JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

ALFRED MILHAUD

Une nouvelle méthode pour étudier les relations entre aptitudes

Journal de la société statistique de Paris, tome 98 (1957), p. 273-293
<http://www.numdam.org/item?id=JSFS_1957__98__273_0>

VIII

VARIETÉ

Une nouvelle méthode pour étudier les relations entre aptitudes

INTRODUCTION

Il nous est apparu que le schéma sur lequel repose l'analyse factorielle contenait une hypothèse qui méritait d'être analysée.

En effet, on assimile l'aptitude à un vecteur qui est la somme de plusieurs composantes affectées chacune d'un coefficient.

On admet aussi que ces composantes sont des réalités absolues, c'est-à-dire des fonctions ayant une existence objective, antérieure par conséquent à la connaissance que nous avons de cette existence.

Cela correspond à un schéma qu'il est commode d'essayer mais dont rien ne garantit l'exactitude en toute circonstance. Il faudrait, pour que ce schéma soit toujours valable, admettre l'existence absolue de l'aptitude, antérieure à l'idée que nous en avons. C'est l'attitude « réaliste » au sens qu'avait ce mot en philosophie scholastique, qui voyait dans les « universaux » des réalités premières. Nous préférons éviter cette hypothèse en considérant une aptitude plus ou moins élevée comme un jeu plus ou moins favorable de facteurs indéterminés.

Il ne nous échappe pas que l'analyse factorielle peut être reconnue utilisable sans adopter l'attitude « réaliste » que nous avons évoquée.

Si l'aptitude Y_1 n'est pas fonction linéaire de la composante x, elle est nécessairement fonction linéaire d'une fonction $f(x)$; et de même, l'aptitude Y_2 est fonction linéaire d'une fonction $g(x)$. Si $f(x)$ et $g(x)$ sont peu différentes, on peut admettre l'existence d'une composante dont Y_1 et Y_2 sont également fonctions linéaires.

Mais on ne saurait manquer de remarquer qu'en pareil cas la composante isolée risque de n'être qu'un « fantôme mathématique ». Quelle est cette composante U, dont Y est fonction linéaire mais qui est une somme de fonctions non monotones de variables correspondant à des réalités plus profondes, par exemple fonctions linéaires de l'envers du poids, du logarithme de capacité vitale et du carré de la taille?

Au surplus, une telle situation n'est réalisée que si Y_1 et Y_2 sont fonctions linéaires l'une de l'autre; et, en l'absence de toute hypothèse métaphysique, rien n'oblige à admettre *a priori* qu'il en sera toujours ainsi. On l'admet toutefois volontiers dans un but de simplifications.

Le problème pratique est alors de savoir si les trois hypothèses simplificatrices :

1° Y_1 et Y_2, fonctions linéaires l'une de l'autre;
2° σ_{Y_1} indépendant de Y_2;
3° σ_{Y_2} indépendant de Y_1 sont sans inconvénient pratique.

Le psychologue ne peut pas l'admettre. Dans un groupe quelconque, la liaison peut être différente entre les moins doués de ce qu'elle est entre les plus doués. Une fonction de régression peut être une courbe à allure convexe ou concave. On peut imaginer diverses formes de courbes de régression. De toutes façons en admettant *a priori* la forme linéaire on se prive de rechercher des informations du plus haut intérêt.

Au surplus la forme de la courbe de régression n'est pas le seul renseignement intéressant.

Pour représenter toutes les situations, nous avons adopté le schéma suivant :

Supposons un nombre indéfini n de variables, et des aptitudes dont chacune dépend de quelques-unes de ces variables.

Chacune de ces n variables est prise comme coordonnée dans un espace à n dimensions.

Chaque individu est représenté par un point dans cet espace et ses aptitudes ont les valeurs qui correspondent aux coordonnées de ce point.

C'est dans ce schéma que nous allons essayer de comprendre les relations probables qui existent entre aptitudes.

Soient 2 aptitudes α et β dépendant de mêmes variables, et par conséquent ayant leurs optimums susceptibles d'être représentés par 2 points P et Q dans le même espace. Nous allons voir des liaisons diverses selon la population étudiée.

Construisons la droite — support de $\overline{PQ}$.

Si, sur cette droite, on se déplace de $-\infty$ vers P, on voit s'élever à la fois l'aptitude α et l'aptitude β jusqu'à ce qu'on atteigne le point P. Ensuite, lorsqu'on dépasse P et que l'on se dirige vers Q, α diminue tandis que β augmente. Une fois que l'on dépasse Q, α et β diminuent ensemble.

Soit une population représentée par un nuage *allongé* avec $\overline{PQ}$ sur un grand axe; α et β ne sont pas fonctions linéaires l'une de l'autre puisque quand α augmente, β peut augmenter ou diminuer.

Supposons par contre, que l'on se déplace le long d'une droite perpendiculaire à $\overline{PQ}$, α et β varieront toujours dans le même sens.

Soit une population représentée par un nuage *allongé* avec un grand axe *perpendiculaire* à $\overline{PQ}$; α et β y sont fonctions monotones l'une de l'autre puisque variant toujours dans le même sens. Mais la fonction n'est pas fatale-

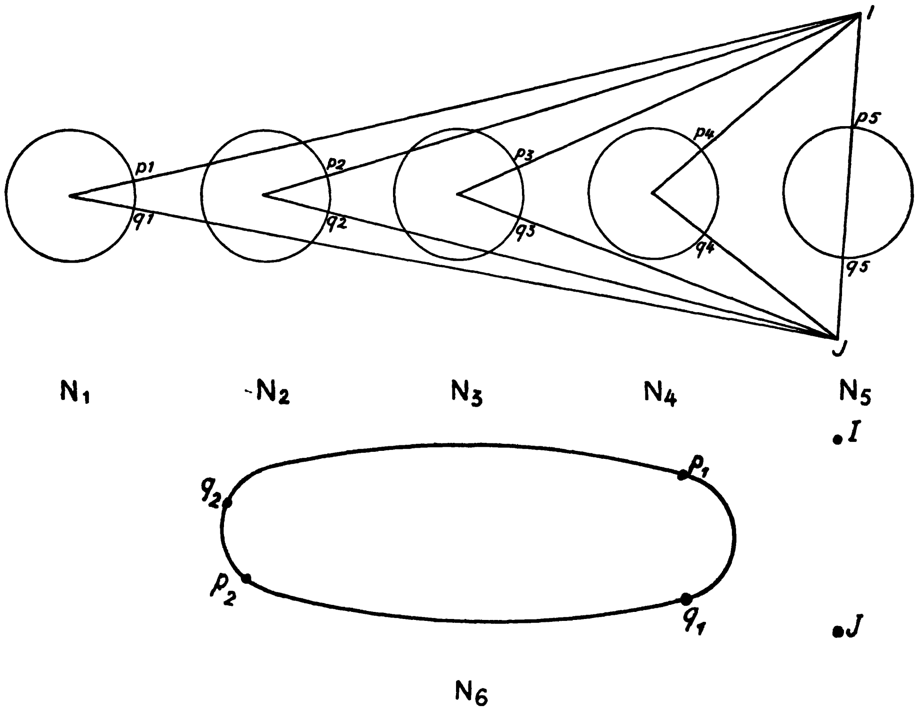

Fig. 1. — En p les conjonctures optimales des aptitudes et en p' celles des inaptitudes, au sens d'une population assez hétérogène.

ment linéaire car la fonction linéaire n'est qu'un cas particulier de la fonction monotone.

