JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

LUCIEN MARCH

Comparaison numérique de courbes statistiques

Journal de la société statistique de Paris, tome 46 (1905), p. 255-277
<http://www.numdam.org/item?id=JSFS_1905__46__255_0>

II

COMPARAISON NUMÉRIQUE DE COURBES STATISTIQUES (1)

I — Liaisons apparentes des faits comparés

1. *Dépendance apparente parfaite.*

Lorsque l'on compare des courbes statistiques, on est naturellement porté — quelle que soit la hardiesse d'une telle entreprise, mais la comparaison n'aurait sans cela aucune utilité — à apprécier, par induction, la dépendance mutuelle des circonstances qui font varier les grandeurs représentées.

La comparaison des courbes met d'ailleurs en évidence des liaisons plus ou moins étroites.

Si deux grandeurs sont liées invariablement, à tout changement de l'une correspond un égal changement de l'autre et les courbes représentatives des variations de ces grandeurs sont parallèles.

Réciproquement, le parallélisme des courbes démontre l'invariabilité du lien qui unit les grandeurs représentées. Nous en inférons que, si ces grandeurs mesurent des faits susceptibles de connexité, ces faits sont dans une étroite dépendance. Par exemple, si pour deux marchés différents les courbes du prix du blé sont parallèles, et si les situations géographiques respectives, l'état des communications, l'importance relative des transactions, etc., ne semblent pas s'opposer à la solidarité de ces deux marchés, le parallélisme des courbes engage à admettre l'étroite dépendance des deux marchés.

Il convient d'y insister, le parallélisme des courbes ne suffit pas pour démontrer la liaison rigide des faits représentés. Il laisse soupçonner des rapports, il fournit des vérifications, il suggère des opinions, mais n'impose pas de certitude.

Les comparaisons graphiques ou numériques auxquelles on a recours en statistique constituent en fait d'excellents instruments de recherche : elles ne dispensent pas de réfléchir; elles appuient et précisent le jugement: à proprement parler, elles ne le conditionnent pas, elles n'en fixent pas la formule. Cependant, tant que

---

1. Communication faite à la Société de statistique de Paris dans la séance du 18 janvier 1905. Voir numéro de février 1905, p. 47. (Suite de la communication insérée dans le numéro de décembre 1904.)

nous ne connaissons pas le mécanisme des faits comparés, tant que nous ne pouvons apprécier leurs rapports que par des effets quantitatifs et des apparences d'ordre numérique, le parallélisme des courbes est un indice — et souvent le seul dont nous disposions — que, dans leur évolution, les phénomènes représentés obéissent aux mêmes influences.

Il est aujourd'hui superflu de dire que, parmi les faits d'observation dont l'enchaînement nous échappe, il n'est point d'exemple de parallélisme parfait, mais les cas de parallélisme approché sont assez fréquents : par exemple dans le domaine des lois physiques. Parmi les faits dont s'occupe la statistique, l'approximation est rarement bien grande : on peut cependant citer, comme exemples de courbes sensiblement parallèles, celles du mouvement des naissances masculines et du mouvement des naissances féminines ; les courbes des cours au plus haut et au plus bas pour les valeurs ou les marchandises sur un même marché ou sur des marchés solidaires ; les courbes qui représentent l'accroissement du réseau des chemins de fer, le tonnage des marchandises, le nombre des voyageurs ; les courbes qui font connaître à diverses époques la fréquentation scolaire et la proportion des individus sachant lire. Ces exemples appellent l'attention sur les interprétations différentes auxquelles peut donner lieu le parallélisme des courbes.

Si la courbe du tonnage des marchandises transportées par voie ferrée est parallèle à la courbe qui représente le développement successif du réseau, ou bien si la courbe qui met en évidence l'accroissement de la proportion des individus sachant lire est parallèle à celle de la fréquentation scolaire, ce parallélisme était prévu et par suite expliqué d'avance : l'un des deux phénomènes comparés est la conséquence naturelle de l'autre.

Mais, quand nous comparons la courbe des cours de la rente 3 °/₀ au plus haut avec la courbe des cours du même titre au plus bas, ou bien la courbe de la natalité masculine avec celle de la natalité féminine, il n'y a plus de lien de causalité, tout au plus a-t-on affaire à des phénomènes qui subissent pareillement l'effet des mêmes influences.

Pour bien marquer la part de l'opération statistique et la part de l'interprétation, on peut dire que le parallélisme des courbes démontre la *concomitance* des mouvements représentés et qu'il permet seulement de soupçonner la *connexité,* la *dépendance* des circonstances qui déterminent ces mouvements. Pour raison d'euphonie, et en vue de l'extension qui va être faite quelques lignes plus loin, nous emploierons le mot de « dépendance » au lieu de celui de « concomitance » qui serait plus exact, étant entendu que la dépendance révélée par la comparaison des courbes est purement formelle. C'est sous ces réserves, et après ces explications, que nous inférerons du parallélisme des courbes à la parfaite dépendance des faits représentés.

Cette dépendance est encore parfaite lorsque deux courbes, au lieu d'être parallèles, sont, comme l'on dit, antiparallèles, c'est-à-dire lorsque, à tout changement de l'une des grandeurs représentées, correspond un changement égal et contraire de l'autre. Dans ce cas, il suffirait de retourner l'une des courbes sur elle-même pour obtenir, avec l'autre demeurée fixe, deux courbes parallèles. Deux courbes dont l'une représenterait le mouvement des élèves dans les écoles et l'autre les changements survenus dans la proportion des illettrés fourniraient un exemple d'antiparallélisme approché.

Dans le cas où deux courbes sont parallèles, la dépendance des grandeurs comparées peut être appelée directe ou positive ; lorsqu'elles sont antiparallèles, la dépendance peut être dite inverse ou négative. Tels les mouvements solidaires de systèmes d'engrenages combinés pour marcher les uns dans le même sens, les autres en sens contraire.

2. *Indépendance apparente complète.*

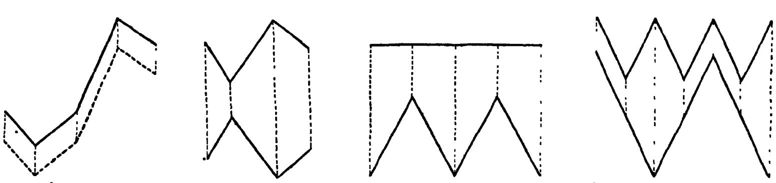
En opposition avec les cas de liaison invariable des mouvements comparés, se placent les circonstances dans lesquelles ces mouvements n'ont entre eux aucun lien, et cela se peut constater sous deux formes distinctes.

