

JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

FRANÇOIS BASTENAIRE

Théorie statistique de la mesure et des comparaisons de productivité

Journal de la société statistique de Paris, tome 93 (1952), p. 140-153

http://www.numdam.org/item?id=JSFS_1952__93__140_0

© Société de statistique de Paris, 1952, 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/>

IX

VARIÉTÉ

Théorie statistique de la mesure et des comparaisons de productivité

PREMIÈRE PARTIE

1. Introduction.

Au cours des études sur la productivité effectuées au Service d'Études et de Mesures du Comité National de la Productivité, divers problèmes se sont posés.

L'un des plus importants consiste à ramener des résultats concernant des produits différents à une commune mesure. Il se pose lorsque, dans une profession déterminée, les produits fabriqués ou les services rendus sont de diverses natures, ce qui est naturellement très général. Ainsi, dans l'industrie de la chaussure, les produits diffèrent par la pointure et par le type (chaussures hommes-ville, femmes, enfants, etc...). De même, dans les chemins de fer, les services rendus peuvent s'exprimer en voyageurs-kilomètres, ou en tonnes-kilomètres.

L'étude théorique que nous présentons vise à résoudre le problème de l'équivalence de productions différentes dans le cas particulier où les produits fabriqués sont d'un seul type, mais diffèrent par un nombre quelconque de caractéristiques continues (les dimensions par exemple). Ce problème se pose dans un grand nombre d'industries : les produits métallurgiques, tréfilés ou laminés, les tubes d'acier, les produits chimiques fabriqués à différents degrés de concentration, etc...

Dans la filature du coton, le fil est caractérisé par son numéro métrique (1) et par sa torsion; dans une étude non encore publiée (2), nous avons pu mettre en évidence l'influence très importante du numéro du fil fabriqué, sur les résultats de productivité par entreprise. Dans cette étude, nous avons calculé la corrélation entre le numéro moyen du fil fabriqué par une entreprise et la productivité du travail (exprimée en kilomètres de fil par heure d'ouvrier) et la corrélation entre le numéro moyen et la productivité des broches, exprimée en kilomètres de fil par broche et par heure.

Le problème de la définition correcte du numéro moyen est excessivement important, car les gammes correspondant aux diverses entreprises sont très étendues et se recouvrent dans une large mesure.

Le problème que nous avons résolu est celui de la détermination de la régression réelle entre la productivité du travail ou d'un facteur quelconque, et les caractéristiques dimensionnelles du produit : la détermination du numéro

(1) Nombre de kilomètres par 500 grammes de fil.

(2) Étude basée sur des données statistiques obtenues par le Syndicat Général de l'Industrie Cotonnière.

moyen et la définition de l'équivalence de productions différentes peuvent alors s'en déduire immédiatement.

2. *Hypothèses.*

Nous admettrons dans la suite, que l'ensemble des conditions économiques, psychologiques et humaines d'une époque détermine une loi de distribution telle que toutes les entreprises d'une branche professionnelle peuvent être considérées comme formant un échantillon tiré au hasard, dans la population définie par cette loi.

Cette hypothèse paraît naturelle lorsque l'on examine l'action des facteurs nombreux, variés et instables, qui peuvent affecter la productivité d'une entreprise.

Il est bien connu, par exemple, que l'exploitation des mines de charbon se présente sous des formes qui rendent l'extraction plus ou moins difficile, que la qualité des matières premières offertes à une industrie quelconque varie avec le temps et le lieu, que le personnel embauché, même lorsqu'il est soumis à des examens scientifiques, a des aptitudes variables, etc...

Le chef d'entreprise lui-même a une formation culturelle qui diffère de celle de son voisin; ils ne peuvent, tous deux, connaître tout des méthodes d'organisation du travail, de la comptabilité, de leur fabrication, et leurs aptitudes et leurs goûts sont nécessairement différents. La réflexion nous conduit donc, logiquement, à cette hypothèse, qui est d'ailleurs confirmée par l'étude expérimentale de la répartition statistique de la productivité.

3. *Principes de la comparaison d'articles de dimensions variables.*

Les seuls obstacles rencontrés lors des comparaisons de productivité concernant des articles identiques, sont de nature matérielle. Au contraire, des difficultés d'interprétation surgissent dès que les articles diffèrent.

Il est toutefois intuitif, qu'entre l'identité parfaite et la non-comparabilité totale, il existe des stades intermédiaires et que la comparaison n'est pas dépourvue de sens lorsque l'analogie est assez étroite entre deux produits.

Il arrive fréquemment que des produits de même nature, fabriqués selon des procédés analogues, ne diffèrent que par des caractéristiques dimensionnelles. Un exemple en est donné par les productions métalliques de laminés, de tréfilés ou de tubes.

