 30{}_k r'_1 + 12{}_k r'_2 - 2{}_k r'_3}$$

$${}_kD_5 = \frac{(14k-42){}_k r'_1 + (8k+96){}_k r'_2 - (13k+81){}_k r'_3 + (6k+32){}_k r'_4 - (k+5){}_k r'_5 - 14k}{70 - 112{}_k r'_1 + 56{}_k r'_2 - 16{}_k r'_3 + 2{}_k r'_4}$$

Utilisation des observations.

En pratique, les grandeurs r'_j et ${}_k r'_j$, représentant des quotients d'espérance mathématique, restent inconnues.

Aussi est-on conduit à introduire dans les formules (16 a) et (16 b), φ'_j , à la place de r'_j et ${}_k r'_j$, à celle de ${}_k \varphi'_j$, les quantités φ'_j et ${}_k \varphi'_j$, étant définies respectivement par les relations suivantes :

$$\varphi'_j = \frac{\binom{n-j}{\sum_{i=1}^{n-j} x'_i x'_{i+j}}}{n-j} : \frac{\binom{n}{\sum_{i=1}^n x_i^2}}{n-1} \quad \text{et} \quad {}_k \varphi'_j = \frac{\sum_{i=1}^{n-k-j} \Delta^k x'_i \Delta^k x'_{i+j}}{n-k-j} : \frac{\sum_{i=1}^{n-k} (\Delta^k x'_i)^2}{n-k}$$

où x', x'_0, \dots, x'_n représentent maintenant les termes de la série statistique (et non les écarts par rapport à la moyenne exprimée par l'espérance mathématique), et aussi où (x_n) remplace $\sum_1^n \frac{x_i^2}{n}$.

Si la relation

$$(7a) \quad E \frac{\sum_{i=1}^{n-k} (\Delta^k x)^2}{n-k} = (-1) {}_k \sigma_x^2 \Delta^{2k} r'_k$$

était exacte, et si, de plus, l'on avait l'identité

$$\frac{\sum_{i=1}^{n-k} (\Delta^k x_i)^2}{n-k} = \frac{\sum_{i=1}^n [x'_i - (x_n)]^2}{n-1} (-1)^k \Delta^{2k} \rho'_{k,}$$

on serait naturellement conduit à estimer :

$$(17) \quad E \left\{ \frac{\sum_{i=1}^{n-k-1} (\Delta^{k+1} x_i)^2}{(2k+2)! [(k+1)!]_2 (n-k-1)} - \frac{\sum_{i=1}^{n-k} (\Delta^k x_i)^2}{2k! (k!)^2 (n-k)} \right\};$$

En réalité, la relation (7 a) n'est satisfaite que si les conditions (3) où l'on a remplacé les y_k par x_k ,

$$(3) \quad \frac{\sum_{i=1}^{n-k-j} x_i x_{i+j}}{n-k-j} = \dots = \frac{\sum_{i=1}^{n-k-j} x_{i+k} x_{i+j+k}}{n-k-j} = \frac{\sum_{i=1}^{n-j} x_i x_{i+j}}{n-j},$$

sont remplies.

Dans l'hypothèse où la suite x'_1, x'_2, \dots, x'_n est non seulement oscillatoire, mais encore régie par une loi de distribution normale, on peut donner l'expression précise de (17), et l'on trouve alors que

$$\sigma_{\sigma_{2k+1}}^2 = \frac{(2k+1)^2}{\mu_{2k}^2} E \{ \sigma_{k+1}^2 - \sigma_k^2 \} 2 \text{ où } \sigma_k = \sqrt{\frac{1}{C_{2k}^k} \frac{\sum_{i=1}^{n-k} (\Delta^k x_i)^2}{n-k}}$$

Examen des derniers mémoires d'Anderson — 1926-1927 (1).

Anderson, revenant sur des travaux antérieurs, les a complétés, en s'efforçant de donner aux statisticiens des moyens de sondage des séries; ce sont ces éléments nouveaux d'information que nous allons mettre en lumière.

Nous avons dit au début de cette étude que l'on substituait à la figure polygonale représentative du phénomène une courbe continue, les écarts de cette courbe par rapport à la ligne polygonale provenant eux-mêmes d'une ou de plusieurs séries secondaires.

Si, dans le domaine de la démographie, l'on peut quelquefois adapter certains types de fonctions à la représentation de cette courbe continue, il n'en est pas tout à fait de même dans le domaine de l'économique, et l'on est réduit en général à utiliser des paraboles d'un certain ordre.

Nous signalons que ce problème, qui n'est autre que celui de l'ajustement, est en lui-même fort important, et nous nous permettons de rappeler que nous lui avons consacré un chapitre spécial dans l'ouvrage intitulé : *Applications de la statistique à la démographie et à la biologie*.

Une suite statistique donnée u_1, u_2, \dots, u_n , résulte alors d'une composante

(1) Anderson a publié dans *Biométrie*, vol. XVIII, p. 293-320; vol. XIX, p. 53-86, deux mémoires importants : *Ueber die Anwendung der Differenzen methode (Variate difference method) bei Reihengleichungen, Stabilitäts untersuchungen und korrelations messungen*.

Voir aussi son important traité : *Einführung in die mathematische statistik* — (Wien — 1935 — Verlag von Julius Springer).

ajustée G et d'un reste s , de sorte que la grandeur u_i fournie par l'observation est telle que $u_i = G_i + s_i$.

Il est admis qu'après avoir pris la $k^{\text{ième}}$ différence, il ne subsiste plus que l'effet de s_i , c'est-à-dire que $\Delta^k u_i = \Delta^k G_i + \Delta^k s_i = \Delta^k s_i$.

Or si l'on se trouve en présence de n observations, on n'a plus que $(n - k)$ fonctions linéaires des grandeurs s_1, s_2, \dots, s_n , qui ne sont autres que :

(18) $\Delta^k s_i = s_i - C_k^1 s_{i+1} + C_k^2 s_{i+2} - \dots + (-1)^k C_k^k s_{i+k}$, et il s'en suit que les grandeurs s_i ($i = 1, 2, \dots, n$) ne peuvent en aucun cas être déduites des différences $\Delta^k s_i$.

Toutefois, du fait de l'introduction de certaines hypothèses sur la nature de la suite s_i , il est possible de déterminer diverses caractéristiques de la suite. Si l'on admet que ces grandeurs s_i sont des variables aléatoires correspondant à une loi de distribution constante afférente à n observations indépendantes les unes des autres, les formules correspondant aux caractéristiques en question se présentent d'une manière simple.

Utilisons les notations de Tschuprow, et formons :

$m_r = E s^r$, $\mu_r = E (s - m_1)^r$, l'écart quadratique

$$\sigma = \sqrt{\mu_2} = \sqrt{m_2 - m_1^2};$$

De l'écart *a priori* σ , rapprochons les deux formes de l'écart empirique

$$\sigma'_{\circ} = \sqrt{\frac{\sum_1^n (s_i - m_1)^2}{n}}, \quad \sigma' = \sqrt{\frac{\sum_1^n [s_i - s_{(n)}]^2}{n-1}}, \quad \text{où } s_{(n)} = \frac{\sum_1^n s_i}{n},$$

$$\sigma'_k = \sqrt{\frac{1}{C_{2k}^k \times (n-k)} \sum_{i=1}^{n-k} (\Delta^k s_i)^2},$$

et rappelons que :

$$E \sigma'_{\circ} = E \sigma' = \sqrt{\mu_2}.$$

et aussi :

$$E \sigma_k'^2 = \sigma^2.$$

Si la grandeur k est telle que l'on puisse appliquer aux factorielles la formule de Stirling, on peut écrire la valeur de $\sigma_k'^2$ ainsi qu'il suit :

$$\sigma_k'^2 = \frac{\sqrt{\pi} k^{n-k} (\Delta^k s_i)^2}{4^k \sum_{i=1}^{n-k} \frac{1}{n-k}}.$$

Nous allons chercher à exprimer de combien la grandeur empirique $\sigma_k'^2$ peut, dans la pratique, s'écarter de son espérance mathématique, et constater en cours de route que la grandeur de $(n - k)$ joue un rôle important.