Ou bien l'une des courbes représentatives se réduit à une ligne droite horizontale parallèle à l'axe fondamental, c'est-à-dire que l'un des phénomènes demeure invariable quels que soient les changements subis par l'autre, ce qui exclut évidemment toute apparence de liaison.

Ou bien, tandis que certains des changements corrélatifs impliqueraient une liaison directe des deux phénomènes, d'autres également importants dénoteraient une liaison inverse ; dans leur évolution, les deux phénomènes se trouveraient tantôt en accord, tantôt en désaccord, comme, au jeu de pile ou face, la pièce de monnaie tombe tantôt sur pile tantôt sur face : les circonstances qui déterminent l'accord ou le désaccord des phénomènes étudiés nous semblent aussi indépendantes les unes des autres que les circonstances auxquelles est imputable la chute de la pièce de monnaie soit sur le côté pile, soit sur le côté face [1].

Comme exemple d'indépendance du premier genre, on peut citer le maintien à peu près invariable de la proportion des naissances masculines au total des naissances, dans certains pays, malgré une baisse considérable de la natalité : ce sont deux faits sans relation apparente.

Les exemples d'indépendance du second genre sont naturellement nombreux : il est plus facile de trouver des groupes de phénomènes indépendants que des groupes de phénomènes corrélatifs et nous aurons l'occasion d'en signaler dans la suite.



Parallélisme · *Anti-Parallélisme* · *Indépendance (1er genre)* · *Indépendance (2e genre)*

3. *Dépendance apparente partielle.*

Nous avons dit qu'en toute rigueur les cas de dépendance parfaite ne se présentent jamais dans l'étude des faits d'observation. Le plus souvent, le parallélisme ou l'antiparallélisme des courbes n'est pas très net ; la comparaison de ces courbes donne l'impression d'un certain accord ou d'un certain désaccord, mais cette impression demeure forcément imprécise et vague, en raison de la multiplicité des points de comparaison, tant que l'œil est seul juge.

---

1. Ce n'est, d'ailleurs, encore qu'une opinion provisoire susceptible d'être modifiée par une connaissance plus complète des faits comparés.

En effet, l'œil peut-il toujours apprécier si le parallélisme de deux courbes est plus ou moins marqué que celui de deux autres ? Ne faut-il pas se méfier des idées préconçues qui font attacher plus d'importance aux cas d'accord qu'à ceux de désaccord ou inversement ? L'impression ne peut-elle varier suivant les observateurs ? N'a-t-on pas cru longtemps, par exemple, que les mariages se faisaient plus rares aux époques de cherté, alors qu'aujourd'hui cette théorie est abandonnée.

Pour éviter ces causes d'incertitude, il convient de suivre les courbes point par point, et d'énumérer les cas d'accord ainsi que ceux de désaccord, avant de porter un jugement sur la dépendance des faits représentés. C'est ce que nous allons faire pour quelques exemples.

II — Indice de dépendance

1. *Détermination d'un indice de dépendance.*

(Mariages, naissances, décès en France depuis 1873.)

Examinons par exemple la courbe des mariages et celle des naissances en France durant la période de 1873 à 1903.

D'après le tableau I annexé à cette notice, en 1874, le nombre des nouveaux mariés pour 10 000 habitants a été plus grand qu'en 1875 ; la courbe s'abaisse à partir de son point de départ ; elle s'abaisse encore de 1874 à 1875 et les années suivantes, pour se relever de 1878 à 1879 et ainsi de suite. Celle des naissances se relève de 1873 à 1874, s'abaisse de 1874 à 1875 et ainsi de suite. En comparant les deux courbes, on remarque que, pour certaines années telles que 1874, 1876, etc., les variations observées par rapport à l'année suivante sont de même sens pour les deux courbes (concordance), tandis qu'à certaines autres époques (1873, 1875, etc.) ces variations sont de sens contraire (discordance).

En comptant le nombre des concordances et le nombre des discordances, on se fera une idée de la similitude des courbes de nuptialité et de natalité et par là on se rendra compte de la liaison des deux phénomènes.

Représentons par le signe + toute variation positive d'une année à la suivante, par le signe — toute variation négative ; la série des signes est relevée pour les deux courbes de nuptialité et de natalité sur le tableau ci-dessous.

ANNÉES	MARIAGES	NAISSANCES	DÉCÈS	ANNÉES	MARIAGES	NAISSANCES	DÉCÈS	ANNÉES	MARIAGES	NAISSANCES	DÉCÈS
1873. . . .	+	—	+	1883. . . .	—	+	—	1893. . . .	0	+	+
1874. . . .	+	+	—	1884. . . .	+	+	+	1894. . . .	+	+	—
1875. . . .	+	—	+	1885. . . .	+	+	—	1895. . . .	—	—	+
1876. . . .	+	+	+	1886. . . .	+	+	+	1896. . . .	0	+	+
1877. . . .	0	+	—	1887. . . .	+	+	+	1897. . . .	+	+	—
1878. . . .	—	+	0	1888. . . .	+	+	+	1898. . . .	—	—	—
1879. . . .	+	+	—	1889. . . .	+	+	—	1899. . . .	—	+	—
1880. . . .	—	—	+	1890. . . .	—	—	—	1900. . . .	—	—	+
1881. . . .	+	+	—	1891. . . .	—	+	+	1901. . . .	+	+	+
1882. . . .	—	0	0	1892. . . .	+	—	+	1902. . . .	0	+	+
								1903. . . .			

Pour compter le nombre des concordances, on notera par le signe + l'association de deux variations de même signe (+ et + ou — et —) et par le signe — l'association de deux variations de signe contraire. On observe alors

18 associations positives (+) ou concordances,
7 associations négatives (—) ou discordances,
5 associations de deux variations dont l'une est nulle ou indifférences,

en sorte que, sur 30 couples donnant soit une concordance, soit une discordance, 18 donnent une concordance, soit 60 °/₀, et 7 une discordance, soit 23 °/₀. On est donc autorisé à conclure qu'en France, de 1873 à 1893, toute variation du nombre des mariages, d'une année à l'autre, est le plus souvent accompagnée d'une variation de même sens du nombre des naissances. Effectuons une comparaison semblable entre les mariages d'une année et les naissances de l'année suivante, nous obtenons sur 29 intervalles

12 concordances,
13 discordances,
4 cas d'indifference,

soit : concordances, 41 °/₀ ; discordances, 45 °/₀.

Entre les mariages d'une année et les naissances de la seconde année qui suit, sur 28 intervalles

16 concordances,
8 discordances,
4 cas d'indifférence,

soit : concordances, 57 °/₀ ; discordances, 29 °/₀.

Entre les mariages d'une année et les naissances survenues trois ans après, sur 27 intervalles on compte

11 concordances,
12 discordances,
4 cas d'indifférence,

c'est-à-dire ce que donnerait une distribution au hasard des concordances et des discordances.