On peut multiplier les hypothèses en supposant des trajets quelconques et des formes quelconques de nuage.

Il existe donc une infinité d'expressions possible sde la fonction de régression, suivant les populations étudiées, pour une même valeur de $\overline{PQ}$.

Les relations probables entre aptitudes peuvent encore différer d'une autre façon d'une population à l'autre.

Soit I et J, les optimums possibles de deux aptitudes, tels que des génies, non rencontrés dans les populations examinées, les incarneraient.

Dans la population représentée par le nuage N_1 les optimums sont en p_1 et q_1 (voir fig. 1). En N_2 ce sont p_2 et q_2, relativement plus distants l'un de l'autre que p_1 et q_1. En N_3 ce seront p_3 et q_3; en N_4 ce seront p_4 et q_4 en N_5 ce seront p_5 et q_5, d'autant plus distants l'un de l'autre que l'on sera plus près de I et de J.

On remarquera que non seulement la liaison devient moins étroite au fur et à mesure que l'on considère des milieux plus doués aux deux points de vue considérés, et que cette liaison peut s'inverser, comme c'est le cas en N_5.

Effectivement, il n'est pas de spécialités, si voisines soient-elles, dont les aptitudes ne soient opposées à un certain échelon, très élevé à l'intérieur d'une population elle-même d'un très haut niveau en ce qui concerne ces aptitudes.

Si par exemple on arrive à considérer des professeurs de Facultés des Lettres, les uns titulaires de la chaire de latin, les autres titulaires de celle de grec, on a une population où les 2 aptitudes s'opposent.

***

Nous pouvons examiner pour chaque nuage les propositions suivantes :

A. — Les individus d'un sous-ensemble (S*e*) présentent pour l'aptitude α une valeur moyenne, supérieure à la valeur moyenne de l'ensemble entier (E*e*) du nuage.

B. — Les individus de ce même sous-ensemble (S*e*) présentent pour l'aptitude β une valeur moyenne, supérieure à la valeur moyenne de l'ensemble entier (E*e*) du nuage.

On vérifie *n* sous-ensembles.

Nous avons les énoncés possibles suivants :

A vrai, A faux, B vrai, B faux, et les compositions suivantes :

1° — A vrai, B vrai;
2° — A faux, B vrai;
3° — A vrai, B faux;
4° — A faux, B faux.

Il y a *conjonction* si les 4 compositions sont possibles. Alors les aptitudes α et β sont indépendantes.

Il y a *homologie* si 2 compositions seulement sont possibles, la 1re et la 4e.

Si A est vrai, B est vrai, et réciproquement.

Si A est faux, B est faux, et réciproquement.

Il y a *contradiction* si 2 compositions seulement sont possibles, la 2e et la 3e.

Si A est faux, B est vrai, et réciproquement.

Si A est vrai, B est faux, et réciproquement.

Mais on ne peut imaginer le cas où il y a *contrariété.* Trois compositions sont possibles : la 2e, la 3e et la 4e :

A faux — B vrai; A vrai — B faux; A faux — B faux.

Une seule composition est impossible dans ce cas, la *première* : A et B ne peuvent pas être également vrais.

Il y aurait donc alors incomptabilité des supériorités seulement.

Dans d'autres cas, la seule composition impossible à rencontrer serait la *deuxième.* On pourrait avoir A vrai B vrai; A vrai B faux, A faux B faux. La seule impossibilité serait : A faux B vrai. Il y aurait donc *implication.*

Il pourrait arriver aussi que la seule composition impossible soit la *troisième :* A vrai B faux. Il y aurait encore *implication.*

Enfin il se pourrait que la seule composition impossible soit la *quatrième.* On ne pourrait pas avoir à la fois A et B faux. On serait alors en présence d'une relation de *disjonction.*

Soient les résultats illustrant les aptitudes α et β dans les épreuves A et B.

On a un moyen, non de mesurer, mais d'échelonner $\overline{PQ}$ qui représente la distance entre les optimums observés pour ces épreuves. Ce moyen est de déterminer le rang que le sujet, classé premier dans chaque épreuve, occupe dans l'autre.

Soit R*g* *ab* le rang en *b* du sujet classé premier en *a.*

Soit R*g* *ba* le rang en *a* du sujet classé premier en *b.*

R*g* *ab* est un nombre probablement d'autant plus élevé que $\overline{PQ}$ est plus grand et par conséquent qui peut servir à l'échelonner. Il en est de même de R*g* *ba.*

Soient *n* lots de 100 sujets.

Déterminons le rang, en épreuve *b*, du sujet classé premier en épreuve *a*, dans chaque lot; puis calculons la moyenne de ces rangs. Nous avons la valeur de V*ab.*

De la même façon, nous pouvons déterminer la valeur de V*ba.*

V*ab* = 0,01 signifie liaison absolue, le 1er sujet en *a* étant le 1er en *b.*

V*ab* = 0,50 signifie absence de liaison, la moyenne des sujets premiers en *a* ayant le rang 50e sur 100 sujets, en *b.*

V*ab* = 1 signifie la liaison inverse, le 1er en *a* ayant un rang du 100e sur 100 sujets en *b.*

La représentation géométrique est la suivante :

Dans le 1er cas, P et Q sont confondus.

Dans le 2e cas, ils sont à telle distance que la moitié de l'espace considéré est formé par des secteurs où l'on ne puisse se rapprocher de l'un de ces points sans s'éloigner de l'autre; et que l'autre moitié est formée par des secteurs où l'on s'approche nécessairement de P en s'approchant de Q et réciproquement.

Dans le 3e cas, P et Q sont aux deux extrémités d'un espace.

Si V*ab* et V*ba* ne sont pas égaux, il est commode de caractériser la distance $\overline{PQ}$ par un nombre unique. On prendra alors la moyenne géométrique.

$$\overline{V} = \sqrt{Vab \cdot Vba}$$

On verra plus loin qu'il est impossible cependant de négliger de tenir compte des valeurs respectives de V*ab* et V*ba*, leur inégalité apportant des renseignements précieux.

***

Le recours aux moyennes géométriques offre d'autres commodités.

Soit à caractériser la liaison d'une aptitude α illustrée par une épreuve a, avec des aptitudes β, γ, δ etc... illustrées par des épreuves b, c, d, etc...

Soit *n* le nombre des aptitudes considérées, on pourra prendre

$$\overline{V}_{a\,(t)} = \sqrt[n-1]{Vab \cdot Vac \cdot Vad \cdot \text{etc...}}$$

Cependant cette formule tient incomplètement compte des liaisons observées, si on a :

$$Vab \neq Vba, \quad Vac \neq Vca, \quad Vad \neq Vda$$

On préférera alors :

$$\overline{V}_{a\,(t)} = \sqrt[2(n-1)]{Vab \cdot Vba \cdot Vac \cdot Vca \cdot Vad \cdot Vda \cdot \text{etc...}}$$

COEFFICIENTS V. ET W

Le raisonnement que nous avons fait en considérant le rang en *b* du sujet classé premier en *a* et réciproquement, pourrait être repris en considérant les sujets classés derniers dans les mêmes épreuves.