La dispersion *a priori* de $\sigma_k'^2$ se détermine au moyen de la formule

$$\sigma_{(\sigma^2, k)}^2 = E (\sigma_k'^2 - \mu_2)^2.$$

Dans l'hypothèse où

$$k \leq \frac{n}{2}.$$

l'on trouve :

$$(19) \quad \sigma_{(\sigma^2, k)}^2 = \frac{\mu_4 - 3\mu_2^2}{n-k} \left[1 - \frac{2 S_{k^2}}{(C_{2k}^k)^2 (n-k)} \right] + \frac{2 \mu_2^2}{n-k} \left[\frac{C_{4k}^k}{(C_{2k}^k)^2} - \frac{k}{2(n-k)} \right],$$

où

$$S_k = \sum_{i=0}^{k-1} (C_i^1 C_k^{i+1})^2 + 2 \sum_{i=1}^{k-2} (C_i^1 C_k^{i+2})^2 + 3 \sum_{i=1}^{k-3} (C_i^1 C_k^{i+3})^2 + \dots + k (C_k^0 C_k^1)^2,$$

[avec $S_1 = 1$, $S_{(2)} = 10$, $S_{(3)} = 138$, $S_{(4)} = 1940$, $S_{(5)} = 28.130$, $S_{(6)} = 414.372 \dots$]

et l'on démontre que toutes les fois que

$$\frac{k}{2(n-k)} \geq \frac{2 S_k}{(C_{i,k}^k)^2 (n-k)}.$$

l'on peut prendre comme valeur approchée de $\sigma_{(\sigma^2, k)}^2$ l'expression

$$(20) \quad \frac{\mu_4 - 3 \mu_2^2}{n-k} + \frac{C_{4,k}^{2k}}{(C_{i,k}^k)^2} \cdot \frac{2 \mu_2^2}{n-k},$$

et si, de plus, la grandeur de k est telle que l'on puisse appliquer avec une erreur relative faible la formule de Stirling, on remarque que l'on peut écrire avec une approximation suffisante :

Valeur approchée de

$$(20 a) \quad \sigma_{(\sigma^2, k)}^2 = \frac{\mu_4 - 3 \mu_2^2}{n-k} + \frac{\mu_2^2 \sqrt{2 k \pi}}{n-k}$$

Nous donnons ci-dessous les expressions de $\sigma_{(\sigma^2, k)}^2$ empruntées au troisième mémoire de Anderson, dans l'hypothèse où (n) a une valeur relativement élevée, pour $k = (0, 1, 2, \dots, 6)$.

TABLEAU I

Ordre des différences	Valeurs de $\sigma_{(\sigma^2, k)}^2$ pour une loi de distribution	
	constante de la variable accidentelle	normale de la variable accidentelle
$k = 0$	$\frac{\mu_4 - \mu_2^2}{n}$	$\frac{2 \mu_2^2}{n}$
$k = 1$	$\frac{(n-1,5) \mu_4 + 0,5 \mu_2^2}{(n-1)^2}$	$\frac{3n-4}{(n-1)^2} \mu_2^2$
$k = 2$	$\frac{(n-2,556) \mu_4 + (0,889n-2,111) \mu_2^2}{(n-2)^2}$	$\frac{(3,889n-9,778) \mu_2^2}{(n-2)^2}$
$k = 3$	$\frac{(n-3,69) \mu_4 + (1,62n-5,79) \mu_2^2}{(n-3)^2}$	$\frac{(4,62n-16,86) \mu_2^2}{(n-3)^2}$
$k = 4$	$\frac{(n-4,792) \mu_4 + (2,253n-10,633) \mu_2^2}{(n-4)^2}$	$\frac{(5,253n-25,012) \mu_2^2}{(n-4)^2}$
$k = 5$	$\frac{(n-5,887) \mu_4 + (2,819n-16,436) \mu_2^2}{(n-5)^2}$	$\frac{(5,819n-34,094) \mu_2^2}{(n-5)^2}$
$k = 6$	$\frac{(n-6,971) \mu_4 + (3,335n-23,096) \mu_2^2}{(n-6)^2}$	$\frac{(6,335n-44,008) \mu_2^2}{(n-6)^2}$

La formule (19) ne fournit aucune indication sur les rapports respectifs des grandeurs empiriques σ_i^2 , σ_{i+1}^2 , qui doivent — il est vrai — être en corrélation

positive nettement accusée; comme les différences $(\sigma'_{k+1} - \sigma'_k)$ doivent différer de zéro d'une façon plus ou moins notable, il importe de connaître leurs dispersions *a priori*, soit :

$$(20) \quad E [\sigma'_{k+1} - \sigma'_k]^2 = E [(\sigma'_{k+1} - \mu_2) - (\sigma'_k - \mu_2)]^2.$$

Or les deux éléments carrés de (20) se déduisent de (19); quant au troisième terme de (20)

$$E (\sigma'_k - \mu_2) (\sigma'_{k+1} - \mu_2),$$

il est défini dans le cas où $k \leq \frac{n}{2}$,

et

$$\frac{k+1}{2} \geq \frac{2 S'_k}{C_{2k}^k C_{2k+2}^{k+1}}$$

par l'expression suivante :

$$(21) \quad E (\sigma'_k - \mu_2) (\sigma'_{k+1} - \mu_2) = \frac{\mu_4 - 3 \mu_2^2}{n - k} \left[1 - \frac{2 S'_k}{C_{2k}^k C_{2k+2}^{k+1} (n - k - 1)} \right] \\ + \frac{2 \mu_2^2}{n - k} \left[\frac{C_{4k+1}^{2k+1}}{C_{2k}^k C_{2k+2}^{k+1}} \cdot \frac{2n - 2k - 1}{n - k - 1} - \frac{k + 1}{2(n - k - 1)} \right],$$

où

$$S'_{(k)} = \sum_{i=0}^{k-1} (C_i C_{i+1})^2 + 2 \sum_{i=0}^{k-2} (C_i C_{i+1})^2 + 3 \sum_{i=0}^{k-3} (C_i C_{i+1})^3 + \dots + k (C_0 C_{k+1})^2. \\ S'_{(1)} = 1, S'_{(2)} = 15, S'_{(3)} = 242, S'_{(4)} = 3.815, S'_{(5)} = 59.724, \dots$$

Ceci étant, si l'on porte les valeurs fournies par les relations (19) et (20) dans (21), on trouve — en négligeant les grandeurs de l'ordre de $\frac{1}{n^3}$, le tableau des valeurs de $E (\sigma'_{k+1} - \sigma'_k)$.