Il est clair que du fait de l'analogie des procédés, les influences des conditions générales s'exercent sur les produits de toutes dimensions et selon une fonction continue des caractéristiques du produit, mais les influences aléatoires subsistant, une description complète du phénomène ne peut être donnée qu'en utilisant la théorie de la régression, celle-ci ayant lieu entre la grandeur étudiée (productivité) et les caractéristiques. La représentation de la productivité par entreprise dans un ensemble économique homogène, qui, pour un produit déterminé (la traverse de chemin de fer, par exemple), se fait sous la forme d'une distribution unidimensionnelle, se fera, dans le cas du fil métallique, sous la forme bidimensionnelle, et pour le tube d'acier, par exemple, tridimensionnelle.

Nous comprenons désormais dans quelles conditions nous effectuerons les comparaisons de productivité, que ce soit dans l'espace ou le temps.

La comparaison de deux échantillons d'entreprises appartenant à deux époques ou à deux pays différents, se ramènera à celle de deux populations unidimensionnelles, s'il s'agit de produits parfaitement définis. Mais s'il s'agit de produits dépendant d'une ou de plusieurs caractéristiques continues, la comparaison sera celle de deux ensembles bi ou multidimensionnels, ayant chacun leurs répartitions marginales propres et leur régression propre.

Lorsque les caractéristiques varient, non plus d'une façon continue, mais par sauts, comme, par exemple, dans le cas de produits normalisés dont les dimensions ne prennent qu'un nombre fini de valeurs, nous admettrons encore l'existence d'une fonction de régression dont seuls quelques points sont connus et les comparaisons entre pays resteront possibles, même si les dimensions type choisies ne coïncident pas (cela peut arriver si les deux pays ne possèdent pas la même unité de mesure).

Ce qui est essentiel, c'est que les comparaisons ne vont plus porter sur les résultats effectivement obtenus par les entreprises du premier et du second échantillon, mais sur les caractéristiques d'ensemble de ces échantillons, et que, par conséquent, ces comparaisons vont rester possibles, même si ces derniers ne s'interpénètrent pas, c'est-à-dire, même si les gammes de produits fabriqués ne se recouvrent en aucun point.

En effet, la connaissance de la répartition multidimensionnelle qui nous est fournie par l'échantillon, s'étend au delà des limites de celui-ci et nous renseigne, en quelque sorte, sur la productivité probable qui serait observée pour des productions encore non réalisées, les conditions économiques (et autres) d'ensemble, restant les mêmes.

Les comparaisons seront donc effectuées sur des ensembles d'après les méthodes de la statistique mathématique et le genre de problème à résoudre sera celui-ci : deux échantillons d'entreprises étant donnés, est-il raisonnable d'admettre que ceux-ci appartiennent, ou non, à une même population infinie théorique. En d'autres termes, est-il possible d'admettre que l'ensemble des conditions économiques, techniques, psychologiques et humaines, qui existent pour l'échantillon n° 1, existent aussi pour l'échantillon n° 2, et ceci, même si les caractéristiques des produits sont différentes en 1 et en 2? Il suffira que les procédés de fabrication soient les mêmes, pour qu'il soit possible d'imputer les différences observées aux autres facteurs de productivité. L'étude se poursuivra normalement, d'ailleurs, quant à l'influence de chacun des facteurs, sous la forme, par exemple, d'une analyse de corrélation.

La méthode reste valable lorsque les procédés sont différents, mais lorsqu'il en est ainsi, il n'est plus possible de savoir si la différence est due aux procédés ou à d'autres facteurs, et la comparaison est seulement globale.

DEUXIÈME PARTIE

DÉTERMINATION ET CALCUL DES FONCTIONS DE RÉGRESSION

1^o Chaque entreprise ne fabrique le produit que dans une dimension seulement.

Pour fixer les idées, nous prendrons l'exemple simple du fil métallique.

Chaque entreprise fournit alors un point sur le diagramme productivité, diamètre, dont l'abscisse (le diamètre) est certaine. On dispose alors de tous les procédés statistiques classiques pour rechercher la régression de la productivité par rapport au diamètre.

En particulier, si l'on suppose que pour chaque valeur de la caractéristique, la productivité suit autour de sa moyenne une loi de Laplace Gauss d'écart type constant, on peut appliquer la méthode des moindres carrés et ajuster un polynôme de degré quelconque aux résultats observés.

2^o Chaque entreprise fabrique une gamme de produits.

2. 1.) Soient Q_1, Q_2, \dots, Q_n , les quantités fabriquées sur les valeurs c_1, c_2, \dots, c_n de la caractéristique.

On pourra choisir arbitrairement l'unité de mesure des quantités produites; il suffit que la caractéristique c restant fixe, le travail dépensé soit proportionnel à ces quantités. De même, l'unité de la caractéristique pourra, dans le cas du fil métallique être le diamètre, son inverse, la section ou le poids au mètre.