Si nous appelons premier le sujet classé dernier dans les mêmes épreuves, second celui classé avant-dernier, etc... nous pouvons considérer des coefficients comparables aux coefficients V*ab* et V*ba*.

Leur usage sera analogue.

Nous pourrons caractériser la liaison des inaptitudes par un nombre compris entre 0,01 et 1

w = 0,01 signifie que le dernier en α est dernier en β et réciproquement.
w = 0,5 signifie que les inaptitudes sont indépendantes
w = 1 signifie que le dernier en α est premier en β et réciproquement.

On verra plus loin que sauf dans le cas où $\overline{W} = 1$, auquel cas $\overline{V} = 1$, la liaison des inaptitudes peut différer de celle des aptitudes. Notre procédé permet d'étudier l'une et l'autre de deux liaisons.

Nous décrirons :

$$\overline{W} = \sqrt{Wab \cdot Wba}$$

$$\overline{W}_{a\,(t)} = \sqrt[n-1]{Wab \cdot Wac \cdot Wad \cdot \text{etc...}}$$

$$\overline{W}_{a\,(t)} = \sqrt[2n-1]{Wab \cdot Wba \cdot Wac \cdot Wca \cdot Wad \cdot Wda, \text{etc...}}$$

L'énoncé d'un nombre exprimant la liaison entre deux aptitudes suppose qu'il y a nécessairement, dans le genre d'expérience que nous imaginons sur deux sous-ensembles (S*e*), constatation d'une relation soit d'homologie, soit de contradiction, soit de conjonction.

Quand les sous-ensembles sont composés chacun d'un petit nombre d'individus, il y a conjonction. Si le nombre de sujets s'élève dans chaque sous-ensemble, on peut, soit voir subsister la conjonction, soit voir apparaître l'homologie ou la contradiction. La situation est alors simple.

Mais il y a d'autres cas possibles. Par exemple, il y a plus de probabilité qu'un bon joueur de tennis soit bon coureur qu'un bon coureur un bon joueur de tennis. Cela signifie que si α est l'aptitude au tennis d'un sous-ensemble assez grand et β celle à la course, alors lorsque A est vrai B est vrai, tandis que la réciproque n'est pas exacte. La relation observée est une *implication.*

Notre schéma permet de représenter toutes les relations possibles y compris l'implication, la contrariété et la disjonction, par l'inégalité de V *ab* et de V *ba* ou par celle de W*ab* et W*ba*.

RELATIONS LOGISTIQUES ENTRE APTITUDES ET ENTRE INAPTITUDES

Supposons que l'aptitude α dépende des mêmes variables que l'aptitude β et de variables supplémentaires.

Si l'optimum de α est un point P, celui de β ne peut pas être représenté par un point. Ce sera une portion d'espace L. Celle-ci sera une droite si β dépend d'une variable de moins que α. Ce sera un plan si la différence est de deux variables. Ce sera un hyperplan à 3 dimensions, c'est-à-dire un espace à 3 dimensions dans un espace à plus de 3, si la différence est de 3 variables; etc...

Or, en tout point P d'un espace à 3 dimensions on peut faire passer un plan à 2 dimensions parallèle à un plan donné, et un seul. En tout point P d'un espace à 4 dimensions, on peut faire passer un hyperplan à 3 dimensions, parallèle à un hyperplan donné et un seul. D'une façon générale dans un espace à n dimensions, on peut toujours faire passer, en un point P, un hyperplan à $(n - 1)$ dimensions, parallèle à un hyperplan donné, et un seul.

Ainsi, quel que soit le nombre des dimensions de l'espace, correspondant chacune à l'une des variables dont dépendent les aptitudes, en P nous pouvons toujours faire passer une portion d'espace L′ parallèle à la portion d'espace L.

Cet artifice nous permet d'étudier les relations logistiques possibles entre aptitudes ou entre inaptitudes.

Plaçons nous au point O et envisageons diverses hypothèses :

1° P est en L (fig. 2).

Il est alors évident que chaque fois que l'on s'approche de P on s'approche de L.

La réciproque n'est pas toujours vraie.

On peut s'approcher de L en s'approchant de P, par exemple en suivant les trajets compris entre $\overline{OP}$ et $\overline{OR}$, si $\overline{OR} \perp$ L.

Mais on peut aussi suivre d'autres trajets tels que $\overline{OS}$ ou $\overline{OT}$. Sur $\overline{OS}$, en se rapprochant toujours de L, on s'approche de P *seulement* jusqu'au moment où on atteint H, ($\overline{PH} \perp \overline{OS}$). De même, sur $\overline{OT}$, en se rapprochant toujours de L, on s'approche de P *seulement* jusqu'au moment où on atteint H′ ($\overline{PH'} \perp \overline{OT}$).

Donc la probabilité de s'approcher de P quand on s'approche de L est moins forte que celle de s'approcher de L quand on s'approche de P.

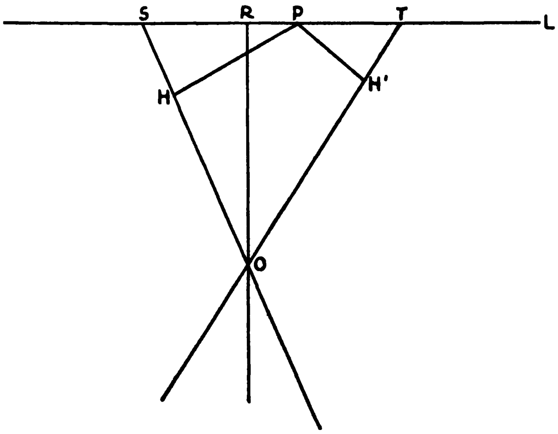

Fig. 2.

Les sujets premiers en épreuve *a* (optimum en P) sont sûrement très bons en *b* V*ab* sera représenté par un nombre très faible; tandis que les premiers en

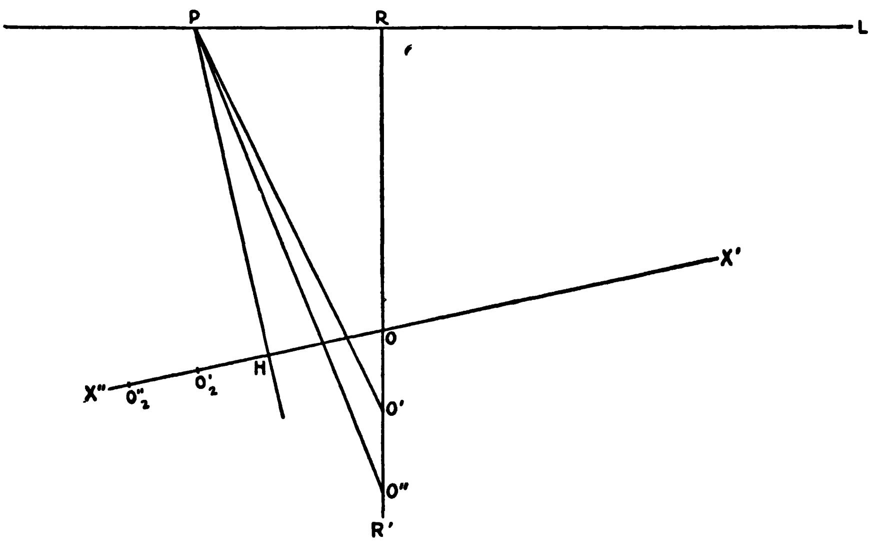

Fig. 2 *bis*.

b ne seront pas toujours très bons en *a*, et V*ba* sera un nombre moins faible : V*ba* > V*ab*.