Ainsi, durant l'intervalle de temps considéré, le nombre des mariages d'une année paraît sans liaison avec le nombre des naissances de l'année suivante, ou de la troisième année suivante ; il semble au contraire en connexion assez étroite avec le nombre des naissances survenues, soit dans l'année de mariage, soit deux ans après.

Admettons que des constatations du même genre ressortent de l'examen des courbes de nuptialité et de natalité prolongées sur un grand nombre d'années, notre confiance dans la réalité des liaisons dont nous venons de soupçonner l'existence s'accroîtra et l'on essaiera de les expliquer : on admettra volontiers par exemple qu'un certain nombre de mariages ayant pour but de légitimer des enfants conçus auparavant, il est naturel que les unions contractées au cours d'une année influent sur les naissances de l'année, et d'autre part, l'influence des mariages d'une année se faisant surtout sentir sur les naissances survenues deux ans après, on en conclura que — pour divers motifs — la première naissance ne suit pas toujours le mariage de très près, en sorte que les mariages ont leur principale répercussion sur les naissances survenues deux ans après leur conclusion.

Nous n'entrerons dans aucune discussion, notre but étant surtout de décrire la méthode.

Appliquons encore cette méthode à la comparaison des naissances et des décès au cours de la même période 1873-1903.

Comparaison des naissances d'une année avec les décès :

	Nombre des concordances	Nombre des discordances	Cas d'indifférence
1° De l'année antéprécédente	12	13	3
2° De l'année précédente	20	6	3
3° De la même année	12	16	2
4° De l'année suivante	13	13	3
5° De la seconde année suivante	13	12	3

Ce petit tableau indique que le nombre des concordances est le plus grand, et n'est d'ailleurs significatif que dans le cas où l'on compare les naissances d'une année avec les décès de l'année précédente, c'est-à-dire que, conformément aux vues que M. Bertillon exposait il y a quelque temps (1), la natalité d'une année paraît soumise à l'influence de la mortalité durant l'année précédente (2). Et comme la liaison ainsi observée existerait entre des faits successifs, il est naturel de regarder l'un comme une cause et l'autre comme un effet et de dire que, d'après les apparences, les changements de la natalité sont en partie les effets de changements antérieurs de la mortalité.

Ces procédés donnent une précision numérique à l'impression visuelle ; à la rigueur, ils peuvent dispenser de construire les courbes.

2. *Formule de l'indice de dépendance.*

(Comparaison de divers articles du bilan de la Banque de France.)

D'après ce que nous venons de voir, si c représente le nombre des concordances, d le nombre des discordances et n le nombre des intervalles, la fraction $\frac{c-d}{n}$ ou, ce qui revient au même quand aucune variation n'est nulle, $\frac{c-d}{c+d} = i$ est l'expression de l'indice de dépendance.

Les conventions algébriques des signes facilitent le calcul de cet indice.

Représentons, pour une série quelconque de nombres, chaque variation d'un nombre au suivant par la lettre V affectée du signe $+$ si le premier nombre est supérieur au second, affectée du signe $-$ si le premier nombre est inférieur au second. On marquera un zéro quand, les deux nombres successifs étant égaux, la

1. *Journal de la Société de statistique de Paris*, numéro d'octobre 1904.
2. La conclusion est la même si la comparaison porte sur les décès de la première année seulement.

variation est nulle. A titre d'exemple, opérons sur différents chapitres des bilans de la Banque de France depuis 1874 jusqu'en 1903 et formons le tableau ci-après :

Articles du bilan de la Banque de France. — Sens des changements annuels de divers comptes.

ANNÉES	ENCAISSE	TAUX DE L'ESCOMPTE	ESCOMPTE	COMPTES COURANTS			ANNÉES	ENCAISSE	TAUX DE L'ESCOMPTE	ESCOMPTE	COMPTES COURANTS			ANNÉES	ENCAISSE	TAUX DE L'ESCOMPTE	ESCOMPTE	COMPTES COURANTS		
				Mouvements	Solde	Virements					Mouvements	Solde	Virements					Mouvements	Solde	Virements
1874.	—	+	+	—	—	—	1884.	—	0	+	+	+	+	1894.	—	+	+	—	—	—
1875.	—	+	+	+	—	+	1885.	—	0	+	—	—	—	1895.	+	+	—	+	—	+
1876.	—	+	—	+	—	+	1886.	+	0	+	+	+	+	1896.	+	0	—	—	+	—
1877.	+	+	—	—	+	—	1887.	+	—	—	—	—	—	1897.	+	—	—	—	+	—
1878.	—	—	+	—	—	—	1888.	—	—	—	—	—	—	1898.	+	—	—	—	+	—
1879.	+	—	—	—	+	—	1889.	—	+	—	—	+	—	1899.	—	—	—	—	—	+
1880.	+	—	—	—	—	—	1890.	—	0	—	—	—	—	1900.	—	+	+	—	—	—
1881.	—	+	+	+	—	+	1891.	—	+	+	+	+	+	1901.	—	0	+	—	+	—
1882.	+	+	+	+	+	+	1892.	—	+	—	—	+	—	1902.	+	0	—	—	+	—
1883.	—	+	+	+	+	+	1893.	—	0	+	—	—	—	1903.						

Dans la première colonne sont portées les années successives et, dans les autres colonnes, les signes des différences entre chaque nombre correspondant à l'année inscrite en regard et le nombre correspondant à l'année suivante.

Six colonnes sont affectées, la première au montant de l'encaisse, la seconde au taux moyen de l'escompte, la troisième au montant des effets escomptés, la quatrième au total des sommes versées ou retirées par les particuliers ayant à la Banque un compte courant, la cinquième au solde moyen des dépôts en comptes courants, la sixième au montant des virements effectués par la Banque.

Pour simplifier la composition du tableau, on n'a pas fait figurer les lettres *v*, qui représentent des variations, on n'a inscrit que les signes.

Si l'on admet, comme nous l'avons fait antérieurement, qu'une discordance détruit une concordance, le nombre des concordances à faire entrer en compte pour juger, par exemple, du parallélisme de la courbe de l'encaisse et de la courbe de l'escompte s'obtiendra en formant la somme algébrique suivante :

$$(-v)\times(+v)+(-v)\times(+v)+(-v)\times(-v)+(+v)\times(-v)+\ldots\ldots+(-v)\times(+v)$$
$$+(+v)\times(-v)$$

En effet, dans toute concordance, les deux variations sont ou toutes deux positives ou toutes deux négatives et dans les deux cas le produit est positif ; dans toute discordance, les deux variations sont de signe contraire, le produit est négatif.

En formant la somme algébrique, les produits négatifs se retranchant des produits positifs, les discordances annulent en quelque sorte automatiquement des concordances en nombre égal.