TABLEAU II

Ordre de la dispersion $E (\sigma'_{k+1} - \sigma'_k)^2$	Valeur de la dispersion <i>a priori</i> , jusqu'au terme de l'ordre $\frac{1}{n^2}$, dans le cas d'une loi de distribution constante de la variable s .	Cas où l'on ne fait intervenir que les termes de l'ordre de $\frac{1}{n}$.
$k = 0; E (\sigma'_1 - \sigma'_0)^2 =$	$\frac{0,5 (\mu_4 - 3 \mu_2^2)}{n (n - 1)} + \frac{(n + 1) \mu_2^2}{n (n - 1)}$	$\frac{\mu_2^2}{n - 1}$
$k = 1; E (\sigma'_2 - \sigma'_1)^2 =$	$\frac{0,2778 (\mu_4 - 2 \mu_2^2)}{(n - 1) (n - 2)} + \frac{(0,2222 n + 1,111) \mu_2^2}{(n - 1) (n - 2)}$	$\frac{0,2222}{n - 2} \mu_2^2$
$k = 2$	$\frac{0,2544 (\mu_4 - 3 \mu_2^2)}{(n - 2) (n - 3)} + \frac{(0,1089 n + 1,0933) \mu_2^2}{(n - 2) (n - 3)}$	$\frac{0,1089 \mu_2^2}{n - 3}$
$k = 3$	$\frac{0,2096 (\mu_4 - 3 \mu_2^2)}{(n - 3) (n - 4)} + \frac{(0,0673 n + 1,0808) \mu_2^2}{(n - 3) (n - 4)}$	$\frac{0,0673 \mu_2^2}{n - 4}$
$k = 4$	$\frac{0,1873 (\mu_4 - 3 \mu_2^2)}{(n - 4) (n - 5)} + \frac{(0,0468 n + 1,0721) \mu_2^2}{(n - 4) (n - 5)}$	$\frac{0,0468 \mu_2^2}{n - 5}$
$k = 5$	$\frac{0,1693 (\mu_4 - 3 \mu_2^2)}{(n - 5) (n - 6)} + \frac{(0,035 n + 1,0656) \mu_2^2}{(n - 5) (n - 6)}$	$\frac{0,035 \mu_2^2}{n - 6}$

Il n'est pas sans intérêt de signaler que si les formules du tableau II ne peuvent remplacer celles de tableau I, toutefois elles les complètent; nous faisons aussi observer qu'à côté de ces σ'_i , il est bon de faire intervenir le coefficient ${}_k r'_j = \frac{\Delta^{2k} r'^{k+j}}{\Delta^{2k} r'^k}$, où j représente un nombre positif compris entre 0 et $(n - k - 1)$, et rappelons que si l'on a

$$E({}_k r'_l) = (-1)^l \frac{(k!)^2}{(k-l)!(k+l)!}$$

pour $l = (1, 2, \dots, j)$, il en résulte

$$E \sigma'_{k+1} = E \sigma'_{k+2} = \dots = E \sigma'_{k+j}$$

Remarques relatives à l'examen de deux séries statistiques (u_i, v_i).

A la série statistique $u_1, u_2, u_3, \dots, u_n$, associons une autre série statistique v_1, v_2, \dots, v_n , que nous traiterons comme la série u , en décomposant v_i au moyen d'une composante ajustée G'_i et d'une composante complémentaire s'_i , de telle façon que pour une certaine valeur de k , l'on ait $\Delta^k v_i = \Delta^k s'_i$, avec $\Delta^k u_i = \Delta^k s'_i$.

Revenons aux notations de Tschuprow :

$$m_{g\ h} = E s_i^g s'^h; m_{g\ 0} = E s_i^g; m_{0\ h} = E s'^h;$$

$$m'_{g\ h} = E s_i^g s'_{i+j}; \mu_{g\ h} = E (s_i - m_{1\ 0})^g (s'_i - m_{0\ 1})^h; \mu'_{g\ h} = E (s_i - m_{1\ 0})^g (s'_{i+j} - m_{0\ 1})^h;$$

et désignons par $p'_{(k)}$ la grandeur numérique :

$$p'_{(k)} = \frac{1}{C_{2\ k}^k} \frac{\sum_{i=1}^{n-k} \Delta^k s_i \Delta^k s'_i}{n - k}$$

et par

$$p'_{j(k)} = \frac{1}{C_{2\ k}^k} \frac{\sum_{i=1}^{n-k} \Delta^k s_i \Delta^k s'_{i+j}}{n - k}$$

On démontre que $E p'_{(k)} = \mu_{1\ 1}$, lorsque les grandeurs s_i et s_{i+j} , pour chaque valeur de $j \neq 0$, sont complètement indépendantes, et par suite $\mu_{1\ 1}^{(j)} = 0$, avec $j \neq 0$.

Quant à $E (p'_{(k)} - \mu_{1\ 1})^2$, on en détermine la valeur en recourant à la méthode qui a servi à l'établissement de la formule (19).

$$(22) \quad E [p'_{(k)} - \mu_{1\ 1}]^2 = \sigma_{(p,k)}^2 = \frac{\mu_{2\ 2} - \mu_{2\ 0} \mu_{0\ 2} - 2 \mu_{2\ 1}^2}{n - k} \left[1 - \frac{2 S_{(k)}}{(C_{2\ k}^k)^2 (n - k)} \right] + \frac{\mu_{2\ 1}^2 + \mu_{2\ 0} \mu_{0\ 2}}{n - k} \left[\frac{C_{1\ k}^k}{(C_{2\ k}^k)^2} - \frac{k}{2(n - k)} \right],$$

et l'on constate que l'on passe de l'une à l'autre par les substitutions de

$$\mu_{2\ 2} \text{ à } (\mu_{2\ 1}^2 + \mu_{2\ 0} \mu_{0\ 2}), \text{ et de } (\mu_{2\ 1} - 3 \mu_{2\ 2}) \text{ à } (\mu_{2\ 2} - \mu_{2\ 0} \mu_{0\ 2} - 2 \mu_{2\ 1}^2);$$

dans le cas d'une surface de corrélation normale, la valeur de

$$E [p'_{(k)} - \mu_{1\ 1}]^2$$

se réduit à

$$\frac{(\mu^2_{1|1} + \mu_{2|0} \mu_{0|2})}{n - k} \left[\frac{C_{1k}^2}{C_{2k}^2} - \frac{k}{2(n - k)} \right].$$

Il résulte de ce qui précède, que l'on peut faire usage du tableau I pour le calcul de $\sigma^2_{(p,k)}$, en ayant soin de faire intervenir les substitutions afférentes à μ^2_2 et $(\mu_4 - 3 \mu^2_2)$ citées ci-dessus; dans les mêmes conditions, on peut utiliser le tableau II pour la détermination de l'espérance mathématique du carré de la différence $(p'_k - p'_{k-1})$ (1).

De quelques observations relatives à la constitution interne d'une série.

Au cours de notre étude, nous avons établi un certain nombre de formules donnant les valeurs de $\sigma'^2_{k-1}, \sigma'^2_k, \dots, p'_{(k-1)}, p'_{(k)}, \dots$ en les déduisant d'un schéma basé sur une loi de distribution constante au cours des n épreuves, et en recourant à la méthode des différences.

Or, l'emploi de cette méthode, ou de toute autre, ne peut nous permettre de déterminer à l'aide d'un système de n équations les $2n$ inconnues (G_i, s_i) , à moins que nous ne possédions sur la suite u_i à l'étude des indications d'ordre analytique, et que nous ne fassions une série d'hypothèses plus ou moins fondées.

Il est évident que l'on peut toujours — lorsque l'on se trouve en présence d'un grand nombre n d'observations, représenter le phénomène au moyen d'une parabole d'ordre $(n - 1)$; mais il est évident qu'une telle solution doit être rejetée, et l'on ne peut également pas songer à faire usage d'un développement en série de Fourier.

L'intuition du chercheur consiste alors à fixer un degré $(k - 1 = 2, 3, 4, \dots)$ pour la parabole, tel que les calculs afférents à l'ajustement ne soient point trop pénibles, et les évaluations qui s'en déduisent mettent en évidence des σ'_j et $p'_{(j)}$, (pour $j = k, k + 1, \dots$), qui soient relativement stables.

Plus les variations des $(\sigma'_k, \dots, \sigma'_{k+l})$, $(p'_{(k)}, \dots, p'_{(k+l)})$ seront faibles, plus le nombre l sera élevé, plus l'hypothèse initiale trouvera une certaine confirmation.

On peut aussi considérer le cas où le phénomène se décompose en trois éléments : composante séculaire, composante sessionnelle et composante accidentelle, et adjoindre comme l'ont fait K. Pearson et Elderton, une fonction périodique aux composantes séculaire et accidentelle.