Il vient naturellement à l'esprit de se poser la question suivante :

— Quelle est la valeur de la caractéristique c telle que le même temps est nécessaire pour fabriquer la quantité totale de produits sur c uniquement, ou pour la fabriquer en diverses dimensions c_1, c_2, \dots, c_n et respectivement en quantités Q_1, Q_2, \dots, Q_n ?

Il s'agit donc de calculer une moyenne possédant la propriété évidente de conserver les temps nécessaires.

Soient :

$$t_1, t_2, \dots, t_n$$

les temps nécessaires à la production d'une unité sur les valeurs

$$c_1, c_2, \dots, c_n$$

de la caractéristique dans une entreprise déterminée. Alors, le temps total est :

$$T = Q_1 t_1 + Q_2 t_2 + \dots + Q_n t_n$$

Soit $t(c)$ le temps nécessaire pour fabriquer l'unité de production sur la caractéristique c .

Alors, notre problème s'écrit :

$$Q t(c) = T$$

où

$$Q = Q_1 + Q_2 + \dots + Q_n$$

T qui est le temps nécessaire à la production du complexe, est supposé connu, mais c est une inconnue.

Diverses hypothèses sont possibles quant à $t(c)$. Nous allons en étudier quelques-unes, en commençant par les plus simples, et nous finirons par l'étude d'une hypothèse très générale qui permet d'étendre les possibilités de la méthode de calcul de la régression par les moindres carrés.

2. 2.) Supposons la fonction $t(c)$ connue.

Nous poserons $t = \varphi(c)$ et nous supposerons que la forme de la fonction φ est certaine.

Cette fonction est évidemment particulière à l'entreprise considérée.

On aura :

$$t_1 = \varphi(c_1) \dots t_n = \varphi(c_n)$$

d'où :

$$T = Q_1 \varphi(c_1) + \dots + Q_n \varphi(c_n)$$

et d'autre part :

$$T = Q \varphi(c)$$

Il vient donc :

$$Q \varphi(c) = \sum_{i=1}^n Q_i \varphi(c_i)$$

équation qui, résolue en c , fournit une racine c_0 qui résout le problème.

On remplacera le complexe :

$$\begin{matrix} Q_1, Q_2, \dots, Q_n, T \\ c_1, c_2, \dots, c_n \end{matrix}$$

par les quantités :

$$Q, T, c_0$$

Remarque : Si les temps t_1, t_2, \dots, t_n étaient connus, on pourrait penser à placer directement sur le diagramme général par entreprise (Productivité en ordonnées, caractéristique en abscisses) les points $(c_1, t_1); (c_2, t_2); \dots (c_n, t_n)$.

On risquerait pourtant de commettre une erreur due au fait que ces temps forment une classe particulière parmi l'ensemble des résultats, puisqu'ils correspondent à une même entreprise. Dans l'estimation des caractéristiques générales d'un échantillon, il en résulterait une pondération involontaire selon le nombre de valeurs de la caractéristique c_1, c_2, \dots, c_n qui donnerait aux diverses entreprises des importances différentes. Il est donc préférable de rechercher à déterminer c_0 .

Pratiquement on ne connaît presque jamais la fonction $\varphi(c)$, mais il n'est pas impossible qu'une entreprise particulière ait procédé à une étude de temps permettant de retrouver φ .

2. 3.) La fonction $\varphi(c)$ est inconnue, mais on peut en obtenir un ajustement.

Si on connaît les résultats $Q_1, Q_2, \dots, Q_n, c_1, c_2, \dots, c_n$, et les temps unitaires correspondants, t_1, t_2, \dots, t_n , on peut ajuster une fonction φ à ces résultats, lorsque n est assez grand.

De cette fonction, on déduira la valeur de c_0 telle que, si l'on avait fabriqué sur cette caractéristique seulement, le temps le plus probable que l'on aurait réalisé aurait précisément été égal à T .

2. 4.) *La fonction $\varphi(c)$ de chaque entreprise n'est pas connue, mais la fonction de régression telle qu'elle apparaît au 1^o est connue.*

Désignons par Ψ cette fonction.

Cette fois, les temps t_1, t_2, \dots, t_n ne sont pas connus, mais on peut chercher à leur attribuer une valeur en utilisant une pondération au moyen de la fonction Ψ .

On peut alors proposer la formule :

$$Q \Psi(c_0) = \sum_{i=1}^n Q_i \Psi(c_i)$$

analogue à celle citée en 2. 2.), mais avec la différence que la fonction Ψ utilisée est la même pour toutes les entreprises. Cela revient à choisir pour t_i la valeur la plus probable $\Psi(c_i)$ ou valeur moyenne.