L'expression ci-dessus sera rendue indépendante de la quantité arbitraire *v* et du nombre des produits si l'on divise la somme algébrique des produits par le nombre de ces produits et par le facteur $v \times v$. L'expression symbolique de l'indice peut donc s'écrire

$$\frac{v\times v+v\times v+v\times v+\ldots\ldots+v\times v}{n\,(v\times v)} \quad \text{ou} \quad \frac{\Sigma v\times v}{n\,(v\times v)}$$

n représentant le nombre des intervalles durant lesquels on note les concordances ou les discordances, et les quantités entre parenthèses au dénominateur étant prises positivement.

Au numérateur, chaque produit $v \times v$ est affecté du signe qui lui convient, tandis que le dénominateur est nécessairement positif.

Pour simplifier l'écriture, on peut poser $v = 1$; le dénominateur se réduit au nombre des intervalles.

On remarquera que si l'une des variations est nulle, ce qui arrive lorsque deux nombres d'une même série sont égaux, le produit correspondant est nul aussi et n'intervient pas dans l'indice. Et, en effet, on se trouve alors dans un cas d'indépendance complète, à une variation dans l'une des séries ne correspond aucun changement dans l'autre.

Les indices qui caractérisent la dépendance réciproque des différents articles du bilan de la Banque de France sont les suivants:

Valeur de l'indice de dépendance entre deux des articles ci-dessous :

	Escompte	Comptes courants — Mouvements	Comptes courants — Solde	Virements	Taux moyen de l'escompte
Encaisse	$\frac{-20+9}{29} = \frac{-11}{29}$	$\frac{+15-14}{29} = \frac{1}{29}$	$\frac{+20-9}{29} = \frac{11}{29}$	$\frac{+14-15}{29} = \frac{-1}{29}$	$-\frac{15-6}{29} = -\frac{9}{29}$
Escompte		$\frac{+20-9}{29} = \frac{11}{29}$	$\frac{+13-16}{29} = \frac{-3}{29}$	$\frac{+19-10}{29} = \frac{9}{29}$	$\frac{15-6}{29} = \frac{9}{29}$
Comptes courants — Mouvement			$\frac{+16-13}{29} = \frac{3}{29}$	$\frac{+28-1}{29} = \frac{27}{29}$	$\frac{15-6}{29} = \frac{9}{29}$
Comptes courants — Solde				$\frac{+15-14}{29} = \frac{1}{29}$	$\frac{11-10}{29} = \frac{1}{29}$
Virements					$\frac{14-7}{29} = \frac{7}{29}$

D'après ces résultats, les groupes pour lesquels l'indice de dépendance possède une valeur significative sont :

1° Mouvement des comptes courants et virements. L'indice est égal à $\frac{27}{29}$, c'est-à-dire très voisin de 1 ; il est dès lors évident, sans qu'il soit nécessaire de tracer les courbes, que celles-ci offrent en presque tous leurs points des variations de même sens. Le montant des sommes versées en compte courant ou retirées subit des fluctuations à peu près parallèles à celles des virements, ce qui laisse supposer que la masse des comptes courants est constituée précisément au moyen des opérations de virement.

2° La masse des opérations en compte courant offre encore un parallélisme assez marqué avec le montant des effets escomptés.

3° Une relation de même genre s'observe entre le montant de l'encaisse et le solde moyen des comptes courants. Dans les années où l'encaisse augmente, le solde moyen des dépôts tend aussi à augmenter, et inversement ; ce sont deux effets concomitants de l'abondance de l'argent.

4° Entre le montant de l'encaisse et le montant de l'escompte [1], l'indice de

1. La dépendance apparaîtrait sans doute encore plus étroite si l'on pouvait isoler l'escompte du papier commercial et l'escompte des bons du Trésor.

dépendance caractérise une relation de même importance que les deux précédentes, mais inverse et non plus directe, c'est-à-dire que si, d'une année à l'autre, l'escompte augmente, l'encaisse tend alors généralement à diminuer, et inversement. Ce sont les mouvements contraires dont M. Juglar a magistralement analysé le mécanisme.

5° On constate encore une certaine dépendance, facile d'ailleurs à expliquer, entre le montant de l'escompte et le montant des virements.

6° Le montant de l'escompte est en rapport direct avec le taux moyen de l'escompte tandis qu'au contraire :

7° Le taux moyen de l'escompte est en rapport inverse avec le montant de l'encaisse. Les autres couples fournissent des indices trop faibles pour qu'on puisse leur attribuer une signification quelconque.

On peut se demander, et c'est même une question du plus haut intérêt, si la dépendance observée entre deux articles, par exemple entre le montant de l'encaisse et le montant de l'escompte, lorsque l'on compare les nombres correspondant aux mêmes années, se manifeste encore lorsque l'on compare des nombres relevés au cours d'années différentes, ainsi que nous l'avons fait dans le cas des mariages, naissances et décès.

Si la dépendance de deux articles paraissait plus grande quand on compare la valeur de l'un dans une certaine année à la valeur de l'autre dans une autre année, on serait à même de prévoir les mouvements de l'un ou de l'autre de ces articles avec quelque chance de succès.

En fait, les comparaisons portant sur des chiffres distants d'une ou de deux années dans un sens ou dans l'autre fournissent presque toujours un indice de faible valeur, en sorte que les changements subis par ces articles, lorsqu'on associe des valeurs relevées au cours d'années différentes, semblent généralement entièrement indépendants les uns des autres.

On relève cependant au moins une exception : le taux moyen de l'escompte au cours d'une année paraît lié dans une certaine mesure au montant moyen de l'encaisse au cours de l'année précédente. Si l'on compare les signes figurant dans la troisième colonne du tableau de la page 261 avec les signes qui figurent dans la seconde colonne en les faisant chevaucher d'une année, rapprochant par exemple le signe du taux de l'escompte en 1875 du montant de l'encaisse en 1874, on relève 17 concordances, 3 discordances et 8 cas d'indifférence (cas où le taux moyen de l'escompte est resté le même d'une année à l'autre). L'indice de dépendance est donc égal à

$$\frac{17-3}{28}=\frac{1}{2}.$$

3. *Précision de l'indice de dépendance.*

Nous avons vu qu'en rapprochant les décès d'une année des naissances enregistrées l'année précédente, pendant la période 1873-1903, on comptait sur 29 intervalles, 20 concordances et 6 discordances, en sorte que l'indice de dépendance est égal à

$$\frac{20-6}{29}=\frac{14}{29},$$

valeur un peu supérieure à celle de l'indice calculé lorsque l'on compare par exemple le mouvement de l'encaisse et le mouvement des escomptes à la Banque de France.