Seules des hypothèses judicieusement choisies à la suite d'un examen critique des conditions d'évolution du phénomène à travers le temps, conduisent à l'adaptation de composantes séculaires et sessionnelles vraiment rationnelles; aussi l'association intellectuelle de l'économiste et du statisticien

(1) On peut comparer les résultats obtenus par Anderson avec ceux figurant dans les derniers travaux du professeur Tschuprow, publiés en russe « sur l'espérance mathématique du quotient de deux variables aléatoires non indépendantes, et le problème de base de la théorie mathématique des différents procédés relatifs aux recherches statistiques d'interdépendance entre deux variables aléatoires.

(Voir aussi : L. SLUTSKY, sur « les schémas de corrélation et sur les erreurs systématiques de coefficients de corrélation ».)

nous semble-t-elle absolument désirable, toutes les fois qu'il s'agit de l'étude de phénomènes économiques plus ou moins complexes.

Revenons maintenant au critère relatif aux séries oscillantes défini par la formule (15 a) et les relations (16 a), et rappelons que si les grandeurs ${}_0D_{i+1}$ sont nulles, c'est-à-dire si :

$$(23) \text{ pour } \begin{cases} i = 0 \\ i = 1 \\ i = 2 \\ \vdots \\ i = k - 1 \end{cases} \left\{ \begin{array}{l} -r'_1 = 0, \\ -r'_1 + r'_2 = 0 \\ -5r'_1 + 8r'_2 - 3r'_3 = 0 \\ \vdots \\ (-1) C_{2k}^k r'_1 + 2^2 C_{2k-2}^{k-1} r'_2 + \dots + (-1)^k k^2 C_{2k}^{k-1} r'_1 = 0 \end{array} \right.$$

ou encore :

$$(23 a) \text{ pour } \begin{cases} i = 1 \\ i = 2 \\ \vdots \\ i = k \end{cases} \left\{ \begin{array}{l} -\nabla r'_1 = 0 \\ -2 \nabla' r'_1 - 3 \nabla'' r'_1 = 0 \\ \vdots \\ (-1) 2 C_{2k-2}^{k-1} \nabla' r'_1 - (2,3) C_{2k-3}^{k-1} \nabla'' r'_1 \dots - k(k-1) C_{k-1}^{k-1} \nabla^{(k)} r'_k = 0 \end{array} \right.$$

l'on a :

$$E \sigma_1'^2 = E \sigma_2'^2 = \dots = E \sigma_{k-1}'^2 = E \sigma_k'^2.$$

Alors que l'on voit apparaître dans le calcul normal des différences les éléments $(r'_1 - r'_0)$, $(r'_2 - 2r'_1 + r'_0)$, ici l'on se trouve respectivement en présence des éléments $(r'_1 - r'_2)$, $(r'_1 - 2r'_2 + r'_3)$; pour cette raison, l'on a été conduit à introduire les symboles $\nabla' r'_1 = r'_1 - r'_2$, $\nabla'' r'_1 = r'_1 - 2r'_2 + r'_3$, et à former le système (23 a) qui n'est que la transposition du système (23).

Les systèmes d'égalités

$$\begin{aligned} E \sigma_1'^2 &= \dots = E \sigma_{k-1}'^2 \\ E \sigma_3'^2 &= E \sigma_4'^2 = \dots = E \sigma_{k-1}'^2 \end{aligned}$$

correspondent respectivement à $i = 1$ et $i = 3$; dans ce dernier cas, toutes les équations des systèmes (23) et (23 a) sont satisfaites à l'exception des deux premières.

Si nous faisons $i = 1$ dans le système (23), nous avons le système des $(k - 1)$ équations ${}_0D_j = 0$, avec $j = (2, 3, \dots, (k - 1))$, pour la détermination des k inconnues r'_1, r'_2, \dots, r'_k , et nous en tirons $r'_1 = r'_2 = \dots = r'_k$, égalités caractérisant les suites uniformes de Tschuprow.

Dans le cas où les ${}_0D_{i+1}$ ne sont nulles qu'à partir de $i = l$ jusqu'à $i = k - 1$, alors on ne dispose plus que de $(k - l)$ équations pour la détermination de k inconnues; $(k - l)$ inconnues sont exprimées au moyen de fonctions linéaires des l inconnues r'_1, r'_2, \dots, r'_l .

Examinons maintenant le cas où la suite $E \sigma_1'^2, E \sigma_2'^2, E \sigma_3'^2, \dots$ ne peut devenir constante, et supposons qu'à partir d'une certaine valeur de i , l'on ait :

$$E \sigma_i'^2 < E \sigma_{i+1}'^2 < E \sigma_{i+2}'^2 \dots;$$

les suites qui jouissent de cette propriété rentrent dans le groupe « Z ».

Les suites pour lesquelles on aurait au contraire à partir d'un certain i :

$$E \sigma_i'^2 > E \sigma_{i+1}'^2 > E \sigma_{i+2}'^2 \dots,$$

appartiennent au groupe « G ».

Étudions les suites répondant aux conditions

$$E \sigma_0'^2 > E \sigma_1'^2 > E \sigma_2'^2 > \dots,$$

et ajoutons à cet ensemble l'inégalité — $r'_1 < 0$; nous remarquons de suite que ce système entraîne les inégalités

$$r'_1 > 0; r'_1 - r'_2 > 0; \quad r'_2 - r'_1 < \frac{3}{2}(r'_1 - 2r'_2 + r'_3),$$

puis un ensemble d'inégalités infiniment plus complexes.

Toutefois, si nous prenons :

$$(24) \quad r'_1 > 0; r'_1 > r'_2; \nabla' r'_1 > \nabla' r'_2; \dots, \nabla^{(k-1)} r'_1 > \nabla^{(k-1)} r'_2,$$

les premiers membres de (23 a) seront certainement inférieurs à zéro.

Il est évident que si l'on fait $r_i = p^i$, avec $0 < p < 1$, les inégalités (24) se trouvent réalisées.

On peut aussi élaborer un système (25) répondant aux conditions d'inégalité :

$$(25) \quad r'_1 < 0; r'_1 < r'_2; \nabla' r'_1 < \Delta' r'_2, \dots, \nabla^{(k-1)} r'_1 < \nabla^{(k-1)} r'_2;$$

il suffit pour cela de prendre :

$$\text{soit} \quad r_1 = p, r_2 = p^2, \dots, r_i = p^i, \text{ avec } 0 > p \geq -1,$$

soit $r_1 = A - a, r_2 = A + a, \dots, r_i = A + (-1)^i a$, si $a > |A|$, et $|A| + a < 1$,

soit encore :

$$r_i = A + (-1)^i (a + ih),$$

en ayant soin toutefois de prendre

$$|kh| < a > 0,$$

si h est négatif, et aussi :

$$|A| + a + |kh| < 1.$$

CHAPITRE II

De l'emploi de la méthode des différences d'après Karl Pearson et Ethel Elderton (1).

Le problème que se sont posés Karl Pearson et Elderson était le suivant : alors que dans l'étude de deux séries statistiques l'on s'est débarrassé de la tendance séculaire, à l'aide d'une parabole d'un certain ordre ou d'une courbe particulière, quel résultat peut-on attendre de la méthode des différences.

D'après ces auteurs, les écarts x et y des deux variables représentatives des deux phénomènes à partir des tendances séculaires, dépendent des facteurs qui sont manifestement d'un caractère non périodique, tels que l'état des connaissances médicales et de l'hygiène et de leur mode d'application, la législation, les transports modernes, l'état de la production industrielle et agricole, les guerres, les disettes d'argent et de produits, l'émigration et l'immigration, la natalité, la nuptialité, la mortalité, etc...

(1) De la méthode des différences. (Exposé fait à la Société de Biometrie et de Statistique.) — Voir *Biométrika*, t. XIV, p. 281-310.

Ils se sont préoccupés — et c'est en cela que réside la grande originalité de leur étude, tout d'abord de faire intervenir à côté de l'élément séculaire, un ou plusieurs éléments périodiques, et enfin un élément accidentel, puis de tenir compte d'une remarque d'ordre capital de Warren-Persons, qui consiste en ce que les séries de variables aléatoires peuvent être à liaisons internes, c'est-à-dire qu'une des variables ne soit pas indépendante de toutes les autres.