S'il est évident qu'une telle pondération est valable lorsque l'entreprise est voisine de la moyenne, il n'en est pas de même lorsque celle-ci s'en éloigne, sauf, peut-être pour certaines formes de distributions conditionnelles de la productivité (distribution statistique de la productivité lorsque c est constant).

D'autre part, la connaissance de la fonction suppose une étude préalable, basée sur des renseignements qu'il est peu probable d'obtenir sous la forme voulue.

2. 5.) *On ne connaît que les quantités Q_1, Q_2, \dots, Q_n fabriquées dans chaque usine les valeurs c_1, c_2, \dots, c_n , ainsi que le temps total T utilisé.*

Hypothèse 1. — Il existe, dans une entreprise déterminée, un grand nombre de facteurs généraux de productivité, tels que les capacités d'organisation de la direction, le système de salaires, le climat de l'entreprise, la valeur de l'équipement. Ces facteurs influent, généralement, sur toute la gamme des produits fabriqués sans qu'un produit déterminé soit favorisé par rapport aux autres.

C'est pourquoi il nous semble naturel d'admettre soit que :

Si la probabilité d'observer un temps unitaire supérieur à t_1 pour une valeur c_1 de la caractéristique, est P , celle d'observer un temps unitaire supérieur à t_2 pour la valeur c_2 de la caractéristique, est aussi égale à P , lorsque t_1 et t_2 représentent des temps unitaires d'une même entreprise.

ou encore que :

Si la densité de probabilité est π en t_1, c_1 , elle sera aussi égale à π en t_2, c_2 , les temps unitaires t_1 et t_2 correspondant à deux valeurs distinctes de la caractéristique, mais à une même entreprise.

Cette dernière hypothèse semblera plus familière si nous faisons remarquer qu'elle ne correspond qu'à attribuer une même valeur de la vraisemblance à tous les résultats observés dans une usine.

Dans le premier cas, si nous posons :

$$\lambda_c = P_r \{ t > t_c/c \}$$

on peut transformer cette relation, et, moyennant certaines conditions de régularité, écrire :

$$t_c = f(c, \lambda_c)$$

donc

$$t_1 = f(c_1, \lambda_1)$$

$$t_2 = f(c_2, \lambda_2)$$

$$t_n = f(c_n, \lambda_n)$$

et notre hypothèse revient à admettre que :

$$\lambda_1 = \lambda_2 \dots = \lambda_n = \lambda$$

On peut remarquer d'ailleurs que la seconde hypothèse s'exprime de la même façon, mais que la fonction f sera, en général, différente.

Nous pourrions donc, en partant de la relation :

$$(2.5.1.) \quad Q_1 t_1 + Q_2 t_2 + \dots + Q_n t_n = T$$

écrire :

$$(2.5.2.) \quad Q_1 f(c_1, \lambda) + Q_2 f(c_2, \lambda) + \dots + Q_n f(c_n, \lambda) = T$$

Cette équation admet en λ une racine λ_0 , et nous prendrons pour valeur de c moyen, la valeur c_m telle que :

$$(2.5.3.) \quad T = Q f(c_m, \lambda_0)$$

Cas particulier :

En utilisant (2.5.2) et (2.5.3), on obtient l'égalité :

$$(2.5.4.) \quad Q_1 f(c_1, \lambda) + Q_2 f(c_2, \lambda) + \dots + Q_n f(c_n, \lambda) = Q f(c_m, \lambda)$$

Il peut arriver que celle-ci ait lieu pour une certaine valeur de c_m quel que soit λ et donc aussi pour λ_0 .

En donnant à λ une valeur fixe quelconque, $f(c, \lambda)$ devient une fonction de c seulement, qui servira au calcul de c_m pour toutes les entreprises, et la solution du problème sera très simplifiée, comme nous le verrons plus loin.

Extension du procédé de calcul de c_m dans le cas d'une hypothèse plus générale (Hypothèse probabiliste).

Admettons l'existence d'une fonction $f(c, \lambda)$ telle qu'à chaque entreprise correspond une valeur de λ bien déterminée (qui exprime sa productivité indépendamment de c).

Supposons que pour une entreprise donnée les temps unitaires t_1, t_2, \dots, t_n correspondant à des caractéristiques c_1, c_2, \dots, c_n , soient déterminés par un processus stochastique ne dépendant pas de λ , mais que :

$$(2.5.5.) \quad E[t_1/\lambda] = f(c_1, \lambda)$$

$$E[t_2/\lambda] = f(c_2, \lambda)$$

$$E[t_n/\lambda] = f(c_n, \lambda)$$

alors l'espérance mathématique du temps total T nécessaire à la production du complexe est :

$$\begin{aligned} E [T/\lambda] &= E [Q_1 t_1 + Q_2 t_2 + \dots + Q_n t_n] \\ &= Q_1 f(c_1, \lambda) + Q_2 f(c_2, \lambda) + \dots + Q_n f(c_n, \lambda) \end{aligned}$$

En remplaçant $E [T/\lambda]$ par la valeur T effectivement réalisée, on obtiendra, en résolvant en λ l'équation :

$$(2. 5. 2.)' \quad Q_1 f(c_1, \lambda) + Q_2 f(c_2, \lambda) + \dots + Q_n f(c_n, \lambda) = T$$

une estimation λ_0 de λ .