Le psycho-physicien Fechner [1], auquel sont dus, je crois, les premières applications de cet indice $i = \frac{c - d}{c + d}$, lui attribuait la valeur d'une probabilité mathématique, c'est-à-dire qu'il assimilait l'observation du signe d'une association de deux valeurs à un tirage de boules dans une urne contenant des boules de deux couleurs.

Si l'on accepte cette assimilation, qui appellerait tout au moins certaines réserves, on peut calculer les limites des prévisions raisonnables.

L'écart probable de la probabilité p d'un événement, lorsque celle de l'événement contraire est $1 - p$, étant donné par l'expression

$$0{,}67 \sqrt{\frac{p\,(1-p)}{n}},$$

on peut estimer par exemple, avec égale chance de se tromper ou d'être dans le vrai, la chance que la mortalité française ayant baissé de 1902 à 1903, le prochain tableau du mouvement de la population de la France pour 1904 fasse ressortir une natalité inférieure. Cette chance serait comprise entre 0,63 et 0,75, l'écart probable étant égal à

$$\frac{0{,}67}{29} \sqrt{\frac{20 \times 9}{29}}.$$

III — Coefficient de dépendance

On peut faire aux indices qui viennent d'être calculés un grave reproche : ils ne tiennent pas compte de la grandeur des variations comparées. Les grandes ont autant d'importance que les petites. On déclare qu'une discordance entre deux variations annule une concordance entre deux autres variations. Cependant, si par exemple les deux premières variations, celles qui sont de même sens, sont beaucoup plus grandes que les dernières, celles de sens contraire, on conçoit que les influences qui déterminent une concordance aient plus d'action que les influences qui déterminent une discordance ; ces influences ne se neutralisent pas, il n'y a donc pas indépendance complète des deux phénomènes.

Lorsque nous avons indiqué le mécanisme algébrique du procédé, nous avons attribué à toutes les variations une même valeur, v. Pour laisser à chacune d'elles son importance, il eût fallu conserver dans la formule les grandeurs mêmes des diverses variations comparées, attribuer à chaque signe un poids égal à la grandeur du changement accompli.

La formule $i = \frac{c - d}{c + d}$, où c représente une somme de produits d'unités de même signe et d une somme de produits d'unités de signe contraire, devrait donc se transformer en $\mathrm{I} = \frac{\mathrm{C} - \mathrm{D}}{\mathrm{C} + \mathrm{D}}$, si l'on désigne par C la somme des produits deux à deux

1. Voir les œuvres posthumes de Fechner publiées par G. Lipp sous le titre : *Kollektivmasslehre* (Leipzig. 1897), page 305.

des variations concordantes et par D la somme des produits deux à deux des variations discordantes. Telle est la modification proposée par Fechner. Si, d'après cette formule, on calcule le coefficient de dépendance entre la courbe des mariages et celle des naissances, la comparaison portant sur les mêmes années, de 1874 à 1903, on obtient la valeur $\mathrm{I} = 0{,}66$, tandis que la simple considération des signes avait fourni un indice $i = 0{,}60$.

Comme le précédent indice, le rapport $\frac{\mathrm{C} - \mathrm{D}}{\mathrm{C} + \mathrm{D}}$ s'annule lorsque les concordances et les discordances sont d'égale valeur $\mathrm{C} = \mathrm{D}$; il s'annule encore si les nombres de l'une des séries comparées ne varient pas. Ce sont les deux cas d'indépendance complète des deux séries : ou bien une variation dans l'une est accompagnée aussi souvent par une variation de même sens que par une variation de sens contraire dans l'autre série, ou bien l'une des séries, étant invariable, n'a de liaison avec aucune autre soumise à variations.

Lorsque toutes les discordances disparaissent, la valeur du rapport est égale à l'unité. Dans ce cas, à toute variation dans l'une des séries correspond une variation de même sens dans l'autre. Le rapport devient égal à -1 quand toutes les concordances disparaissent ; à toute variation dans l'une des séries correspond une variation de sens contraire dans l'autre : c'est le cas d'antiparallélisme. Dans les autres circonstances, le rapport a une valeur comprise entre -1 et 0 ou entre 0 et $+1$.

Mais on remarquera que ce rapport est égal à $+1$, ou à -1, lorsque les couples de variations associées sont tous, soit de même sens, soit de sens contraire, quelle que soit leur valeur, et par conséquent ce rapport ne tient pas encore suffisamment compte de la grandeur des variations comparées [1].

Observons aussi que ce rapport, comme l'indice précédemment choisi $\frac{c - d}{c + d}$, est appliqué directement aux courbes qui représentent les observations. Par suite les grandeurs comparées dépendent des unités adoptées. Pour qu'il n'en soit pas ainsi, il faut fixer les unités par une convention uniforme et nous avons proposé, après M. Cheysson et d'autres auteurs, de prendre pour unité des nombres de chaque série leur valeur moyenne, ou une quantité liée invariablement à cette valeur moyenne.

Par conséquent, avant de former le tableau des variations des points des deux courbes considérées, il convient de les rendre comparables en rapportant les ordonnées de chacune d'elles à leur valeur moyenne.

Opérons par exemple sur les mariages et les naissances enregistrés en France de 1874 à 1903, dont les nombres pour 10 000 habitants sont inscrits dans les colonnes du tableau I annexé.

---

1. On pourrait encore adopter comme coefficient de dépendance la moyenne des rapports des variations correspondantes des deux séries, c'est-à-dire l'expression

$$\frac{1}{n} \Sigma \left(\frac{v}{v'}\right)$$

pourvu que l'on place toujours au numérateur la plus petite des deux variations (comptée en valeur absolue) et au dénominateur la plus grande. Toutefois, ce coefficient offrirait l'inconvénient de donner lieu à des calculs non symétriques et surtout il laisse encore une égale influence aux petites et aux grandes variations.

Le nombre moyen des nouveaux mariés au cours de la période 1874-1903 étant de 150 (p. 10 000 habitants) et le nombre moyen des naissances durant la même période étant de 234 (p. 10000 habitants), on divisera chacun des nombres de la colonne 2 du tableau par 150, chacun des nombres de la colonne 5 par 234, en multipliant tous les quotients par 100 pour éviter les décimales. Les résultats sont portés dans les colonnes 5 et 6.

On détermine ensuite pour chaque colonne la différence entre chaque nombre et le suivant. Pour la colonne des mariages, les différences avec leurs signes sont inscrites dans la colonne 8, pour les naissances dans la colonne 9.

Supposons maintenant que l'on construise deux courbes à l'aide de ces différences et sur les mêmes axes.

Si les courbes primitives, représentant le mouvement des mariages et celui des naissances, s'appliquent à des faits en dépendance tellement étroite qu'à toute variation dans l'une des séries correspond une variation égale et de même sens dans l'autre, les deux courbes dérivées des premières, construites sur les mêmes axes en portant en ordonnées les différences successives des nombres de chaque série, devront coïncider.