Rappelons à ce propos que si les grandeurs observées d'une série sont annuelles, par exemple, et qu'elles ne sont pas indépendantes, on peut, à titre d'hypothèse simplificatrice, supposer que la probabilité d'une valeur x_i dépend seulement de x_{i-1} , puis en élargissant le domaine d'action que x_i dépendra de $x_{i-1}, x_{i-2}, \dots, x_{i-p}$, sans omettre que les liaisons par groupes se répètent le long de la série (1).

Présence d'un terme périodique.

Karl Pearson et Elderton font tout d'abord remarquer que si la variable périodique peut être de la forme $A \sin (nt + \alpha)$, l'on ne possède dans la plupart des cas comme données que son intégrale dans l'unité de temps, et c'est cette dernière intégrale qu'ils font intervenir.

$$(26) \quad x_p = \int_{(p-1)h}^{ph} A \sin (nt + \alpha) dt = A' \sin \left[\alpha + 2\pi \left(p - \frac{1}{2} \right) \frac{h}{\tau} \right],$$

où h est l'intervalle et $\tau = \frac{2\pi}{n}$ la période, avec

$$A' = \frac{A\tau}{\pi} \sin \frac{\pi h}{\tau}.$$

Formant la différence centrale

$$\delta^2 x_p = x_{p+1} - 2x_p + x_{p-1} = -A' 2^2 \sin \frac{2\pi h}{\tau} \sin \left[\alpha + \frac{2\pi \left(p - \frac{1}{2} \right) h}{\tau} \right],$$

puis

$$\delta^{2q} x_p = (-1)^q \left(2^2 \sin^2 \frac{\pi h}{\tau} \right)^q x_p,$$

ces statisticiens calculent la valeur moyenne de $\int_{x_p}^{2q}$ qu'ils représentent par le symbole $[\overline{\delta^{2q} x_p}]$.

$$(27) \quad [\overline{\delta^{2q} x_p}] = (-1)^q \left(2^2 \sin^2 \frac{\pi h}{\tau} \right)^q [x_p] = (-1)^q \left(2^2 \sin^2 \frac{\pi h}{\tau} \right)^q \frac{A' \sin \left(\alpha + \frac{\pi \lambda h}{\tau} \right) \sin \frac{\pi \lambda h}{\tau}}{\lambda \sin \frac{\pi h}{\tau}},$$

où λ est le nombre d'intervalles considéré.

(1) M. Darmois a présenté d'une manière fort judicieuse une synthèse des études relatives aux séries aléatoires régulièrement enchaînées dans un article du *Métron*, vol. VIII, 1929, intitulé : « Analyse et comparaison des séries statistiques qui se développent dans le temps » ; le lecteur y trouvera toute une série de remarques intéressantes au sujet de l'emploi de la méthode des différences et de la fonction caractéristique.

Si λh est petit par rapport à τ , l'expression (27) aura une limite égale à $\sin \alpha$; si τ est petit par rapport à λh (intervalle total d'observation), l'expression sera de l'ordre de $\frac{1}{\lambda}$, et tend à devenir négligeable. Ce second cas correspond à celui de l'introduction des termes à courte période.

Quant à l'écart standard afférent à la $q^{\text{ième}}$ différence centrale, il a pour valeur moyenne :

$$(28) \quad [(\delta^{2q} \overline{x_p})^2] = \left(2^2 \sin^2 \frac{\pi h}{\tau}\right)^{2q} A'^2 \left[\sin^2 \left(\alpha + \frac{2\pi \left(p - \frac{1}{2}\right) h}{\tau} \right) \right];$$

comme $\sin^2 z = \frac{1}{2} (1 - \cos 2z)$, on remarque que pour les termes à courte période, l'expression (28) se réduit à :

$$(28)' \quad [(\delta^{2q} \overline{x_p})^2] = \frac{1}{2} \left(2^2 \sin^2 \frac{2\pi h}{\tau}\right)^{2q} A'^2 \frac{\tau^2}{\pi^2} \sin^2 \frac{2\pi h}{\tau}.$$

Calculons maintenant la corrélation entre deux termes périodiques x_p et $y_p = B \sin(n't + \beta)$, avec $n' = \frac{2\pi}{\tau'}$; pour cela, formons :

$$[(\delta^{2q} \overline{x_p})^2] \cdot [(\delta^{2q} \overline{y_p})^2],$$

et

$$(29) \quad [\delta^{2q} \overline{x_p} \delta^{2q} \overline{y_p}] = \left(2^2 \sin^2 \frac{2\pi h}{\tau}\right)^q \left(2^2 \sin^2 \frac{2\pi h}{\tau'}\right)^q A B \frac{\tau \tau'}{\pi^2} \sin \frac{\pi h}{\tau} \sin \frac{\pi h}{\tau'} \times$$

$$\left[\sin \left\{ \alpha + \frac{2\pi \left(p - \frac{1}{2}\right) h}{\tau} \right\} \sin \left\{ \beta + \frac{2\pi \left(p - \frac{1}{2}\right) h}{\tau'} \right\} \right].$$

Le dernier crochet (valeur moyenne d'un produit de sinus), peut être facilement calculé et a pour valeur :

$$\frac{1}{2\lambda} \left[\frac{\cos \left\{ \alpha - \beta + \pi \lambda h \left(\frac{1}{\tau} - \frac{1}{\tau'} \right) \right\} \sin \pi h \lambda \left(\frac{1}{\tau} - \frac{1}{\tau'} \right)}{\sin \pi h \left(\frac{1}{\tau} - \frac{1}{\tau'} \right)} - \frac{\cos \left\{ \alpha + \beta + \pi \lambda h \left(\frac{1}{\tau} + \frac{1}{\tau'} \right) \right\} \sin \pi h \lambda \left(\frac{1}{\tau} + \frac{1}{\tau'} \right)}{\sin \pi h \left(\frac{1}{\tau} + \frac{1}{\tau'} \right)} \right].$$

Dans le cas où les périodes sont courtes, l'expression précédente est petite en raison du facteur $\frac{1}{\lambda}$; toutefois, si $\tau = \tau'$, on remarque que l'expression (29) se réduit à :

$$[\delta^{2q} \overline{x_p} \delta^{2q} \overline{y_p}] = \left(2^2 \sin^2 \frac{2\pi h}{\tau}\right)^{2q} A B \frac{\tau^2}{\pi^2} \sin^2 \frac{\pi h}{\tau} \cdot \frac{1}{2} \cos(\alpha - \beta),$$

et que la corrélation entre deux termes à périodes courtes est égale à $\cos(\alpha - \beta)$, c'est-à-dire à 1, s'il n'y a pas de différence de phase.

On constate, de plus, que ce résultat est indépendant de 1, si l'on opère avec les différences centrales.

Il est évident que si le calcul nous conduit à une corrélation négligeable, on est amené à conclure qu'il ne se trouve pas un terme simple commun aux deux variables, ou que $\alpha - \beta = \frac{\pi}{2}$.

Non existence de termes périodiques.

Supposons maintenant que nous n'ayons que des FLUCTUATIONS NON PÉRIODIQUES; la méthode de Student nous donne :

$$\begin{aligned} [\overline{\delta^{2q} X_p}]^2 &= \frac{4q!}{(2q!)^2} \sigma_x^2, \\ [(\overline{\delta^{2q} Y_p})^2] &= \frac{4q!}{(2q!)^2} \sigma_y^2, \\ [\overline{\delta^{2q} X_p} \overline{\delta^{2q} Y_p}] &= \frac{4q!}{(2q!)^2} \sigma_x \sigma_y r_{XY}, \end{aligned}$$

et par suite r_{XY} = corrélation de $\delta^{2q} x_p$ et de $\delta^{2q} y_p$; dans ce cas spécial, on n'a pas besoin d'écarter la tendance séculaire.