Dans la figure 2 *bis* qui correspond au même cas que la figure 2, nous pouvons nous déplacer sur $\overline{OR'}$ prolongement de $\overline{RO}$.

Déplaçons-nous de O′ à O″ sur $\overline{OR'}$. Nous nous éloignons, par rapport à L, de la distance $\overline{O'O''}$.

Nous nous éloignons, par rapport à P, de la distance $(\overline{PO''} - \overline{PO'})$.

Or $\overline{(PO''^2} = (\overline{PO'})^2 + (\overline{O'O''})^2 - 2\,\overline{PO'} \cdot \overline{PO''} \cdot \cos \widehat{PO'R'}$.

Mais $\widehat{PO'R'}$ est sûrement obtus puisque $\widehat{PO'R}$ est nécessairement aigu.

Donc

$$(\overline{PO''})^2 > (\overline{P\,O'})^2 + (O'O'')^2$$

$$(\overline{PO''})^2 - (\overline{PO'})^2 > (\overline{O'O''})^2$$

Or

$$(\overline{PO''} - \overline{PO'})^2 > (\overline{PO''})^2 - \overline{PO''})^2$$

Donc *a fortiori*

$$(\overline{PO''} - \overline{PO'})^2 > (\overline{O'O''})^2$$

$$(\overline{PO''} - \overline{PO'}) > \overline{O'O''}$$

Nous nous éloignons, par rapport à P, plus que par rapport à L.

Mais au fur et à mesure que nous nous déplaçons sur une droite $\overline{OX'}$ l'éloignement par rapport à P sera supérieur à celui par rapport à L.

Le cas est différent si nous nous déplaçons sur $\overline{OX''}$.

Si nous nous déplaçons le long de $\overline{XX''}$, tandis que nous nous éloignons de L, nous nous rapprochons de P jusqu'à ce que nous ayons atteint H, puis nous nous éloignons de P en même temps que de L et $\widehat{PO_2'O_2''}$, étant nécessairement obtus, nous nous retrouvons dans le cas précédent où nous nous éloignons plus par rapport à P que par rapport à L.

Ainsi il y a *certains* cas où nous nous approchons de P en nous éloignant de L, mais une *infinité* d'autres où nous nous éloignons de P en nous éloignant de L.

De même quand on s'éloigne de L, on s'éloigne généralement de P. Les exceptions correspondent aux mêmes trajets.

De toutes façons le déplacement par rapport à P est supérieur, *en valeur absolue*, au déplacement par rapport à L.

Dans ces conditions, la moyenne des déplacements possibles quand on veut s'éloigner de P ou de L est un éloignement par rapport à la fois à P et à L, mais *plus* par rapport à P que par rapport à L.

Éloignons-nous donc de P, c'est-à-dire considérons des sujets de plus en plus près des derniers en épreuve α. Nous nous éloignerons de L, mais moins que de P; les sujets seront de plus en plus mauvais en b mais moins mauvais qu'en a.

Par contre éloignons-nous de L, c'est-à-dire considérons des sujets de plus en plus mauvais en épreuve b; nous nous approchons très vite de sujets qui sont en même temps les derniers en a.

Cela nous permet de poser $Wab > Wba$ d'où $Wba < Wab$.

Ainsi quand une aptitude α et une aptitude β sont dans des relations telles que β dépende d'un certain nombre de variables, tandis que α dépend de ces mêmes variables et d'autres en plus, on a :

1° En ce qui concerne les *aptitudes*, l'inégalité

$$Vba > Vab \quad \text{d'où} \quad Vab < Vba$$

2° En ce qui concerne les *inaptitudes*, l'inégalité

$$Wba < Wab \quad \text{d'où} \quad Wab > Wba$$

Considérons maintenant P sur une droite, un plan, ou un hyperplan L' parallèle à L, et plaçons-nous en dehors de L (fig. 3), la situation est la même.

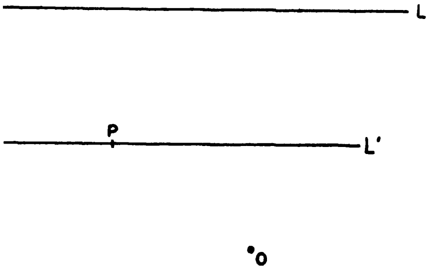

Fig. 3.

Si on se place entre L et L' (fig. 3 *bis*), quand on s'approche de L, on s'éloigne de L', et quand on s'approche de L' ou s'éloigne de L.

La probabilité de s'approcher de L égale celle de s'éloigner de L'.

Traçons XY // L et L', et passant en O, et lisons les lettres qui signalent un angle, dans le sens habituel qui est le sens des aiguilles d'une montre.

Tout trajet en $\widehat{YOX}$ nous éloigne de L' et nous approche de L.

La plupart des trajets nous éloignent de P. Cependant certains d'entre eux, dans une partie de leur parcours, nous rapprochent de P. Par exemple, soit le trajet $\overline{OT}$; menons $\overline{PH} \perp \overline{OT}$, nous constatons que de O en H, nous nous appro-

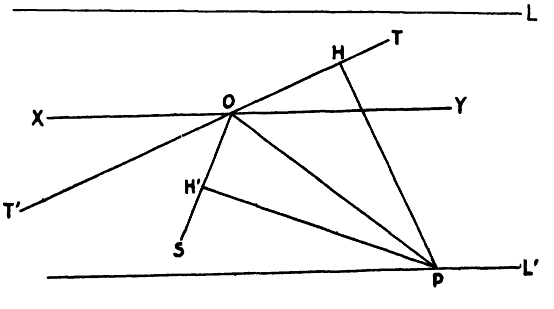

Fig. 3 *bis*.

chons de L et de P à la fois. Par contre, il n'est pas possible dans ce secteur de s'éloigner de P sans se rapprocher de L.

Ainsi en $\widehat{YOX}$, quand on s'approche de L on s'éloigne en général de P et réciproquement, donc

$$Vab > 0{,}5, \quad Vba > 0{,}5, \quad Wab > 0{,}5 \quad \text{et} \quad Wba > 0{,}5,$$

mais on ne s'éloigne pas de P d'une façon constante quand on s'approche de L. Les bons en *b* sont en général mauvais en *a*, mais il y a beaucoup d'exceptions, tandis que les bons en *a* sont plus constamment mauvais en *b*, d'où

$$Vba < Vab \quad \text{et} \quad Wba > Wab$$

En $\widehat{XOY}$, tous les déplacements possibles nous éloignent de L et nous approchent de L', mais tous les déplacements ne nous approchent pas de P. En $\overline{OP}$, nous nous approchons de P sur toute la longueur du trajet. Mais en $\overline{OS}$, nous nous approchons de P jusqu'à H' et nous nous éloignons ensuite. En $\overline{OT'}$ prolongement de $\overline{OT}$, nous nous éloignons de L sans jamais nous approcher de P. Les trajets plus éloignés nous en éloignent aussi.