Si elles ne coïncident pas, c'est que la dépendance des termes des deux séries n'est pas parfaite; elle l'est d'autant moins que les deux courbes s'écartent davantage l'une de l'autre. Il est donc permis de regarder la liaison entre les deux séries primitives, c'est-à-dire, dans notre exemple, entre la série des mariages et celle des naissances, comme d'autant moins étroite que l'écartement des courbes dérivées est plus grand. Cet écartement peut être mesuré en prenant la moyenne des carrés des écarts des différents points, afin que le sens de chaque écart n'influe pas sur le résultat.

Le maximum du coefficient de dépendance étant toujours supposé égal à 1, on pourra mettre ce coefficient sous la forme

$$1 - \lambda \Sigma (v - v')^2,$$

si v et v' représentent deux ordonnées correspondantes des deux courbes dérivées.

Ce coefficient devient égal à 1 quand $\Sigma (v - v')^2 = 0$, c'est-à-dire quand les deux courbes dérivées coïncident, et réciproquement.

Pour qu'il s'annule lorsque les deux séries de nombres primitifs sont entièrement indépendantes, c'est à-dire quand la somme des produits vv' positifs est compensée par la somme des produits négatifs, il faut que l'on ait dans ce cas :

$$\lambda \Sigma (v - v')^2 = 1.$$

Comme par hypothèse $\Sigma vv' = 0$, cette égalité entraîne :

$$\lambda = \frac{1}{\Sigma v^2 + \Sigma v'^2}.$$

Le coefficient devient alors :

$$1 - \frac{\Sigma v^2 + \Sigma v'^2 - 2\Sigma vv'}{\Sigma v^2 + \Sigma v'^2} = \frac{\Sigma vv'}{\frac{\Sigma v^2 + \Sigma v'^2}{2}} = j.$$

Le maximum de ce coefficient est 1, valeur atteinte quand les v' sont respectivement égaux aux v.

Pour calculer ce coefficient dans le cas des mariages et des naissances, formons les produits deux à deux, avec leurs signes, des nombres contenus dans les colonnes 8 et 9 du tableau I annexé. Ces produits sont inscrits dans la colonne 11 ; leur somme algébrique obtenue en retranchant le total des produits négatifs du total des produits positifs est égale à 74.

D'autre part, la somme des carrés des nombres de la colonne 8 est égale à 163.

La somme des carrés des nombres de la colonne 9 est égale à 136.

La moyenne des deux est 149,5.

La valeur du coefficient est donc $\frac{74}{149,5} = 0,49$ tandis que, par la simple considération des signes, la méthode de Fechner avait fourni la valeur 0,60.

Le coefficient précédent est égal à l'unité lorsque les changements observés dans les séries de faits comparés sont égaux et de même sens, ce qui caractérise la dépendance parfaite des deux séries.

On peut élargir un peu la conception de dépendance absolue sans sortir d'ailleurs de l'analogie avec les systèmes rigides, en l'étendant aux cas où chaque variation dans l'une des séries de faits est non pas égale à la variation correspondante dans l'autre, mais se trouve avec cette dernière dans un rapport constant. Deux points invariablement liés subissent des déplacements égaux quand on les déplace dans une direction constante, mais ils subissent seulement des déplacements proportionnels à leur distance au centre quand ils sont soumis à des mouvements de rotation : cependant leur liaison est aussi étroite dans un cas que dans l'autre.

Pour que le coefficient précédent demeure égal à 1 dans le cas de dépendance parfaite, c'est-à-dire que $\Sigma vv' = \frac{\Sigma v^2 + \Sigma v'^2}{2}$ même lorsque l'on remplace v et v' par deux grandeurs proportionnelles $\frac{v}{m}, \frac{v'}{m'}$, il faut que

$$mm' \Sigma vv' = \frac{m'^2 \Sigma v^2 + m^2 \Sigma v'^2}{2};$$

d'où

$$\Sigma v^2 + \Sigma v'^2 = \frac{m'^2 \Sigma v^2 + m^2 \Sigma v'^2}{mm'},$$

ce qui implique $m'\Sigma v^2 [m' - m] = m \Sigma v'^2 [m' - m]$. Comme m' est différent de m, cette égalité entraîne la suivante

$$\frac{\Sigma v^2}{m} = \frac{\Sigma v'^2}{m'} \quad \text{ou} \quad \frac{m}{\sqrt{\Sigma v^2}} = \frac{m'}{\sqrt{\Sigma v'^2}}.$$

Si donc le coefficient doit être égal à 1 dans le cas où $\frac{v}{m} = \frac{v'}{m'}$, c'est-à-dire dans le cas où m est l'unité de mesure de v et m' l'unité de mesure de v', il faut que m et m' satisfassent à la relation précédente et il est facile de voir que, de plus, chacun des rapports $\frac{m}{\sqrt{\Sigma v^2}}$ $\frac{m'}{\sqrt{\Sigma v'^2}}$ doit nécessairement être égal à 1.

Ainsi, en regardant la dépendance des faits numériques observés comme parfaite lorsque, dans les deux séries, les changements subis sont simplement proportionnels, le coefficient précédent sera égal à 1 dans ce cas de dépendance parfaite, si l'on a soin de rapporter les variations v de l'une des séries à la quantité $\sqrt{\Sigma v^2}$ et les variations de l'autre série à la quantité $\sqrt{\Sigma v'^2}$.

En adoptant ces unités, l'expression du coefficient devient :

$$\frac{\Sigma vv'}{\Sigma v^2\, \Sigma v'^2\, \dfrac{\left[\dfrac{\Sigma v^2}{\Sigma v^2}+\dfrac{\Sigma v'^2}{\Sigma v'^2}\right]}{2}} = \frac{\Sigma vv'}{\sqrt{\Sigma v^2\, \Sigma v'^2}} = k.$$

Ce coefficient diffère du précédent, j, en ce que, au dénominateur, la moyenne arithmétique $\dfrac{\Sigma v^2 + \Sigma v'^2}{2}$ est remplacée par la moyenne géométrique $\sqrt{\Sigma v^2\, \Sigma v'^2}$.

Les deux coefficients j et k ne diffèrent pas beaucoup l'un de l'autre lorsque les sommes Σv^2 et $\Sigma v'^2$ ne sont elles-mêmes pas très différentes. A mesure que celles-ci se différencient, leur moyenne géométrique s'écarte de plus en plus de leur moyenne arithmétique, dans les conditions où une parabole s'écarte de sa tangente.