Nous rappelons à ce propos que ce que nous venons de dire subsiste si X_p et Y_p , sont en corrélation [($p' - p$) restant constant]; cela équivaut à la considération du phénomène P_2 décalé dans le temps de ($p' - p$) unités par rapport au phénomène P_1 .

Lorsque l'on se trouve en présence de séries en chaîne, l'on doit alors écrire :

$$\begin{aligned} [(\overline{\delta^{2q} X_p})^2] &= \frac{4q!}{(2q!)^2} \Phi(\rho') \sigma_x^2, \\ [(\overline{\delta^{2q} Y_p})^2] &= \frac{4q!}{(2q!)^2} \Phi(\rho'') \sigma_y^2, \\ (30) \quad [\overline{\delta^{2q} X_p} \overline{\delta^{2q} Y_p}] &= \frac{4q!}{(2q!)^2} \Phi(\rho) \sigma_x \sigma_y r_{XY}, \end{aligned}$$

d'où il résulte :

$$r_{XY} = \frac{\sqrt{\Phi(\rho') \Phi(\rho'')}}{\Phi(\rho)} \times \text{corrélation de } (\delta^{2q} X_p \text{ et } \delta^{2q} Y_p).$$

COMBINAISON COMPLEXE

Si l'on se trouve en présence d'une combinaison, où figurent une composante séculaire, un ensemble de termes périodiques et une fluctuation non périodique, on a alors :

$$\begin{aligned} x_p &= a_0 + a_1 t + \dots + a_{n-1} t^{n-1} + S_\tau \left\{ A'_\tau \sin \left[\alpha_\tau + \frac{2\pi}{\tau} \left(p - \frac{1}{2} \right) h \right] \right\} + X_p, \\ y_p &= b_0 + b_1 t + \dots + b_{m-1} t^{m-1} + S_\tau \left\{ B'_\tau \sin \left[\beta_\tau + \frac{2\pi}{\tau} \left(p - \frac{1}{2} \right) h \right] \right\} + Y_p, \end{aligned}$$

et l'on obtient pour les valeurs moyennes des éléments apparaissant dans le

calcul du coefficient de corrélation les expressions ci-dessous, en admettant — avec les auteurs — la théorie de Student valable :

$$\begin{aligned} [\overline{\delta^{2q} x_p \delta^{2q} y_p}] &= S. \left\{ \frac{1}{2} A'_{\tau} B'_{\tau} \left(2^2 \sin^2 \frac{\pi h}{\tau} \right)^{2q} \cos(\alpha_{\tau} - \beta_{\tau}) \right\} + \frac{4q!}{(2q!)^2} \sigma_x \sigma_y r_{x_p y_p}, \\ [\overline{\delta^{2q} x_p^2}] &= S. \left\{ \frac{1}{2} A'^2 \left(2^2 \sin^2 \frac{\pi h}{\tau} \right)^{2q} \right\} + \frac{4q!}{(2q!)^2} \sigma_x^2, \\ [\overline{\delta^{2q} y_p^2}] &= S_{\tau} \left\{ \frac{1}{2} B'^2 \left(2^2 \sin^2 \frac{\pi h}{\tau} \right)^{2q} \right\} + \frac{4q!}{(2q!)^2} \sigma_y^2. \end{aligned}$$

Karl Pearson et Elderton ont montré que dans l'hypothèse de l'intervention d'un seul terme à courte période, tant pour x_p que pour y_p , ce même terme n'a aucune influence si la période τ est de 4 fois ou plus l'intervalle choisi h ; à moins que la fluctuation non périodique ne soit très petite en comparaison du terme périodique, l'on constate la disparition du terme périodique à la 8^e ou la 10^e différence.

Cas des liaisons en chaîne.

Revenons maintenant aux liaisons en chaîne, et désignons respectivement par :

ρ'_{+s} } la corré- } X_p et X_{p+s} , ρ''_{+s} } la corré- } Y_p et Y_{p+s} , ρ_s } la corré- } X_p et Y_{p+s} ,
 ρ'_{-s} } lation de } X_p et X_{p-s} , ρ''_{-s} } lation de } Y_p et Y_{p-s} , ρ_{-s} } lation de } X_p et Y_{p-s} ,

et remarquons que l'on obtient, en suivant un processus de calcul dû à M. Es. Pearson (1921) :

$$(31) \frac{[\overline{\Delta^n X_p \Delta^n Y_p}]}{\sigma_x \sigma_y} = \frac{(2n)!}{(n!)^2} \left\{ (-1)^n \frac{(n!)^2}{(2n)!} \rho_{-n} + \dots + \frac{n!(n-1)}{(n+1)(n+2)^{\rho-2}} - \frac{n}{n+1} \rho_{-1} \right. \\ \left. + \rho_0 - \frac{n}{n+1} \rho_1 + \frac{n(n-1)}{(n+1)(n+2)} \rho_2 \dots \dots + (-1)^n \frac{(n!)^2}{(2n)!} \rho_n \right\} = \frac{(2n)!}{(n!)^2} \Phi(n, \rho_0, \rho),$$

et, de même :

$$(31)' \quad \frac{[\overline{\Delta^{n+1} X_p \Delta^{n+1} Y_p}]}{\sigma_x \sigma_y} = \frac{(2n+2)!}{[(n+1)!]^2} \Phi(n+1, \rho_0, \rho),$$

et par suite :

$$(32) \quad \frac{[\overline{\Delta^{n+1} X_p \Delta^{n+1} Y_p}]}{[\overline{\Delta^n X_p \Delta^n Y_p}]} = \left(4 - \frac{2}{n+1} \right) \frac{\Phi(n+1, \rho_0, \rho)}{\Phi(n, \rho_0, \rho)};$$

On trouve immédiatement par analogie :

$$(33) \quad \frac{[\overline{(\Delta^{n+1} X_p)^2}]}{[\overline{(\Delta^n X_p)^2}]} = \left(4 - \frac{2}{n+1} \right) \frac{\Phi(n+1, 1, \rho')}{\Phi(n, 1, \rho')},$$

$$(34) \quad \frac{[\overline{(\Delta^{n+1} Y_p)^2}]}{[\overline{(\Delta^n Y_p)^2}]} = \left(4 - \frac{2}{n+1} \right) \frac{\Phi(n+1, 1, \rho'')}{\Phi(n, 1, \rho'')}.$$

Si maintenant l'on pose $\bar{\rho}_s = \frac{\rho_s}{\rho_0}$, l'on peut écrire :

$$\frac{[\overline{\Delta^n X_p \Delta^n Y_p}]}{\sqrt{[\overline{\Delta^n X_p^2}] \cdot [\overline{\Delta^n Y_p^2}]}} = \rho_0 \frac{\Phi(n, \bar{\rho})}{\sqrt{\Phi(n, \rho') \Phi(n, \rho'')}},$$

et l'on est alors conduit à considérer les séries ($\varrho, \varrho', \varrho''$).

En désignant par :

$$Z_{n,o} \left(4 - \frac{2}{n+1} \right), Z'_{n,o} \left(4 - \frac{2}{n+1} \right) \text{ et } Z''_{n,o} \left(4 - \frac{2}{n+1} \right)$$

les seconds membres des équations (32), (33) et (34), pour bien spécifier que l'on fait intervenir $\Delta^n x_p$ et $\Delta^n y_p$ et non $\Delta^n x_p$ et $\Delta^n y_{p+s}$, nous avons alors les 3 équations :

$$Z_{n,o} = \frac{\Phi(n+1, \bar{\rho})}{\Phi(n, \bar{\rho})}, Z'_{n,o} = \frac{\Phi(n+1, \rho')}{\Phi(n, \rho')}, Z''_{n,o} = \frac{\Phi(n+1, \rho'')}{\Phi(n, \rho'')},$$

grâce auxquelles l'on déterminera $\bar{\rho}, \rho'$ et ρ'' .

