à la base des relations entre les séries Y et X. D'après elle, l'intensité de liaison, caractérisée par le coefficient de corrélation pour les mêmes deux séries, apparaît dans la mesure de + 0·70. La seconde hypothèse, B, ressort également de la supposition du caractère linéaire de la liaison, cependant elle compare la série Y avec la série Z et trouve le coefficient de corrélation + 0·90.

Par conséquent, dans le premier cas nous aurions :

$$\frac{\sigma_\xi}{\sigma_x} \cdot \frac{\sigma_\psi}{\sigma_y} = 0{\cdot}70$$

Ici sont possibles quelques suppositions: si la componente e manque, $\xi = x$; $\frac{\sigma_\xi}{\sigma_x} = 1$ et, par conséquent, $\frac{\sigma_\psi}{\sigma_y} = 0{\cdot}70$; si c'est la componente ε qui manque, alors $\psi = y$ et $\frac{\sigma_\psi}{\sigma_y} = 1$ et, par conséquent, $\frac{\sigma_\xi}{\sigma_x} = 0{\cdot}70$; si, enfin, $\frac{\sigma_\xi}{\sigma_x} = \frac{\sigma_\psi}{\sigma_y}$, alors chacune des deux fractions est égale à $\sqrt{0{\cdot}70}$, c.-à-d. presque à 0·84. Ainsi, la véritable grandeur du rapport $\frac{\sigma_\psi}{\sigma_y}$ réside quelque part entre 0·70 et 1·0, et on peut dire que l'hypothèse A est en état d'expliquer au-moins 70 % des variations de Y au moyen de celles de X et de ses componentes. Par contre, par l'admission de l'hypothèse B nous pourrions expliquer au-moins 90 % des variations de Y. Dans l'hypothèse A de réactions latérales, on peut compter jusqu'à 30 % des variations, tandis que dans l'hypothèse B — pas plus que 10 %. Il résulte de là que dans *d'autres conditions égales,* l'hypothèse B est préférable à l'hypothèse A. Par là, cependant, nous n'aurons guère démontré la certitude de la première. *)

En se servant du coefficient de corrélation comme d'un outil de l'analyse causale, nous ne devons pas oublier que la formule

$$|r_{12}| = q_1\, q_2$$

(où r_{12} désigne le coefficient de corrélation à priori, à la différence du coefficient empirique r_{12}') est déduite par la supposition que non seulement la dépendance de ξ et de ψ soit linéaire, mais que les variables ξ et ε soient entièrement indépendantes l'une de l'autre. Cette supposition n'est légitime que lorsque ψ est vraiment une fonction linéaire de ξ; si cependant la liaison entre ξ et ψ n'est pas réellement linéaire, ou ne l'est pas entièrement, notre supposition peut paraître tout à fait fausse. Dans ce cas, comme on le voit par l'appendice au présent article,

$$|r_{12}| = q_1\, q_2 + R,$$

où R est le terme résiduel qui représente, pour ainsi dire, l'erreur systématique de la formule considérée. Si cependant R est trop petit par sa valeur absolue en comparaison avec $q_1\, q_2$, le coefficient de corrélation $|r_{12}|$ pourra toujours être considéré comme première approximation de la mesure H; d'autant plus que le coefficient empirique $|r_{12}'|$, que nous avons calculé, n'est qu'une approximation du coefficient à priori $|r_{12}|$.

Bien entendu, on peut déduire des formules également pour les cas où la dépendance de ξ et de ψ, bien que non linéaire, est entièrement déterminée quant à sa forme. Par exemple quand ψ par rapport à ξ représente une parabole, une hyperbole, une fonction potentielle, etc. On a accompli dans ce sens beaucoup de travail préparatoire (voir particulièrement dans Tchouproff). Pour ne pas trop élargir le volume du présent article, nous délaisserons cette question sans l'examiner; d'autant plus que le choix du type de la dépendance fonctionnelle de ψ et de ξ devrait être dicté par les nécessités de la théorie économique, mais tout n'est pas très bien sous ce rapport.

Au lieu d'examiner les types de la dépendance non linéaire, nous passons à présent au cas où la variable Y est corréliée en même temps à *plusieurs* variables: $X^{(1)}$, $X^{(2)}$, $X^{(3)}$ . . . . . et en outre toutes les dépendances sont des fonctions linéaires. C'est le cas de la corrélation dite „multiple" dont la théorie est beaucoup plus complexe au point de vue aussi bien mathématique que logique. L'application de nos procédés analytiques constitue, ce pendant, ici aussi un apport pour une clarté logique plus grande. Nous réussissons en outre à découvrir les sources de quelques erreurs commises par les statisticiens qui appliquent en pratique les formules de la corrélation multiple.

II.

Considérons (n + 1) séries à variables aléatoires, dont chacune se compose de N termes:

$$\left.\begin{array}{l} \text{série du zéro: } X_1^{(0)},\ X_2^{(0)},\ X_3^{(0)} \ldots\ldots X_N^{(0)}; \\ \text{première série: } X_1^{(1)},\ X_2^{(1)},\ X_3^{(1)} \ldots\ldots X_N^{(1)}; \\ \text{deuxieme serie: } X_1^{(2)},\ X_2^{(2)},\ X_3^{(2)} \ldots\ldots X_N^{(2)}; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ n\text{-}^{e} \text{ série: } X_1^{(n)},\ X_2^{(n)},\ X_3^{(n)} \ldots\ldots\ldots X_N^{(n)}. \end{array}\right\} [14]$$

Supposons que chaque terme de la série du zéro soit lié par des liaisons libres (stochastiques), avec les termes homonymes des autres séries, c'est-à-dire:

$$\left.\begin{array}{l} X_1^{(0)} = f\,(X_1^{(1)},\ X_1^{(2)},\ X_1^{(3)} \ldots X_1^{(n)}) + E_1 \\ X_2^{(0)} = f\,(X_2^{(1)},\ X_2^{(2)},\ X_2^{(3)} \ldots X_2^{(n)}) + E_2 \\ X_3^{(0)} = f\,(X_3^{(1)},\ X_3^{(2)},\ X_3^{(3)} \ldots X_3^{(n)}) + E_3 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ X_N^{(0)} = f\,(X_N^{(1)},\ X_N^{(2)},\ X_N^{(3)} \ldots X_N^{(n)}) + E_N \end{array}\right\} [15]$$

A la différence du cas de la corrélation de deux variables, nous supposons ici, suivant l'exemple général, que les variables aléatoires figurent dans la fonction f avec toute leur grandeur, c.-à.-d. que la composante ξ dans

*) Dans certains cas le résultat peut être amélioré par l'union des deux hypothèses en une et par l'explication des variations de la série Y au moyen d'une action simultanée des grandeurs X et Z. Cela nous conduit cependant au cas de la „corrélation multiple" (voir partie II).