Lorsque l'une des sommes tend vers zéro, c'est-à-dire quand l'un des phénomènes comparés est à peu près constant, le coefficient j tend aussi vers zéro, comme il convient, puisque l'on se rapproche du cas de complète indépendance. Au contraire, le coefficient k tend vers une valeur différente de zéro (1).

A cet égard, les deux coefficients n'ont pas la même signification. Tandis que le coefficient j devient très petit quand l'une des séries comparées est à peu près constante, c'est-à-dire quand on approche de l'un des cas d'indépendance complète que nous avons considérés, le coefficient k tend vers une valeur fixe, petite, il est vrai, quand le nombre des termes de chaque série est assez grand, mais qui ne s'annule pas.

Le choix de l'un ou de l'autre coefficient est lié à l'idée que l'on se fait de la dépendance ou de l'indépendance. Considérons par exemple la courbe de la nuptialité en Angleterre et la courbe du chômage. Si nous construisons ces courbes avec méthode, c'est-à-dire en adoptant pour unité de mesure des ordonnées une unité uniforme, la moyenne des termes de chaque série, la courbe des mariages présentera des oscillations très petites par rapport aux oscillations de la courbe du chômage ; en ce sens on peut dire que d'importants changements dans la proportion du chômage n'étant accompagnés que de très faibles changements de la proportion des mariages, l'influence réciproque est minime ; mais, si la comparaison minutieuse des deux courbes révèle néanmoins un synchronisme à peu près parfait entre les fluctuations du chômage et celles des mariages, nous pourrons légitimement inférer qu'il existe une étroite dépendance entre la proportion des mariages et la proportion des chômeurs, quelque petits que soient les changements de la nuptialité qui accompagnent les mouvements beaucoup plus considérables du chômage.

Le coefficient k semble donc devoir être préféré au coefficient j sur lequel il possède d'ailleurs un autre avantage, celui d'éviter le calcul des rapports à la moyenne de chaque série.

---

1. Il est d'ailleurs facile de déterminer les valeurs du rapport de Σv^2 à $\Sigma v'^2$ pour lesquelles l'un des coefficients j ou k devient plus petit ou plus grand que l'autre.

En effet, le numérateur et le dénominateur de la fraction étant homogènes en v et v', peu importe que v ou v' représente les variations des rapports à la moyenne des nombres de chaque série ou les variations de ces nombres eux-mêmes.

On aurait pu déduire directement le coefficient k de la formule proposée par Fechner mise sous la forme indiquée dans notre première communication (voir Journal, numéro de décembre 1904, p. 419). Dans cette formule, toutes les variations entrent avec une valeur hypothétique égale. Si chacune entre avec sa valeur propre, c'est-à-dire avec son poids, on obtient le numérateur des coefficients j ou k.

Pour éviter que le coefficient ne change par le seul fait d'un changement dans la valeur moyenne des variations comparées, on rapportera chaque variation à sa valeur moyenne.

Soit m la valeur moyenne des variations v de la première série, m' la valeur moyenne des variations v' de la seconde série. La valeur d'une concordance ou d'une discordance sera représentée avec son poids et son signe par le produit

$$\frac{v}{m} \times \frac{v'}{m'}.$$

En prenant la moyenne de ces concordances ou discordances, on obtiendra le coefficient de dépendance

$$k = \frac{1}{n} \frac{\Sigma vv'}{mm'}.$$

Pour que k soit égal à 1 quand les v et les v' sont tous égaux deux à deux et de même signe, on doit avoir

$$m^2 = \frac{\Sigma v^2}{n} \qquad m'^2 = \frac{\Sigma v'^2}{n},$$

ce qui permet de mettre le coefficient sous la forme déjà obtenue par un autre procédé :

$$k = \frac{\Sigma vv'}{\sqrt{\Sigma v^2 \times \Sigma v'^2}}.$$

Appliquons cette formule aux exemples pour lesquels l'indice de dépendance calculé précédemment s'est trouvé assez élevé. Voici le tableau des coefficients calculés :

		k
Mariages comparés aux naissances de la même année		+ 0,49
Mariages comparés aux naissances survenues deux ans après		+ 0,315
Décès comparés aux naissances de l'annee suivante		+ 0,33
Banque de France (Comparaison des chiffres de la même année).	Encaisse et escomptes	— 0,615
	Escomptes et comptes courants (mouvement)	+ 0,345
	Comptes courants et virements	+ 0,86
	Escomptes et virements	+ 0,215
	Encaisse et solde des comptes courants	+ 0,46
	Escomptes et taux de l'escompte	+ 0,535
	Encaisse et taux de l'escompte	— 0,32
Encaisse comparée au taux d'escompte de l'année suivante		— 0,655

Le tableau ci-dessus confirme avec plus de précision les conclusions tirées du tableau des indices de dépendance, page 261.

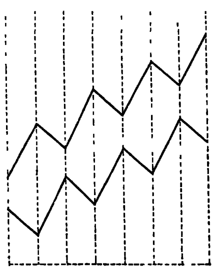
Les articles du bilan de la Banque qui dépendent le plus étroitement l'un de l'autre sont le mouvement des comptes courants et le montant des virements. Le coefficient de dépendance est encore fort élevé, mais négatif quand on compare l'encaisse d'une année au taux moyen de l'escompte l'année suivante : le taux de l'escompte semble dépendre beaucoup plus du montant antérieur de l'encaisse que du montant antérieur des escomptes.

IV — Décomposition par interpolation des changements comparés

(Mariages, naissances, chômage en Angleterre depuis 1851.)

Dans les recherches précédentes, nous avons toujours supposé qu'il s'agissait de suivre des changements annuels. Mais il arrive fréquemment que, dans les changements subis par les faits statistiques, on puisse distinguer des phases diverses. On distinguera des changements annuels, des changements polyannuels (décennaux par exemple), des changements séculaires, sans parler des périodes de temps plus courtes qu'une année.

La plupart des statistiques sont de date trop récente pour que l'on ait à s'occuper des changements séculaires, mais il est important, dans l'étude des liaisons auxquelles peuvent être soumis les faits statistiques, de séparer les modifications observées d'une année à l'autre de celles qui subsistent après une période de plusieurs années lorsque les variations annuelles ont subi des compensations.



On se rendra compte de l'importance de cette distinction en comparant les deux courbes de l'encaisse de la Banque et du montant des effets escomptés, courbes dont l'allure générale accuse un certain parallélisme, tandis que les variations annuelles sont antiparallèles, mouvements figurés sur le schéma ci-dessus. Il en résulte que le coefficient dont nous avons donné la formule est négatif.

Il convient donc, si l'on veut analyser plus complètement les conditions de dépendance de deux séries statistiques, de calculer à la fois le coefficient de dépendance des variations annuelles et un coefficient de dépendance à plus longue période, par exemple celui qui correspond à des variations décennales.