Il est essentiel de remarquer que l'on voit figurer dans $Z_{n,o}$ les inconnues

$$\bar{\rho}_{-(n+1)}, \dots, \bar{\rho}_{-1}, \bar{\rho}_1, \dots, \bar{\rho}_{n+1},$$

et dans $Z_{n+1,o}$ apparaissent en outre $\bar{\rho}_{-(n+2)}$ et $\bar{\rho}_{n+2}$.

Si l'on suppose le terme séculaire écarté par la 4^e différence, on peut, en raison de la présence du facteur $\frac{(n!)^2}{(2n!)}$ écarter $\bar{\rho}_4$ et $\bar{\rho}_{-4}$, et les $\bar{\rho}$ d'indices supérieurs à 4 en valeur absolue, il s'en suit que l'on se trouve en présence des inconnues : $\bar{\rho}_{-3}, \bar{\rho}_{-2}, \bar{\rho}_{-1}, \bar{\rho}_1, \bar{\rho}_2, \bar{\rho}_3$, dont le calcul nécessitera 6 équations.

Il faudra donc faire état des valeurs de \bar{Z}_n , correspondant à $n = 4, 5, \dots, 9$, et faire intervenir par suite les différences d'ordre 10.

Si les hypothèses initiales de Student sont exactes, l'on doit trouver 0 pour tous les ϱ , sauf un.

Cette méthode repose sur la précision des différences d'ordre supérieur, et celle-ci n'est pas toujours réalisée; aussi les auteurs ont alors songé à introduire des corrélations d'un type tel que le risque dû aux différences d'ordre élevé fût éliminé, et pour cela ont fait appel aux corrélations afférentes à $\Delta^n X_p$ et $\Delta^n Y_{p+s}$.

C'est ainsi que :

$$(35) \frac{[\Delta^n X_p \Delta^n Y_{p+1}]}{\sigma_x \sigma_y} = \frac{(2n)!}{(n!)^2} \left[(-1) \frac{n(n!)^2}{(2n)!} \rho_{-(n-1)} + \dots + \frac{n(n-1)}{(n+1)(n-2)} \rho_{-1} \right. \\ \left. - \frac{n}{n+1} \rho_0 + \rho_1 - \frac{n}{n+1} \rho_2 + \frac{n(n-1)}{(n+1)(n+2)} \rho_3 + \dots + (-1) \frac{n(n!)^2}{(2n)!} \rho_{n+1} \right],$$

et l'on forme de même :

$$(35') \frac{[\Delta^n X_p \Delta^n Y_{p-1}]}{\sigma_x \sigma_y} = \frac{(2n)!}{(n!)^2} \left[(-1) \frac{n(n!)^2}{(2n)!} \rho_{-(n+1)} \dots + \frac{n(n-1)}{(n+1)(n+2)} \rho_{-3} \right. \\ \left. - \frac{n}{n+1} \rho_{-2} + \rho_{-1} - \frac{n}{n+1} \rho_0 + \dots + (-1)^n \frac{n!}{(2n)!} \rho_{n-1} \right].$$

Ces fonctions ne convergent pas aussi rapidement que (31) et (31)'; toutefois, si nous voulions comme primitivement connaître les grandeurs

$$(\bar{\rho}_{-4}, \bar{\rho}_{-3}, \bar{\rho}_{-2}, \bar{\rho}_{-1}, \bar{\rho}_1, \dots, \bar{\rho}_4),$$

nous serions amené, *en supposant que la tendance séculaire disparaisse à partir*

de la différence quatrième, à recourir à 8 équations en poussant jusqu'aux différences septièmes, et en ayant soin de contrôler les résultats au moyen d'une équation supplémentaire.

Rappelant que $\bar{\rho}_s = \frac{\rho_s}{\rho_0}$, et introduisant l'opérateur différentiel fini E agissant sur ρ_s et non sur ρ_0 , on remarque que :

$$\frac{[\Delta^{n+1} X_p \Delta^{n+1} Y_{p+1}]}{[\Delta^n X_p \Delta^n Y_{p+1}]} = \left(4 - \frac{2}{n+1}\right) \frac{E \Phi(n+1, \rho)}{E \Phi(n, \rho)};$$

Ceci étant, si l'on désigne par $Z_{n,1} \left(4 - \frac{2}{n+1}\right)$ le premier membre de l'équation précédente, on remarque qu'elle s'écrit :

$$(36) \quad Z_{n,1} = \frac{E \Phi(n+1, \bar{\rho})}{E \Phi(n, \rho)}$$

et l'on a de même :

$$(36)' \quad Z_{n,-1} = \frac{E^{-1} \Phi(n+1, \bar{\rho})}{E^{-1} \Phi(n, \bar{\rho})}$$

et les équations semblables :

$$(37) \quad \left\{ \begin{array}{l} Z'_{n,1} = \frac{E \Phi(n+1, \rho')}{E \Phi(n, \rho')}, \quad Z'_{n,-1} = \frac{E^{-1} \Phi(n+1, \rho')}{E^{-1} \Phi(n, \rho')}, \\ Z''_{n,1} = \frac{E \Phi(n+1, \rho'')}{E \Phi(n, \rho'')}, \quad Z''_{n,-1} = \frac{E^{-1} \Phi(n+1, \rho'')}{E^{-1} \Phi(n, \rho'')}, \end{array} \right.$$

qui donnent respectivement les ρ' et ρ'' .

K. Pearson et Elderton font observer à juste titre que « seule l'expérience pratique peut montrer jusqu'à quel point les équations proposées sont suffisantes pour donner des résultats utiles ».

C'est ici qu'interviennent diverses hypothèses qui permettent de simplifier la tâche du calculateur :

1° Le nombre des équations est réduit environ de moitié, si l'on suppose que $\rho'_s = \rho'_s$, relation qui correspond à l'hypothèse de la représentation d'une oscillation par un terme harmonique, cette oscillation ne pouvant être asymétrique.

Il est un cas particulièrement intéressant, c'est celui relatif qui est à $\bar{\rho}_s = \bar{\rho}_s = \rho'_s$, pour toutes les valeurs de s ; en l'occurrence, la corrélation entre $\Delta^n X_p$ et $\Delta^n Y_{p+1}$ est égale à celle de X et de Y .

La relation $\rho'_s = \bar{\rho}_s$ qui s'écrit encore : $\rho'_s \rho_0 = \rho_s$, signifie que la corrélation entre X_p et Y_{p+1} est égale à la corrélation de X_p et de Y_p multipliée par la corrélation de X_p et X_{p+1} ; comme la corrélation de (X_p, Y_p) est aussi celle de (X_{p+1}, Y_{p+1}) , il s'en suit :

Corrélⁿ $(X_p, Y_{p+1}) = \text{corrél}^n (X_{p+1}, Y_{p+1}) \times \text{corrél}^n (X_p, Y_{p+1})$, et en conséquence la corrélation de (X_p, Y_{p+1}) est nulle pour X_{p+1} constant.

On peut généraliser en supposant que les trois groupes de corrélations décroissent avec le même rythme ; alors on a :

$$\rho_s = \varepsilon^s \rho_0, \rho'_s = \varepsilon'^s, \rho''_s = \varepsilon''^s.$$

Si l'on désigne par $\Phi(n, \varepsilon)$ la série

$$(-1)^n \frac{(n!)^2}{(2n)!} \varepsilon^n + \dots + \frac{n(n-1)}{(n+1)(n+2)} \varepsilon^2 - \frac{n}{n+1} \varepsilon + 1,$$

on constate que l'on peut écrire :

$$\frac{[\Delta^{n+1} X_p \Delta^{n+1} Y_p]}{[\Delta^n X_p \Delta^n Y_p]} = \left(4 - \frac{2}{n+1}\right) \cdot \frac{2\Phi(n+1, \varepsilon) - 1}{2\Phi(n, \varepsilon) - 1},$$

$$\frac{[(\Delta^{n+1} X_p)^2]}{[(\Delta^n X_p)^2]} = \left(4 - \frac{2}{n+1}\right) \cdot \frac{2\Phi(n+1, \varepsilon') - 1}{2\Phi(n, \varepsilon') - 1},$$

$$\frac{[\Delta^{n+1} Y_p]^2]}{[\Delta^n Y_p]^2]} = \left(4 - \frac{2}{n+1}\right) \cdot \frac{2\Phi(n+1, \varepsilon'') - 1}{2\Phi(n, \varepsilon'') - 1},$$

et l'on remarque que la connaissance des premiers membres de ces 3 relations conduit à celle de $\varepsilon, \varepsilon', \varepsilon''$.