On a donc la même situation que dans le secteur $\widehat{YOX}$, d'où

$$Vba < Vab \quad \text{et} \quad Wba > Wab$$

En résumé, nous concluons :

« Lorsqu'une aptitude α dépend de variables en plus de variables communes à elle et à une aptitude β, deux cas peuvent se présenter :

1° Vab, Vba, Wab, $Wba < 0,5$ on a
$Vba > Vab$ et $Wba < Wab$
2° Si Vab, Vba, Wab $Wba > 0,5$ on a
$Vba < Vab$ et $Wba > Wab$

De toutes façons, l'inégalité des W est de sens opposé à celle des V.

En logistique, si nous opposons les aptitudes α et β aux inaptitudes $\bar{\alpha}$ et $\bar{\beta}$, c'est une notion classique que $(\alpha \longrightarrow \beta) \rightleftarrows (\bar{\beta} \longrightarrow \bar{\alpha})$.

Il est intéressant cependant d'en donner une démonstration. Si nous disons que l'aptitude α dépend des mêmes variables que l'aptitude β et d'autres variables en plus, (par exemple l'aptitude au tennis dépend des variables déterminant l'aptitude à la course et d'autres variables en plus) nous pouvons appeler B l'ensemble de ceux qui réussissent en b (par exemple, les bons cou-

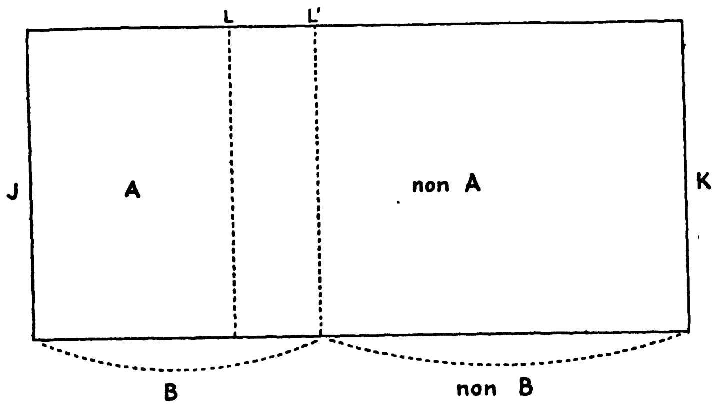

Fig. 4.

reurs) et appeler A l'ensemble de ceux qui excellent en α (par exemple les champions de tennis).

Nous dirons que l'ensemble B contient l'ensemble A et une partie de non-A.

L'ensemble des non-B est un ensemble des sujets à la fois non-B et non-A.

Enfin l'ensemble des sujets non-A est plus large que celui des non-B donc, il le contient (fig. 4).

Ainsi l'inégalité des V et celle des W seront de signes opposés.

Du point de vue logistique, on observera ce qui suit :

1° Il y a *conjonction*

$$\text{Si } (Vab = Vba = Wab = Wba) = 0{,}5$$

Les optimums sont 2 points dans le même espace, et distants l'un de l'autre d'une longueur égale à la moyenne des trajets possibles.

2° Il y a *homologie*

$$\text{Si } (Vab = Vba \text{ et } Wab = Wba) < 0{,}5$$

Les optimums sont 2 points rapprochés l'un de l'autre.

3° Il y a *contradiction*

$$\text{Si } (Vab = Vba \text{ et } Wab = Wba) > 0{,}5$$

Les optimums sont 2 points éloignés l'un de l'autre.

4° Il y a *implication* si les 2 aptitudes ont des variables communes, et que l'une, qui implique l'autre, a des variables supplémentaires.

Si l'aptitude α implique l'aptitude β, on a :

$$Vab < Vba, \text{ si les deux sont } < 0{,}5$$
$$Vab > Vba, \text{ si les deux sont } > 0{,}5.$$

L'inégalité des W sera de sens opposé à celle des V.

La *contrariété* signifie :

$$(Vab = Vba) > 0{,}5 \quad (Wab = Wba) > 0{,}5$$

Les deux aptitudes n'ont qu'une variable commune, toutes les autres étant différentes. Les deux optimums sont dans deux espaces n'ayant qu'une dimension commune, sur laquelle leurs projections sont *éloignées.*

La *disjonction* correspond au cas où

$$(Vab = Vba) < 0{,}5 \quad \text{et} \quad (Wab = Wba) > 0{,}5$$

Les deux aptitudes n'ont qu'une variable commune, toutes les autres étant différentes. Les deux optimums sont dans deux espaces n'ayant qu'une dimension commune, mais sur laquelle leurs projections sont *voisines.*

Il faut pour appliquer ce schéma connaître les approximations avec lesquelles dans nos recensements nous exprimons les V et les W, pour savoir si les inégalités trouvées :

$$\begin{array}{ll} V & \gtrless 0{,}5 \\ W & \gtrless 0{,}5 \\ Vab & \gtrless Vba \\ Wab & \gtrless Wba \end{array}$$

peuvent être retenues pour leur appliquer les conclusions conformes à la théorie qui vient d'être exposée.

Il résulte de notre schéma (voir fig. 1) qu'en cas d'homologie ou implication, on a $\overline{V} \geqslant \overline{W}$, et en cas de contradiction $\overline{V} \leqslant \overline{W}$.

Nous avons toujours envisagé le cas d'efficiences variant en fonction l'une de l'autre de façon, non pas linéaire comme dans le schéma des corrélations, mais toutefois monotone.

En disant que si $Vab < Vba$, $Wab < Wba$ nous avons donné non seulement une démonstration géométrique qui, évidemment, ne s'applique plus en cas de fonction non monotone, mais aussi une démonstration logistique.

Mais il faut reprendre celle-ci.

Nous avons dit : Si le groupe B des bons en *b* contient le groupe A des bons en *a*, le groupe « non-A » des mauvais en *a* contient celui « non-B » des mauvais.

Nous l'avons démontré par un schéma.

Ce schéma contient une hypothèse implicite. Nous avons admis qu'il existait un ensemble tel qu'en se déplaçant de l'une de ses extrémités « J » vers l'autre extrémité « K », on trouvait des sujets de plus en plus détériorés, à la fois en *a* et en *b*. C'était admettre que *a* et *b* étaient fonctions monotones.

Ceci posé, il était possible de poser une limite « L » entre le groupe A et le groupe non-A, et une limite L′ entre le groupe B et le groupe non-B, et le premier coup d'œil nous montrait que si B contient A, non-A contient non-B (voir fig. IV).

Ces limites ne signifient plus rien si *a* et *b* ne se détériorent pas toujours quand on va de J à K.

Examinons ce qui se passe si *a* et *b* varient l'un en fonction de l'autre d'une façon discontinue.

IMPLICATION DES APTITUDES (fig. 5)

On a une variable x avec une origine éloignée de la zone optimale de l'échantillon étudié, et orientée de telle façon que les efficiences *a* et *b* s'améliorent quand x augmente.

a s'améliore quand x augmente, et cela dans toute la zone considérée.

