Ce n'est qu'en passant à l'espérance mathématique de la grandeur $X_i^{(1)}$ du système [33] que nous obtenons ce qui nous est nécessaire.

D'autre part, en appliquant la formule [29] ou [30], nous pouvons trouver les valeurs suivantes des *coefficients partiels à priori de la corrélation :*

$$r_{01\cdot 2} = +\frac{1}{\sqrt{2}};\ r_{02\cdot 1} = +\frac{1}{\sqrt{2}};\ r_{12\cdot 0} = -\frac{1}{2}$$

Il n'est pas difficile de concevoir le sens des deux premiers coefficients. Nous aurions obtenu exactement le même résultat, si nous avions supposé que la composante W_i manque à l'expression $X_i^{(0)}$ (voir formule 31) et si, après cela nous avions calculé le coefficient ordinaire de la corrélation de $X_i^{(0)} = U_i + T_i$, d'une part, avec $X_i^{(1)} = U_i$, de l'autre ; ou bien, si nous avions supprimé U_i dans $X_i^{(0)}$ et après cela nous avions calculé le cofficient de corrélation de $X_i^{(0)} = W_i + T_i$ avec $X_i^{2} = W_i$. Que désigne, cependant le coefficient négatif de la corrélation de $X_i^{(1)} = U_i$ avec $X_i^{(2)} = W_i$, lorsque nous savons que ces grandeurs sont par elles-mêmes absolument indépendantes l'une de l'autre ? Le résultat obtenu n'est pas accidentel, puisque si nous supprimons entièrement la composante T_i et si nous posons

$$X_i^{(0)} = W_i + U_i;\ X_i^{(1)} = W_i;\ X_i^{(2)} = U_i,$$

nous obtenons la valeur $r_{12.0} = -1$, une expression indiquant une proportionnalité entièrement inverse entre deux grandeurs qui sont, nous le savons avec certitude, indépendantes l'une de l'autre ! La réponse à cette question consiste en ce que la mise elle-même du problème chez nous a un défaut logique : si la composante $X_i^{(0)}$ représente la somme des valeurs des deux variables aléatoires $X_i^{(1)}$ et $X_i^{(2)}$, nous n'avons pas le droit d'admettre logiquement, dans la mesure de l'intensité de la liaison entre ces dernières grandeurs, que la composante $X_i^{(0)}$, c'est-à-dire leur somme, reste constante quand les différentes quantités additives changent de valeur. Ce qui est tout à fait convenable et juste au point de vue de l'analyse formelle d'une équation algébrique, peut paraître n'avoir aucun sens quand nous tendons à découvrir les dépendances de causalité entre les phénomènes. Supposons, par exemple, que $X_i^{(1)}$ désigne la quantité de pluies tombées en m. m. au cours d'un mois quelconque à Varna, que $X_i^{(2)}$ montre le nomre des oranges consommées par le Président des Etats Unis au cours du même mois. L'algèbre ne nous empêche pas de faire la somme

$$X_i^{(0)} = X_i^{(1)} + X_i^{(2)}$$

et de faire subir à $X_i^{(0)}$, $X_i^{(1)}$, $X_i^{(2)}$ différentes opérations mathématiques, en calculant entre autres la valeur

$$\frac{r_{12} - r_{01}\, r_{02}}{\sqrt{(1 - r_{01}^2)\,(1 - r_{02}^2)}}$$

Mais si nous affirmons, sur la base du résultat obtenu $r_{12.0} = -1$, *qu' „en éliminant l'influence de la composante* $X_i^{(0)}$" il existera en effet entre $X_i^{(1)}$ et $X_i^{(2)}$ une dépendance de raison inverse et que, par conséquent, plus il pleuvra à Varna, moins d'oranges seront consommées par le Président à Waschington et inversement, le logicien commencera à protester énergiquement.

Aucun statisticien ne fera certainement une pareille erreur évidente. Cependant il est souvent difficile de déterminer en pratique laquelle des séries examinées représente *en réalité* la somme des fonctions linéaires des autres séries et laquelle ne la représente pas ; laquelle des séries reproduit le complexe des causes qui nous intéressent et laquelle — celui des effets. En outre, peu nombreuses sont les personnes utilisant les coefficients partiels de la corrélation, qui se rappellent à l'aide de quelles considérations et admissions sont déduits ces coefficients, et nous pouvons énumérer nombre d'exemples dans la littérature statistique où nous rencontrons, sous une forme voilée, le cas de la pluie de Varna et des oranges américaines. Nous recommandons encore une fois une prudence maximum lors du calcul et particulièrement lors de l'interprétation des coefficients partiels de corrélation.

Revenons maintenant à la formule [18] ou [24].

On sait que dans cette formule les valeurs des écarts-type et des coefficients b sont des valeurs à priori, c.-à.-d. ne peuvent être déterminées exactement dans la plupart des cas. Ordinairement on ne peut trouver pour elles que des valeurs empiriques, approchées. Ainsi, dans la formule [17c] on substitue aux espérances mathématiques des termes de la série les moyennes arithmétiques de leurs valeurs ; aux coefficients de corrélation à priori — les coefficients empiriques calculés d'après la formule [1] ; aux écarts-type à priori — les écarts empiriques calculés d'après la formule [2] [avec ou sans correction pour (N — 1)], etc.

On ne connaît non plus les véritables valeurs des grandeurs e. Une question se pose, à savoir si on peut déterminer le degré d'approximation réalisé par la formule [24], c.-à.-d. si on peut se rendre compte jusqu'à quelle mesure la liaison entre les écarts de $x^{(0)}$ et ceux des autres séries sur leurs espérances mathématiques est-elle obscurcie par l'influence du terme résiduel e. Ce dernier, comme on sait, représente par lui — même non seulement l'ensemble des influences accidentelles et non prises en considération, mais reflète également les inexactitudes admises lors de la recherche de la forme et des constantes de la formule elle-même [24].

Si on introduit les notations :

$$\bar{x}_i = b_{01}\, x_i^{(1)} + b_{02}\, x_i^{(2)} + b_{03}\, x_i^{(3)} + \ldots\ldots + b_{0n}\, x_i^{(n)} \qquad [34]$$

l'égalité [18] peut prendre la forme suivante :

$$x_i^{(0)} = \bar{x}_i + e_i;\ \text{ou}\ \bar{x}_i = x_i^{(0)} - e_i \qquad [35]$$

Il résulte de là qu'on peut répondre à la question posée, en premier lieu, à l'aide du coefficient ordinaire de la corrélation de $\bar{x}_i$ avec $x_i^{(0)}$. En passant à la symbolisation de la première partie du présent article (voir page 278 et sui-