λ_0 et T étant des estimations de λ et de $E [T/\lambda]$ l'équation :

$$(2. 5. 3.)' \quad T = Q f(c_m, \lambda_0)$$

résolue en c_m fournira alors une estimation de c_m .

Cas particulier :

On peut écrire que l'espérance mathématique du temps total T' nécessaire à la production d'une quantité Q quelconque pour la valeur c_m de c , est :

$$E [T'/\lambda] = Q f(c_m, \lambda)$$

Si, en particulier, $Q = Q_1 + Q_2 + \dots + Q_n$, le complexe équivalent T' , Q , c_m tel que :

$$E [T'/\lambda] = E [T/\lambda]$$

sera tel que :

$$(2. 5. 4.)' \quad Q_1 f(c_1, \lambda) + \dots + Q_n f(c_n, \lambda) = Q f(c_m, \lambda)$$

Si cette égalité est indépendante de λ du fait de la forme de la fonction $f(c, \lambda)$, la valeur de c_m sera déterminée au moyen d'une fonction de c seulement comme dans le cas particulier précédent.

2. 6.) *Traitement mathématique des cas particuliers du paragraphe (2. 5.) :*

Cherchons à quelle famille de fonctions, $f(c, \lambda)$ doit appartenir pour que la relation (2. 5. 4.) soit valable quel que soit λ .

L'hypothèse est la suivante :

Étant donné un ensemble $Q_1, Q_2, \dots, Q_n, c_1, c_2, \dots, c_n$, il existe une valeur c telle que :

$$(2. 6. 1.) \quad \sum_{i=1}^n Q_i [f(c_i, \lambda) - f(c, \lambda)] = 0$$

a lieu identiquement en λ .

Nous allons montrer que si une fonction $f(c, \lambda)$ est telle que :

$$(2. 6. 2.) \quad Q_1 f(c_1, \lambda) + Q_2 f(c_2, \lambda) = (Q_1 + Q_2) f(c, \lambda)$$

a lieu quel que soit λ pour une certaine valeur de c quand Q_1, Q_2, c_1, c_2 sont donnés arbitrairement, cette fonction peut satisfaire (2. 6. 1.).

En effet, si (2. 6. 2.) est réalisé, on peut écrire :

$$Q_1 f(c_1, \lambda) + Q_2 f(c_2, \lambda) + Q_3 f(c_3, \lambda) = (Q_1 + Q_2) f(c, \lambda) + Q_3 f(c, \lambda)$$

mais puisque la propriété est supposée vraie pour $n = 2$, le second membre se transforme et s'écrit :

$$(Q_1 + Q_2 + Q_3) f(c', \lambda) \quad \text{c. q. f. d.}$$

et la démonstration s'étend à n quelconque.

Soit :

$$(2. 6. 3.) \quad Q_1 f(c_1, \lambda) + Q_2 f(c_2, \lambda) = Q f(c, \lambda)$$

Pour

$$\lambda = \lambda_1$$

on a :

$$(2. 6. 4.) \quad Q_1 f(c_1, \lambda_1) + Q_2 f(c_2, \lambda_1) = Q f(c, \lambda_1)$$

et enfin on a :

$$(2. 6. 5.) \quad Q_1 + Q_2 = Q$$

Les équations (2. 6. 3.), (2. 6. 4.), (2. 6. 5.), forment un système homogène admettant une solution en Q_1 et Q_2 , donc le déterminant

$$\Delta = \begin{vmatrix} f(c_1, \lambda_1) & f(c_2, \lambda_1) & f(c, \lambda_1) \\ f(c_1, \lambda) & f(c_2, \lambda) & f(c, \lambda) \\ 1 & 1 & 1 \end{vmatrix} = 0$$

Cette dernière condition, lorsque Q_1, Q_2, c_1, c_2 , sont donnés arbitrairement devant être remplie pour au moins une valeur de c quel que soit λ .

En mettant (2. 6. 4.) sous la forme :

$$\frac{Q_1}{Q} f(c_1, \lambda_1) + \frac{Q_2}{Q} f(c_2, \lambda_1) = f(c, \lambda_1)$$

et en remarquant que :

$$0 \leq \frac{Q_1}{Q} \leq 1 \quad 0 \leq \frac{Q_2}{Q} \leq 1$$

on peut démontrer, en supposant pour fixer les idées :

$$f(c_1, \lambda_1) < f(c_2, \lambda_1)$$

que :

$$f(c_1, \lambda_1) \leq \frac{Q_1}{Q} f(c_1, \lambda_1) + \frac{Q_2}{Q} f(c_2, \lambda_1) \leq f(c_2, \lambda_1)$$

Ce qui prouve que si la fonction $f(c, \lambda_1)$ est continue en c en tout point, elle admet toujours une racine satisfaisant (2. 6. 4.).