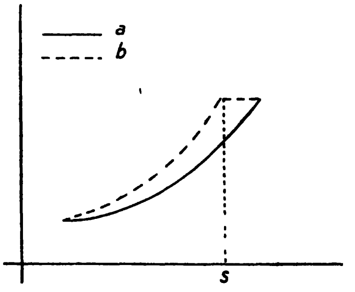

Fig. 5.

b ne s'améliore quand x augmente que jusqu'à ce que x ait atteint un seuil s. Les possibilités représentées par $x > s$ ne sont pas utilisées dans l'efficience *b*.

Tous les sujets qui représentent des valeurs de x telles que $x > s$, réalisent en ce qui concerne x la conjoncture optimale pour *b*, tandis qu'ils ne la réalisent pas tous pour *a*.

Il y a là tout un groupe de sujets qui seront généralement bons en *a* et en *b*, mais surtout en *b*, ce qui entraîne $Vab < Vba$.

Plus x est faible, plus les inaptitudes sont voisines $Wab = Wba$.

IMPLICATION DES INAPTITUDES (fig. 6)

On a une variable x avec une origine o, éloignée de la zone optimale de l'échantillon étudié. L'efficience a s'améliore quand x augmente et cela dans

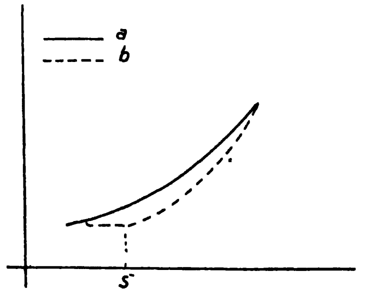

Fig. 6.

toute la zone considérée. L'efficience b est très basse tant que x n'a pas atteint un certain seuil s.

Si x est faible, l'inaptitude b est plus marquée que l'inaptitude a, ce qui entraîne Wab $<$ Wba.

Plus x est élevé, plus les deux aptitudes sont voisines, Vab $=$ Vba.

DOUBLE IMPLICATION (fig. 7)

Les deux efficiences s'améliorent quand x augmente.

Au début, c'est surtout a qui s'améliore, b restant très mauvais jusqu'en S_1. Avec x faible, on a donc des sujets mauvais en a et pires en b, ce qui entraîne

$$\mathrm{W}ab < \mathrm{W}ba$$

Quand x atteint une certaine valeur S_1, son augmentation continue à améliorer a, mais augmente aussi b, et *plus rapidement* que a.

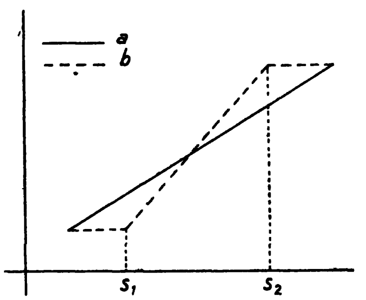

Fig. 7.

Au-delà d'une valeur encore plus forte S_2, les augmentations sont à nouveau sans influence sur b alors que a s'améliore encore.

Tous les sujets qui présentent des valeurs de x telles que $x > S_2$ réalisent en ce qui concerne x la conjoncture optimale pour b, alors qu'ils ne la réalisent pas tous pour a, donc les sujets bons en a sont encore meilleurs en b, ce qui entraîne $Vab < Vba$.

On a donc alors à la fois

$$Vab < Vba$$
$$Wab < Wba.$$

VALEUR SIGNIFICATIVE DES INÉGALITÉS

Supposons une population où nous pouvons former un nombre très grand de lots de 100 sujets.

Supposons qu'en déterminant Vab et Vba sur ce grand nombre de lots, on trouve $Vab = Vba = \overline{X}$.

Si nous formons des groupes de 10 lots, nous trouverons des valeurs diverses de Vab, avec pour mode la valeur de $\overline{X}$. Nous trouverons aussi des valeurs diverses de Vba, avec pour mode la valeur $\overline{X}$. Quand sur un lot de 10 groupes de 100 sujets on trouve $Vab \neq Vba$, il est possible d'admettre cette inégalité dans le cas seulement où nous sommes sûrs qu'elle n'est pas due aux erreurs d'échantillonnage.

Malheureusement, nous ignorons tout de la distribution des Vab ou des Vba autour de $\overline{X}$.

Nous devons nous contenter de déterminer quelles sont les inégalités qui sont s gnificatives, quel que soit le mode de distribution.

Nous ignorons la forme de la courbe de fréquence des rangs en b des sujets premiers en a.

Mais nous pouvons nous placer dans une hypothèse invraisemblablement défavorable, à savoir que cette courbe est rectangulaire et que tous les rangs sont également probables.

Avec des séries de 100, la somme des carrés des écarts est alors

$$2\Sigma(1^2 + 2^2 + \ldots\ldots 50^2) = \frac{2\,(50+1)\cdot(100+1)\cdot 50}{6} = 85850$$

La moyenne des carrés des écarts devient $\frac{85\,850}{100} = 858{,}50$

La moyenne des carrés des centièmes d'écart, puisque c'est cela que nous considérons devient 0,08585

Le σ de chaque série devient $\sqrt{0{,}08585} = 0{,}29$.

Sur 10 séries le σ est de $\frac{0{,}29}{\sqrt{10}}$ et le σ de la moyenne des séries, est $\frac{0{,}29}{\sqrt{10}\sqrt{10}} = \frac{0{,}29}{10} = 0{,}029$.

Nous recourons alors au raisonnement suivant :

L'écart observé entre Vab et Vba est-il fortuit?

Cela revient à demander la probabilité d'observer cet écart en le contrôlant sur un ensemble parent plus vaste où $Vab = Vba$.

Cette probabilité est maxima dans un cas bien connu, à savoir lorsque dans cet ensemble parent, on a Vab = Vba = moyenne des valeurs observées.

Quelle que soit la forme de la courbe de fréquence, en vertu du théorème de Bienaymé et Tchebicheff, un écart de 2 σ présente une probabilité $\varpi < \frac{1}{4}$.

Deux écarts de 2 σ ont une probabilité $\varpi < \left(\frac{1}{4}\right)^2$, c'est-à-dire $\varpi < \frac{1}{16}$.

Un écart de 4 σ suppose 2 écarts de 2 σ de sens opposés, ce qui a une probabilité $\varpi < \frac{1}{32}$.

Ainsi, un écart de (0.029).4 = 0,116, (so t en chiffres ronds 0,12) est sûrement significatif.

Il l'est en effet en adoptant 'es 2 hypothèses les plus défavorables pour son caractère s'gnificatif, à savoir :

1° que la courbe de fréquence des rangs en a des sujets premiers en b soit étalée au maximum et rectangulaire,

2° qu'i en soit de même de a courbe de fréquence des rangs en b des sujets premiers en a.

Soit un ensemble parent où le rang moyen en b des sujets premiers en a est égal au rang moyen en a des sujets premiers en b. On procède au contrôle sur deux lots, chaque lot comprenant 10 groupes de 100 sujets.

On obtient 2 valeurs différentes pour le rang en a des sujets premiers en b.

soit Vab et V$'ab$

Posons $M = \frac{Vab + Vba}{2}$, et $M' = \frac{V'ab + V'ba}{2}$.

Si un écart de 0,12 est peu vraisemblable entre Vab et Vba, il est peu vraisemblable aussi entre Vab et V$'ab$.

Si un écart de 0,12 est peu vraisemblable entre Vba et Vab, il est peu vraisemblable aussi entre Vba et V$'ba$.