Grâce à une table de la fonction du second membre de ces mêmes relations établie par M. Henderson pour $n = (1, 2, 3, \dots, 10)$ et pour des valeurs de ε variant de -1 à $+1$ par dixième d'unité, il est possible de calculer par interpolation $\varepsilon, \varepsilon', \varepsilon''$.

Ceci étant, de la relation

$$r_{\Delta^n X_p \Delta^n Y_p} = \frac{[2\Phi(n, \varepsilon) - 1] r_{XY}}{\sqrt{[2\Phi(n, \varepsilon') - 1][2\Phi(n, \varepsilon'') - 1]}}$$

l'on tire la valeur de r_{XY} .

Examen de la fonction $\Phi(n, \varphi)$ dans le cas d'existence d'un terme périodique unique dans chacune des séries à l'étude.

Si les termes X_p sont respectivement représentés par :

$$X_p = A \frac{\tau}{\pi} \sin \frac{\pi h}{\tau} \sin \left[\alpha + 2\pi \left(p - \frac{1}{2} \right) \frac{h}{\tau} \right],$$

$$X_{p+s} = A \frac{\tau}{\pi} \sin \frac{\pi h}{\tau} \sin \left[\alpha + 2\pi \frac{s h}{\tau} + 2\pi \left(p - \frac{1}{2} \right) \frac{h}{\tau} \right],$$

leur corrélation sera définie par

$$(38) \quad \rho_{s'} = \cos \frac{2\pi s h}{\tau},$$

expression d'où l'on déduit $\varphi'_{-} = \varphi'$.

Ceci étant, posons $\gamma = \frac{2\pi h}{\tau}$, et revenons à l'expression de

$$[(\Delta^n X_p)^2] = \frac{2n!}{(n!)^2} \Phi(n, 1, \gamma'),$$

qui se déduit facilement de (31); on voit alors que l'on a :

$$(39) \quad \Phi(n, \gamma) = 1 - \frac{n}{n+1} \cos \gamma \frac{n(n-1)}{(n+1)(n+2)} \cos 2\gamma \dots + (-1)^n \frac{(n!)^2}{(2n)!} \cos n\gamma.$$

Dans ces conditions, l'on trouve :

$$(40) \quad 2 \Phi(n, \gamma) - 1 = \frac{(n!)^2}{(2n!)} \left(2 \sin \frac{\gamma}{2}\right)^{2n}$$

et par suite :

$$\left(4 - \frac{2}{n+1}\right) \frac{2 \Phi(n+1, \gamma) - 1}{2 \Phi(n, \gamma) - 1} = 2(1 - \cos \gamma);$$

il résulte de cette dernière relation que :

$$\frac{[(\Delta^{n+1} X_p)^2]}{[(\Delta^n X_p)^2]} = 2(1 - \cos \gamma),$$

et enfin :

$$(41) \quad \sin \frac{\gamma}{2} = \sin \frac{\pi h}{\tau} - \frac{1}{2} \left\{ \frac{[(\Delta^{n+1} X_p)^2]}{[(\Delta^n X_p)^2]} \right\}^{1/2}$$

Le rapport des écarts standards successifs des différences tend vers une limite, d'où nous déduisons la période dans le cas où il existe un terme périodique unique.

Revenant à l'expression de

$$[(\Delta^n X_p)^2],$$

nous remarquons qu'elle est égale à :

$$\sigma_x^2 \frac{(2n!)}{(n!)^2} \left\{ 2 \Phi(n, \gamma) - 1 \right\} = \sigma_x^2 \left(2 \sin \frac{\gamma}{2}\right)^{2n},$$

d'où l'on tire :

$$(42) \quad \sigma_x^2 = \frac{[(\Delta^n X_p)^2]}{2 \sin \frac{\gamma}{2}}.$$

Si donc une forme de réduction nous a donné σ_x^2 , il est alors possible de se rendre compte des fluctuations à partir de la tendance séculaire par l'introduction d'un terme périodique.

Pour la deuxième série, on détermine la période τ' grâce à la relation

$$\sin \frac{\pi h}{\tau'} = \frac{1}{2} \left\{ \frac{[(\Delta^{n+1} Y_p)^2]}{[(\Delta^n Y_p)^2]} \right\}$$

et expliquer ainsi la présence de σ_y^2 .

Dans l'hypothèse où il existe un terme périodique commun à x et à y , DE MÊME AMPLITUDE ET DE MÊME PÉRIODE, il est évident que :

$$x_p = A \frac{\tau}{\pi} \sin \frac{\pi h}{\tau} \sin \left\{ \alpha + 2\pi \left(p - \frac{1}{2}\right) \frac{h}{\tau} \right\},$$

$$y_p = A \frac{\tau}{\pi} \sin \frac{\pi h}{\tau} \sin \left\{ \beta + 2\pi \left(p - \frac{1}{2}\right) \frac{h}{\tau} \right\}.$$

les corrélations afférentes à

$$X_p \text{ et } Y_{p+1},$$

$$X_p \text{ et } Y_{p-1},$$

ont pour valeur :

$$\rho_s = \cos \left(\alpha - \beta - 2 \pi \frac{s \hbar}{\tau} \right),$$

$$\rho_{-s} = \cos \left(\alpha - \beta + 2 \pi \frac{s \hbar}{\tau} \right).$$

Faisant état de la relation (31), on constate que :

$$\frac{[\Delta^n X_p \Delta^n Y_p]}{\sigma_x \sigma_y} = \frac{(2n)!}{(n!)^2} \cos(\alpha - \beta) \{ 2 \Phi(n, \gamma) - 1 \} = \left(2 \sin \frac{\gamma}{2} \right)^{2n} \cos(\alpha - \beta);$$

et l'on déduit de là le résultat important suivant :

$$(43) \quad \frac{[\Delta^{n+1} X_p \Delta^{n+1} Y_p]}{[\Delta^n X_p \Delta^n Y_p]} - \left(2 \sin \frac{\gamma}{2} \right)^2 = \frac{[(\Delta^{n+1} X_p)^2]}{[(\Delta^n X_p)^2]} = \frac{[(\Delta^{n+1} Y_p)^2]}{[(\Delta^n Y_p)^2]},$$

qui permet de vérifier si les fluctuations peuvent être attribuées à un terme périodique unique.

On remarque enfin que la *différence de phase* résulte de l'expression

$$r_{\Delta^n X \Delta^n Y} = \cos(\alpha - \beta).$$

Lorsque l'on fait abstraction d'un terme périodique tant dans la constitution de x_p que dans celle de y_p , on sait d'après (32) que

$$\frac{[\Delta^{n+1} X_p \Delta^{n+1} Y_p]}{[\Delta^n X_p \Delta^n Y_p]} = \left(4 - \frac{2}{n+1} \right) \frac{\Phi(n+1, \rho_o, \rho)}{\Phi(n, \rho_o, \rho)}$$

alors que dans le cas précédent, ce même rapport est théoriquement égal à 4 pour $\gamma = \pi$, c'est-à-dire lorsque la période a pour valeur deux fois l'unité fondamentale de temps.

(A suivre.)

R. RISSER