En donnant à Q_1 et Q_2 , toutes les valeurs positives possibles, c peut même prendre toutes les valeurs de l'intervalle c_1, c_2 .

Comme il est nécessaire que les équations (2. 6. 3.), (2. 6. 4.), (2. 6. 5.), soient satisfaites non pas pour un seul système de Q_1, Q_2, c_1, c_2 , mais pour tous les systèmes possibles, il faut donc aussi qu'elles admettent une solution quel que soit c .

En développant le déterminant selon les éléments de la troisième colonne, on obtient une condition de la forme :

$$\varphi(\lambda) f(c, \lambda_1) + k f(c, \lambda) + \Psi(\lambda) = 0$$

Ce qui veut dire que $f(c, \lambda)$ possède nécessairement la forme :

$$f(c, \lambda) = \theta(\lambda) \mu(c) + \tau(\lambda)$$

On vérifie immédiatement que cette condition est suffisante en substituant $f(c, \lambda)$ dans (2. 6. 1.).

Remarque : Si la répartition des entreprises autour de la courbe de régression est normale avec un écart type constant, on peut écrire :

$$f(c, \lambda) = \varphi(c) + g(\lambda) \quad .$$

et nous sommes dans les conditions d'application des développements précédents :

$$\sum_{i=1}^{i=n} Q_i [f(c_i, \lambda) - f(c, \lambda)] = 0$$

s'écrit :

$$\sum_{i=1}^{i=n} Q_i \varphi(c_i) = \varphi(c_m) \sum_{i=1}^{i=n} Q_i$$

2. 7.) La fonction $f(c, \lambda)$ est inconnue, mais on suppose qu'elle peut s'écrire $f(c, \lambda) = \varphi(c) + g(\lambda)$ (Hypothèse 1 non probabiliste).

Les éléments que l'on suppose connus, concernant les usines sont les suivants :

Pour chaque entreprise, le complexe :

$$Q_i, c_1; Q_i, c_2; \dots; Q_i, c_n$$

relatif à une période définie, et le temps total T nécessaire à la production de ce complexe.

On supposera d'autre part que $g(\lambda)$ définit une répartition normale, et que l'on peut donc employer la méthode des moindres carrés.

Notations :

Soit i l'indice de l'entreprise

$$1 \leq i \leq n.$$

Soit j l'indice relatif à une quantité fabriquée sur une certaine valeur de la caractéristique dans l'usine i

$$1 \leq j \leq k_i.$$

Caractéristique : c_i

Quantités : Q_i

Temps unitaire global : $\tau_i = \frac{T_i}{\sum_j Q_{ij}}$

Détermination de la fonction $\varphi(c)$:

La caractéristique moyenne \bar{c}_i relative à l'entreprise i est alors déterminée par l'équation (voir remarque du paragraphe (2. 6.) :

$$(2. 7. 1.) \quad \sum_{j=1}^{j=k_i} Q_{ij} \varphi(c_{ij}) = \varphi(\bar{c}_i) \sum_{j=1}^{j=k_i} Q_{ij} \quad (i = 1, 2, \dots, n)$$

et l'application de la méthode des moindres carrés revient à rechercher quelle fonction φ appartenant à une famille définie rend minimum l'expression :

$$(2. 7. 2.) \quad \sum_{i=1}^{i=n} [\tau_i - \varphi(\bar{c}_i)]^2$$

Comme on peut d'après (2. 7. 1.) écrire :

$$\varphi(\bar{c}_i) = \frac{1}{\sum_{j=1}^{j=k_i} Q_{ij}} \sum_{j=1}^{j=k_i} Q_{ij} \varphi(c_{ij})$$

on obtient en substituant dans (2. 7. 2.), la condition :

$$(2. 7. 3.) \quad \sum_i [\tau_i - \sum_j \frac{Q_{ij}}{Q_i} \varphi(c_{ij})]^2 \text{ minimum}$$

où la notation Q_i est employée pour :

$$Q_i = \sum_j Q_{ij}$$

Le problème est alors résolu en principe lorsque la famille proposée pour φ est connue.

Exemple :

φ appartient à la famille des fonctions linéaires de c :

Cherchons à ajuster une fonction linéaire aux observations.

$$\text{Soit : } \varphi = a c + b$$

Notre équation s'écrit :

$$\sum_i [\tau_i - \sum_j \frac{Q_{ij}}{Q_i} (a c_{ij} + b)]^2 \text{ minimum.}$$

On voit aisément qu'en dérivant par rapport à a puis à b , on obtient deux équations linéaires, comme cela a lieu par la méthode ordinaire des moindres carrés.