S'il est peu vraisemblable entre Vab et V$'ab$ et entre Vba et V$'ba$, *a fortiori* sera-t-.l peu vraisemblab e entre les moyennes M et M'.

Donc soient 2 aptitudes. Pour l'une on a déterminé, sur 10 lots de 100 sujets, Vab et Vba.

Puis on considère 2 autres aptitudes et on détermine Vcd et Vdc. .

On a
$$M = \frac{Vba + Vab}{2}$$

et
$$M'' = \frac{Vcd + Vdc}{2}$$

Si M et M″ diffèrent de 0,12, cette divergence n'est pas fortuite.

Remarquons que ces coefficients M sont des moyennes arithmétiques, et que nos coefficients $\overline{V}$ sont des moyennes géométriques. Or la moyenne géométrique de 2 termes est toujours plus petite que leur moyenne arithmét que et quand les termes varient, leur moyenne géométrique varie moins que leur moyenne arithmétique

Posons $\overline{V} = \sqrt{Vab \cdot Vba}$ et $\overline{V}'' = \sqrt{Vcd \cdot Vdc}$. Si un écart de 0,12 est significatif entre M et M″ il est, *a fortiori*, significatif entre V et V″.

Il en résulte que si un de nos contrôles donne $\overline{V} \leqslant 0{,}38$, nous pouvons affirmer que nous n'aurions pas pu trouver dans un autre contrôle, relatif aux mêmes aptitudes, $\overline{V} \geqslant 0{,}50$.

De même, si un de nos contrôles donne $\overline{V} \geqslant 0{,}62$, nous pouvons affirmer que nous n'aurions pas pu trouver dans un autre contrôle, relatif aux mêmes aptitudes, $\overline{V} \leqslant 0{,}50$

En résumé, avec 10 lots de 100 sujets on obtiendra :

$\overline{V} < 0{,}50$ est certain seulement si le contrôle donne $\overline{V} \leqslant 0{,}38$. De même
$\overline{V} > 0{,}50$ est certain seulement si le contrôle donne $\overline{V} \geqslant 0{,}62$.
$Vab \neq Vba$ est certain seulement si le contrôle donne une inégalité dépassant 0,12.

On pourrait faire un calcul en supposant plus de 10 lots et un nombre de sujets inférieurs à 100.

Plus le nombre de lots est grand, meilleure est l'approximation, même si le nombre de sujets par lots diminue un peu. Mais l'expérience nous a montré qu'avec des petits lots, la détermination du rang en *b* du sujet premier en *a*, ou du rang en *a* du premier en *b*, devient très difficile à cause des sujets *ex aequo*. Aussi conseillerons-nous, pous savoir si une différence entre 2 valeurs de $\overline{V}$, ou entre Vab et Vba est significative, de ne pas utiliser moins de 1.000 sujets partagés en lots de 100.

Le même raisonnement s'appliquera aux valeurs de $\overline{W}$.

Si on les a déterminées sur 10 lots de 100 sujets, on retiendra que

$\overline{W} < 0{,}50$ est certain seulement si le contrôle donne $\overline{W} \leqslant 0{,}38$
$\overline{W} > 0\ 50$ est certain seulement si le contrôle donne $\overline{W} \geqslant 0{,}62$
$Wab \neq Wba$ est certain seulement si le contrôle donne une inégalité dépassant 0,12.

Une autre considération est à retenir.

Le schéma sur lequel se fonde la notion de corrélation suppose que chaque aptitude varie de façon linéaire en fonction de chaque variable. Or ceci sera sûrement vrai si les variations considérées sont très petites par rapport aux variations possibles, parce qu'un très petit segment de courbe peut toujours être assimilé à une droite.

Si nous considérons une population très homogène à ce point de vue que le niveau de chaque aptitude est très éloigné de l'optimum humainement possible, toute la théorie classique des corrélations sera applicable avec les équivalences que l'on démontrera plus loin.

Au sein d'une population quelconque nous pourrons isoler les plus mauvais dans deux aptitudes et considérer une population composée d'individus comme eux. Le schéma des corrélations est applicable à cette population.

Si dans une population, le schéma classique des corrélations est applicable, deux cas peuvent se présenter :

1° *Les aptitudes considérées dépendent des mêmes variables.*

Nous sommes dans le schéma des nuages (fig. 1) qui nous montre une corrélation d'autant plus forte que l'on est plus éloigné des points représentatifs des individus géniaux.

En réalité pour représenter l'humanité entière, il faudrait un point pessimum représentant l'idiot anencéphale. Mais aucun point ne devrait être plus éloigné d'un optimum quelconque que ce pessimum. Ce pessimum devrait être représenté par 2 points symétriques, 2 pôles d'une hypersphère. Dans le plan équatorial de celle-ci se situeraient des optimums. Ce serait un espace courbe, non euclidien.

Les déplacements près du pôle sont nécessairement presque parallèles au plan équatorial. Il en résulte que, dans cette population de très mal doués, les liaisons entre les aptitudes sont très faibles.

Pour raisonner dans un espace qui, bien que multidimensionnel, reste euclidien, nous avons rejeté le point pessimum à l'infini, comme on peut s'en rendre compte dans notre schéma des nuages. Cela signifie qu'on a pu s'éloigner des individus géniaux, sans se rapprocher sensiblement de l'anencéphale. C'est une hypothèse. Elle peut cesser d'être légitime sur une population de sujets extrêmement déficients. On cessera alors de trouver des corrélations de plus en plus fortes au fur et à mesure que l'on s'éloignera des optimums.

2° *Elles ne dépendent pas des mêmes variables.*

En ce cas le raisonnement qui vient d'être fait ne vaudra que pour les variables communes.

Les coordonnées correspondant aux optimums de différentes aptitudes auront des valeurs très voisines pour chaque variable commune, mais ces aptitudes ne dépendent pas uniquement des mêmes variables.

Soient des variables $\vec{j}, \vec{k}, \vec{l}, \vec{m}, \vec{n}$

Le schéma classique est applicable. C'est pourquoi les individus les plus doués sont représentés par des points, définis par des vecteurs ayant comme composantes les variables susmentionnées. Mais de plus, on pourra considérer qu'il y a des composantes $= 1$ et des composantes $= 0$.

Plusieurs aptitudes peuvent avoir le même nombre de composantes $= 1$, et le même nombre de composantes $= 0$, sans que ces valeurs concernent les mêmes composantes.

Il sera alors possible d'énumérer ces vecteurs de façon à réaliser une substitution circulaire.

Le 1er sujet en a, est représenté par l'extrémité du vecteur

$$1\vec{i} + 1\vec{j} + 1\vec{k} + 1\vec{l} + 0\vec{m} + 0\vec{n}$$

Le 1^er^ sujet en b est représenté par l'extrémité du vecteur

$$1\vec{j} + 1\vec{k} + 1\vec{l} + 1\vec{m} + 0\vec{n} + 0\vec{i}$$

Le 1^er^ sujet en c est représenté par l'extrémité du vecteur

$$1\vec{k} + 1\vec{l} + 1\vec{m} + 1\vec{n} + 0\vec{i} + 0\vec{j}$$

Le 1^er^ sujet en d est représenté par l'extrémité du vecteur

$$1\vec{l} + 1\vec{m} + 1\vec{n} + 1\vec{i} + 0\vec{j} + 0\vec{k}$$

Le 1^er^ sujet en e est représenté par l'extrémité du vecteur

$$1\vec{m} + 1\vec{n} + 1\vec{i} + 1\vec{j} + 0\vec{k} + 0\vec{l}$$

Le 1^er^ sujet en f est représenté par l'extrémité du vecteur

$$1\vec{n} + 1\vec{i} + 1\vec{j} + 1\vec{k} + 0\vec{l} + 0\vec{m}$$

Les individus de rang n en a, de rang n en b, etc... sont représentés par d'autres points sur ces mêmes vecteurs à une même distance de l'origine fonction inverse de n.

