Le coefficient r'_{12} s'appelle *le coefficient empirique de corrélation*, et b_{12}' et b'_{21} — *les coefficients empiriques de régression.*

L'étude des propriétés mathématiques du coefficient de corrélation r_{12}' nous amène aux déductions suivantes: sa valeur maximum est de $+1$; on n'y arrive que lorsqu'entre les termes correspondants des deux séries il existe un rapport de proportionnalité directe, c'est-à-dire lorsque $\frac{x_1}{x_1} = \frac{x_2}{y_2} = \frac{x_3}{y_3} = \ldots \frac{x_N}{y_N} = d,$

où d est supérieur à 0. Si les mêmes proportions restent en vigueur, mais donnent un d négatif, nous avons alors une proportionnalité inverse entre les termes des deux séries, et le coefficient de corrélation r_{12}' atteint son minimum -1. En général, plus la liaison des termes correspondants des deux séries est éloignée de la liaison „idéale" de la proportionnalité (directe ou inverse), plus le coefficient r_{12}' tend vers zéro, valeur qui est atteinte, enfin, quand

$$\sum_{i=1}^{N} x_i\, y_i = 0.$$

Cependant, comme nous allons voir plus bas, la valeur de $r_{12}' = 0$ ne désigne pas encore qu'entre x et y il n'y ait aucune dépendance.

La déduction exposée de la formule relative au coefficient de corrélation (particulièrement si cette déduction est compliquée par l'examen des séries et des colonnes du tableau dit des corrélations à intervalles de classes), cache en elle de grands risques, puisqu'elle crée facilement une idée exagérée sur l'importance de ce coefficient et sur les limites de son application, limites qui paraissent plus larges qu'elles ne le sont en réalité.

Même au cas où la loi de liaison (de dépendance) entre x_i et y_i peut être représentée avec succès sous forme d'équation de premier degré et n'est affectée que par des causes *accidentelles* insignifiantes, le coefficient de corrélation ne donne pas d'une façon générale une caractéristique complète de toutes les particularités de cette dépendance. Avec les autres 4 paramètres, il ne la donne que lorsque la loi d'interdépendance des deux séries représente la „surface normale de la corrélation", c'est-à-dire seulement dans un cas particulier *).

C'est ainsi que la mesure de la longueur du navire, bien qu'elle représente une caractéristique technique très importante, ne nous donne encore une idée exacte ni sur sa largeur, ni sur son tonnage, ni sur sa vitesse et ses autres qualités maritimes. Si cependant nous connaissons le type exact du navire (par exemple superdreadnought, cuirassé de ligne et autres), alors la même mesure de la longueur parle beaucoup plus au spécialiste.

*) Comparer avec mon ouvrage cité plus haut: „Korrelationsrechnung", page 104.

Si la dépendance de y et de x n'a pas un caractère linéaire, le coefficient de corrélation peut amener à des déductions tout à fait inexactes. De plus, on ne peut pas souligner d'une manière suffisante la circonstance que le coefficient de corrélation ne peut guère s'appliquer en dehors des limites de liaisons libres, stochastiques, c'est-à-dire en dehors des cas où il y a lieu de mesurer l'intensité ou le degré de liaison entre deux ou plusieurs séries. La question de l'„étroitesse" (c'est ainsi que nous appellerons l'intensité, la solidité — note du traducteur) de la liaison entre deux fonctions mathématiques de la même variable n'a aucun sens: la liaison entre x et $Sn\, x$ étant exactement aussi „étroite" que celle entre x et x^2, ou x et kx. Cette limitation ne ressort pas directement de la déduction exposée de la formule, mais sa certitude peut être aisément démontrée par quelques exemples.

Considérons deux séries:

$$1,\ 2,\ 3,\ 4,\ 5,\ 6\ \ldots .\ N \text{ et}$$
$$1k,\ 2k,\ 3k,\ 4k,\ 5k,\ 6k\ \ldots .\ N_k$$

Il existe proportionnalité entre les termes de ces deux séries, et le coefficient de corrélation prend, comme nous l'avons noté plus haut, la valeur $+1$, si k est positif, et -1, si k est négatif. Prenons maintenant la corrélation de la série des premiers degrés des nombres naturels avec celle de leurs carrés:

$$1,\ 2,\ 3,\ 4,\ 5.\ \ldots\ldots\ N$$
$$1^2,\ 2^2,\ 3^2,\ 4^2,\ 5^2\ \ldots\ldots\ N^2.$$

Il en résulte que leur coefficient de corrélation change de grandeur en dépendance du dernier nombre N, par lequel la série se termine. Dans la limite, lorsque N tend vers l'infini, ce coefficient est égal à $+0{,}968$.

Pour le coefficient de corrélation des séries $1, 2, 3, 4, 5, 6 \ldots . N$ et $1^3, 2^3, 3^3, 4^3, 5^3, 6^3 \ldots . N^3$ on obtient dans les mêmes conditions: $r_{12}' = \pm 0{,}9512$ *).

Mais si dans les cas, où nous avons affaire avec une dépendance fonctionnelle, parfaite par sa nature, de x et de y, le coefficient de corrélation suscite chez nous une impression de „liaison pas tout à fait complète", il donne tout de même une grandeur positive assez considérable. Cependant, on peut obtenir aussi un résultat beaucoup plus mauvais. Ainsi, par exemple, nous calculous $r_{12} = 0$ pour la corrélation de x avec $Sn\, x$, quand x représente une progression arithmétique $\alpha, \alpha + h,\ \alpha + 2h \ldots .\ \alpha + Nh$, et, par conséquent, la seconde série est $Sn\, \alpha,\ Sn\, (\alpha + h),\ Sn\, (\alpha + 2h), \ldots\ Sn\, (\alpha + N\, h)$, à condition que la dernière série renferme un nombre entier de „vagues" complètes, ou bien, pourvu que N tende vers l'infini.

De même, le coefficient de corrélation est nul pour les séries des „erreurs" et de leurs probabilités dans chaque loi symétrique des erreurs, par exemple celle de *Gauss*, bien que la proba-

*) Voir Korrelationsrechnung, page 39.