Ainsi, on a :

$$(2. 7. 4.) \quad \sum_i \sum_j \tau_i \frac{Q_{ij} C_{ij}}{Q_i} = a \sum_i \left(\sum_j \frac{Q_{ij} C_{ij}}{Q_i} \right)^2 + b \sum_i \sum_j \frac{Q_{ij} C_{ij}}{Q_i}$$

et :

$$(2. 7. 5.) \quad \sum_i \sum_j \frac{\tau_i Q_{ij}}{Q_i} = a \sum_i \sum_j \frac{Q_{ij} C_{ij}}{Q_i} + b n$$

Il en résulte que la connaissance des c_{ij} et Q_{ij} permet donc de déterminer quand même la régression $\varphi = a c + b$ et que l'on retombe sur un cas analogue au cas classique.

Nous remarquerons enfin que l'on obtient encore des équations linéaires si l'on désire ajuster un polynôme de degré n quelconque aux observations.

(2. 8.) *Détermination de la fonction de régression dans le cas de l'hypothèse probabiliste :*

Soit τ_i le temps unitaire global correspondant à l'entreprise i . On a alors :

$$E [\tau_i | \lambda] = E \left[\frac{T_i}{Q_i} \right] = \frac{1}{Q_i} E [T_i] = f(c_m, \lambda)$$

où c_m est déterminé par (2. 5. 4.)'.

Nous pouvons écrire :

$$\tau_i = x + y$$

où

$$\begin{aligned} x &= \tau_i - f(c_m, \lambda) \\ y &= f(c_m, \lambda) \end{aligned}$$

Les variables x et y étant telles que :

$$E [x] = 0 \qquad E [y] = \int_0^1 f(c_m, \lambda) d\lambda$$

on en déduit que :

$$E [\tau_i] = \int_0^1 f(c_m, \lambda) d\lambda$$

Supposons que :

$$(2. 8. 1.) \qquad f(c, \lambda) = \theta(\lambda) \mu(c) + v(\lambda)$$

Dans ce cas :

$$(2. 8. 2.) \qquad E [\tau_i] = \int_0^1 [\theta(\lambda) \mu(c_m) + v(\lambda)] d\lambda$$

et l'on voit que $E(\tau_i)$ ne dépend que de c_m .

Appliquer la méthode des moindres carrés revient à rechercher la fonction $\mu(c)$ choisie dans un ensemble, qui rendra l'expression :

$$(2. 8. 3.) \qquad \sum_i [(\tau_i - E[\tau_i])^2] \text{ minimum}$$

Mais pour que cette méthode soit applicable, il faut que :

$$E[(\tau - E[\tau])^2 / c]$$

soit indépendante de c .

Or d'après (2. 8. 1) et (2. 8. 2) :

$$\tau - E[\tau] = \theta(\lambda) \mu(c) + v(\lambda) - \mu(c) \int_0^1 \theta(\lambda) d\lambda - \int_0^1 v(\lambda) d\lambda$$

d'où :

$$\begin{aligned} (\tau - E[\tau])^2 &= [\theta(\lambda) - \int_0^1 \theta(\lambda) d\lambda]^2 \mu^2(c) - 2 \mu(c) [\theta(\lambda) - \int_0^1 \theta(\lambda) d\lambda] \cdot \\ &\quad [v(\lambda) - \int_0^1 v(\lambda) d\lambda] + [v(\lambda) - \int_0^1 v(\lambda) d\lambda]^2 \end{aligned}$$

En prenant l'espérance mathématique de cette expression, on peut voir que si la fonction :

$$\theta(\lambda) - \int_0^1 \theta(\lambda) d\lambda$$

n'est pas nulle en moyenne (1),

$$E[(\tau - E[\tau])^2 / c]$$

(1) Nulle en tout point sauf sur un ensemble de mesure nulle.

est de la forme :

$$a \mu^a(c) + b \mu(c)$$

fonction qui, quel que soit a non nul et b (même nul) ne pourrait être constante que si la fonction $\mu(c)$ ne pouvait prendre qu'une ou deux valeurs particulières.

Ainsi donc en général :

$$E[(\tau - E[\tau])^2/c]$$

ne sera indépendante de c que si la fonction :

$$\theta(\lambda) = \int_0^1 \theta(\lambda) d\lambda$$

est nulle presque partout.

En se restreignant pour $\theta(\lambda)$ à une fonction continue, il faut donc que :

$$\theta(\lambda) = \int_0^1 \theta(\lambda) d\lambda = 0$$

ce qui entraîne :

$$\theta(\lambda) = \text{constante.}$$

et l'on vérifie que la condition est suffisante.