La corrélation entre 2 aptitudes est proportionnelle au cosinus de l'angle que forment les vecteurs correspondants, donc au module du vecteur qui est le produit scalaire de deux vecteurs.

Deux vecteurs qui ont une composante commune non nulle ont un produit scalaire avec une composante = 1 et les autres composantes nulles, leur produit scalaire a un module égal à $\sqrt{1} = 1$.

Deux vecteurs qui ont 2 composantes communes non nulles donnent un produit scalaire avec deux composantes non nulles; ce produit scalaire a un module = $\sqrt{2}$ etc...

Dans le cas particulier que nous avons envisagé (cf. p. 000) la corrélation entre deux aptitudes est proportionnelle à la racine carrée du nombre de composantes communes.

On trouvera alors en énumérant les aptitudes dans l'ordre correspondant à la substitution circulaire de leurs composantes, une répartition périodique.

Telle est l'explication des corrélations réparties en « circumplex », découvertes par Guttman.

Le schéma classique des corrélations correspond à un cas particulier de notre schéma, celui où l'on est très loin à la fois de l'optimum humainement possible et des déficients extrêmes.

Quand cette situation est presque parfaitement réalisée, on peut en outre envisager un cas plus particulier encore, où les aptitudes considérées dépendent à peu près du même nombre de variables, bien que ces variables diffèrent d'une aptitude à l'autre. C'est alors que l'on observe la répartition en « circumplex » de Guttman.

POSSIBILITÉ DE TRANSFORMATION DU COEFFICIENT D'ÉLOIGNEMENT V EN COEFFICIENT DE CORRÉLATION R

Nous disons qu'il y a corrélation stricto sensu entre une aptitude α et une aptitude β si toutes les hypothèses admises dans la démonstration de la formule de Bravais-Pearson, et par conséquent dans celle de Spearman, sont remplies, c'est-à-dire si $\overline{Y} = f(x)$ et $\overline{X} = f(y)$ sont des fonctions linéaires, et si σx, y est indépendant de y, et σy, x indépendant de x. Nous avons vu que ceci est un cas particulier de notre schéma.

Quand il y a *corrélation stricto sensu*

$$Vab = Vba,\ Wab = Wba,\ \overline{V} = \overline{W}$$

On peut dire réciproquement que si

$Vab = Vba$; $Wab = Wba$; $\overline{V} = \overline{W}$, il y a ***corrélation stricto sensu.*** En pareil cas les méthodes classiques d'analyse factorielle nous paraissent légitimes.

Il nous semble d'ailleurs qu'une vérification est possible. Étant donné que

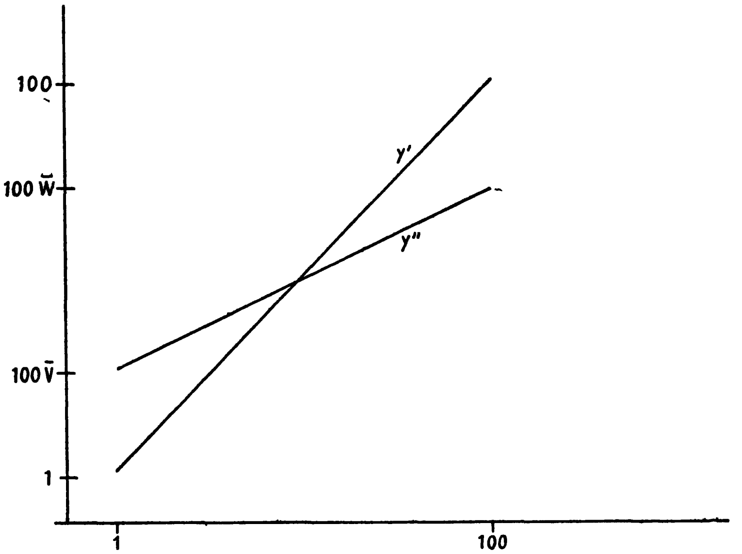

Fig. 8.

le ρ de Spearman peut être calculé à partir de $\overline{V}$, on peut vérifier l'identité du résultat déterminé par ce procédé avec le ρ calculé directement.

Voici comment on peut faire cette transformation à condition qu'elle soit légitime,

Construisons $\overline{Y} = f(X)$ (fig. 8).

Soient 2 modalités de Y :

Y' qui correspond à l'éventualité où les rangs en b représentés en Y égalent rangs en a représentés en X.

Y'' correspond aux véritables rangs en b dans l'éventualité rencontrée. Les différences de rang correspondent en $y'' - y' = l$.

En pareil cas $\Sigma\, l^2$ peut être assimilé au volume d'un solide, volume égal à

$$\int Y^2\, d(x) \quad \text{ou} \quad Y = l^2$$

Nous pouvons appliquer la formule du prismoïde

$$\int_s^t = \frac{h}{6}\left[f\,(s) + 4\,f\left(\frac{s+t}{2}\right) + f\,(t)\right]$$

$$\text{or } f\,(s) = (1 - 100\,\overline{V})^2 = f\left(\frac{s+t}{2}\right) = 0$$

$$f\,(t) = -\,(100 - 100\,\overline{V} - 100)^2 = -\,(100\,\overline{V})^2$$

et $$h = (100)^2 = 10^4$$

$f\,(t)$ est de signe opposé à $f\,(s)$ parce que symétrique de $f\,(s)$ par rapport à un point, ce qui nous donne

$$\Sigma\, l^2 = \frac{10^4}{6}\left|\, -200\,\overline{V}\,\right| = -\frac{10^6}{3}\,\overline{V}$$

Avec $n = 100$, on obtient

$$\frac{6\,\Sigma\, l^2}{n\,(n^2 - 1)} = -\,\frac{10^6 \cdot 6\,\overline{V}}{999\,200 \cdot 3}$$

ou à très peu près $(-\,2\,\overline{V})$.

La formule de Spearman devient alors $\rho = 1 - 2\,\overline{V}$ formule valable à condition que $\overline{V}$ soit exact, et que l'on ait

$$\overline{V} = \overline{W};\ Vab = Vba;\ Wab = Wba$$

auquel cas d'ailleurs on peut aussi bien écrire $\rho = 1 - 2\,\overline{W}$.

Même dans une population où le schéma classique des corrélations n'est pas valable, c'est-à-dire où la fonction de régression n'est pas linéaire, on peut considérer qu'il existe une sous-population des plus mauvais à l'interview de laquelle le schéma classique des corrélations sera applicable si les sujets ne sont pas trop déficients.

Dans cette sous-population, on aura alors $\rho = 1 - 2\,\overline{W}$.