La seconde condition d'application de la méthode des moindres carrés est que τ soit réparti lorsque c est constant, selon une loi de Laplace-Gauss.

Dans ces conditions, nous écrirons en convenant que \bar{c}_i désigne la valeur de c_m relative à l'entreprise i :

$$E[\tau_i] = K_1 \mu(\bar{c}_i) + K_2$$

où \bar{c}_i est déterminé par :

$$(2. 8. 4.) \quad \sum_{j=1}^{j=\bar{K}_1} Q_{ij} \mu(c_{ij}) = \mu(\bar{c}_i) \sum_{j=1}^{j=\bar{K}_1} Q_{ij}$$

car quels que soient K_1 et K_2 on peut effectuer la simplification comme à la « Remarque » du paragraphe (2. 6.).

La condition (2. 8. 3) s'écrit alors comme (2. 7. 2.) :

$$\sum_{i=1}^{i=n} [\tau_i - (K_1 \mu(\bar{c}_i) + K_2)]^2 \text{ minimum.}$$

Comme la fonction qui nous intéresse est :

$$\varphi(c) = E[\tau/c] = K_1 \mu(c) + K_2$$

et que l'équation (2. 8. 4.) ne change pas quand on multiplie $\mu(c)$ par K_1 et qu'on ajoute K_2 quelconque, on retrouve exactement la condition (2. 7. 2.), et le calcul pratique se poursuit de la même façon.

2. 9.) Généralisation au cas où le produit dépend de plusieurs caractéristiques.

Désignons par $\alpha, \beta, \dots, \theta$ les caractéristiques au nombre de p .

Pour simplifier nous désignerons symboliquement cet ensemble par la lettre c , qui représente alors un point dans un espace à p dimensions.

Les considérations développées au paragraphe (2. 5.) restent valables, mais le système des équations (2. 5. 2.) et (2. 5. 3.) ne permet plus de déterminer

c_m d'une manière unique. En général, c_m pourra appartenir à un ensemble défini par $p - 1$ paramètres.

De même dans le cas de l'hypothèse probabiliste, on peut encore poser les équations (2. 5. 5.), c_1, c_2, \dots, c_n , représentant n ensembles de p paramètres.

Cependant le traitement mathématique du cas particulier, où l'identité :

$$Q_1 f(c_1, \lambda) + Q_2 f(c_2, \lambda) + \dots + Q_n f(c_n, \lambda) = Q f(c_m, \lambda)$$

a lieu en λ pour un c_m donné reste valable. En effet, le problème s'exprime toujours par (2. 6. 1.), la démonstration par récurrence peut être généralisée, et les équations (2. 6. 3.), (2. 6. 4.) et (2. 6. 5.), donnent lieu aux mêmes conditions. Le déterminant doit encore être nul quel que soit c , car en faisant varier Q_1, Q_2, c_1, c_2 , la variable c décrit son domaine de variation.

Le développement du déterminant redonne la condition :

$$f(c, \lambda) = \theta(\lambda) \mu(c) + \tau(\lambda)$$

Dans le cas de l'hypothèse non probabiliste, (2. 7. 1.) et (2. 7. 2.) doivent être vérifiés et conduisent à (2. 7. 3.).

Lorsque la fonction φ dépend linéairement de ses paramètres, on est encore conduit à un système d'équations linéaires.

Les mêmes considérations s'étendent au cas où l'on fait l'hypothèse probabiliste.

Conclusion :

Plusieurs méthodes d'ajustement de données ont été proposées.

1° Déterminer pour chaque usine une fonction $t = \varphi(c)$ qui caractérise la régression dans l'usine de t par rapport à c . Utiliser $\varphi(c)$ pour la pondération des résultats et le calcul de c moyen.

2° Utiliser une fonction $\varphi(c)$ (se rapprochant le plus possible de la fonction de régression réelle) obtenue d'après les résultats d'un calcul de régression.

3° A la rigueur, utiliser une fonction $\varphi(c)$ théorique basée le plus objectivement possible sur les faits. Par exemple, pour la production du fil de fer, le temps est d'autant plus élevé que le diamètre est plus fin (car il faut un plus grand nombre de passes). On peut donc proposer par exemple, une fonction du type :

$$t = \frac{1}{d^\alpha}$$

4° Faire le calcul de régression d'après la méthode que nous avons indiquée en dernier lieu.

- a) Si les Q_v ne sont pas trop nombreux, en particulier si les k_v sont faibles, calculer directement.
- b) Si les Q_v sont nombreux, chercher pour chaque entreprise la répartition des quantités fabriquées dans des intervalles ou classes choisis à l'avance et prendre pour valeur des c_v les centres de ces intervalles.

François BASTENAIRE.

Statisticien au Comité National de la Productivité.