Ces deux coefficients ont leur importance. On est parfois tenté de s'en tenir à la comparaison des variations décennales en négligeant les variations annuelles regardées comme accidentelles. C'est un tort, parce qu'il faudrait d'abord prouver que ces variations annuelles sont bien accidentelles, puis parce que la dépendance de deux phénomènes est mieux établie lorsque l'on divise les variations étudiées, par exemple, en 50 groupes au lieu de 5, de même que l'expérience détermine d'une manière plus précise la composition d'une masse de boules blanches et de boules noires lorsque ces boules sont réparties en un certain nombre d'urnes que si elles sont toutes réunies en une seule.

A titre d'application, comparons les mariages et les naissances observés en Angleterre de 1851 à 1900 (tableau III annexé).

Voici le tableau des résultats par périodes décennales :

Années	Sur 10 000 habitants		Rapports à la moyenne supposée égale à 100		Différences successives		Carrés des différences		Produits des différences
	Nouveaux mariés	Naissances	Mariages	Naissances	Mariages	Naissances	Mariages	Naissances	
1851-1860	169	341	108,6	106,8					
					+ 3	— 11	9	121	— 33
1861-1870	166	352	105,6	117,8					
					+ 4	— 2	16	4	— 8
1871-1880	162	354	101,6	119,8					
					+ 13	+ 29	169	841	+ 377
1881-1890	149	325	88,6	90,8					
					— 7	+ 26	49	676	— 188
1891-1900	156	299	95,6	64,8					
Moyennes.	160,4	334,2					243	1 642	+ 154

Le coefficient de dépendance est égal à

$$k = \frac{154}{942} = +\,0{,}16$$

Le coefficient est positif et de faible valeur ; les circonstances qui déterminent les mouvements à longue période, à la fois des mariages et des naissances, semblent avoir peu d'influence. Opérons maintenant sur les chiffres annuels. La somme algébrique des produits des variations de la nuptialité et de la natalité est égale à 105. La somme des carrés de ces variations est, pour la nuptialité, 1 177, pour la natalité, 1 785 Le coefficient k est donc égal à

$$\frac{105}{\sqrt{1177 \times 1785}} = \frac{105}{1\,450} = 0{,}072.$$

Cette fois le coefficient de dépendance est positif et fort petit. Il ne semble donc exister aucune relation appréciable entre les mariages d'une année et les naissances de la même année.

En comparant les mariages de l'année aux naissances des années suivantes jusqu'à la cinquième, on observe que le coefficient applicable à chaque cas est maximum lorsque les mariages sont comparés aux naissances survenues trois ans après ; par exemple, quand on compare les mariages de 1851 aux naissances de 1854.

Les produits des variations ainsi échelonnées sont inscrits dans la colonne 12 du tableau III, leur somme algébrique est + 285 Après avoir retranché des sommes de carrés précédemment trouvées les chiffres de trois années qui ne doivent pas y entrer, on obtient pour k la valeur

$$k = \frac{285}{\sqrt{1160 \times 1688}} = \frac{285}{1400} = 0{,}236.$$

Voici d'ailleurs les valeurs successives du coefficient quand on compare les mariages d'une année aux naissances survenues les années suivantes :

Même année	0.072
1re année suiv	0,047
2e — —	0,175
3e — —	0,236
4e — —	0
5e — —	0,091

Ce ne sont pas encore des valeurs bien grandes; il est donc assez difficile de saisir la répercussion des mariages sur les naissances. Les influences qui gouvernent les unes semblent très différentes des influences qui commandent les autres.

Cependant, on peut poursuivre l'analyse avec une plus grande précision.

Les variations annuelles successives de la natalité, par exemple, résultent pour la plus grande part de circonstances qui changent notablement, et en sens divers, d'une année à l'autre, par exemple du plus ou moins grand nombre d'enfants décédés au cours d'une année, de conditions climatériques, etc., mais, pour une partie aussi, de circonstances qui changent fort peu d'une année à l'autre et se modifient lentement dans un certain sens : par exemple, développement des moyens de communication, progrès de l'instruction, de l'aisance, etc.

Pour donner à l'analyse plus de précision, il convient de décomposer les changements étudiés, ce à quoi l'on parvient en traçant une courbe interpolée entre les points qui correspondent aux observations.

Cela fait, on peut déterminer le coefficient de dépendance : 1° entre deux de ces courbes interpolées ou courbes moyennes; 2° entre les écarts des nombres observés par rapport à ces courbes moyennes.

Dans les deux premières colonnes du tableau ci-après (voir p. 273), on a inscrit les écarts, à partir des lignes moyennes, des taux observés de nuptialité et de natalité en Angleterre, depuis 1851.

Ces écarts, rapprochés des nombres observés reproduits dans les colonnes 2 et 4 du tableau III, permettent de reconstituer aisément les lignes moyennes interpolées.

La comparaison de ces deux lignes permet de distinguer trois périodes : De 1851 à 1873, la nuptialité et la natalité ont varié en sens inverse, la nuptialité tendant à diminuer tandis que la natalité tendait à augmenter. De 1873 à 1886, les deux coefficients ont varié dans le même sens, la nuptialité continuant à diminuer et la natalité commençant à diminuer aussi. Depuis 1886, les changements sont de nouveau de sens différents, la natalité continuant à décroître, tandis que la nuptialité se relevait.

En raison de ces alternances, le coefficient général de dépendance est très faible; sans donner le détail des calculs, nous dirons qu'il est seulement égal à 0,07, tandis que le coefficient calculé page 271 en procédant par périodes décennales était 0,16.

Afin de donner une nouvelle application de la méthode, nous détaillerons les calculs nécessaires à la détermination du coefficient de dépendance des variations annuelles.

Remarquons d'abord que les écarts autour de la ligne moyenne se distribuant alternativement tantôt dans un sens, tantôt dans l'autre, autour de cette ligne, sans aucune tendance dans l'un ou l'autre sens, il n'est pas utile de déterminer les variations successives de ces écarts pour calculer le coefficient de dépendance. On peut opérer directement sur ces écarts. La grandeur du coefficient de dépendance de ces écarts indiquera si de grands écarts positifs de la nuptialité sont généralement associés à de grands écarts négatifs de la natalité et inversement. Si e, e' représentent deux écarts associés, à la formule $k = \frac{\Sigma vv'}{\sqrt{\Sigma v^2 \, \Sigma v'^2}}$ on peut substituer la suivante $k = \frac{\Sigma ee'}{\sqrt{\Sigma e^2 \, \Sigma e'^2}}$.

A la suite des deux colonnes où sont inscrits les écarts, par rapport aux lignes moyennes, des taux de nuptialité et de natalité, deux autres colonnes du tableau ci-dessous renferment les carrés des écarts portés dans les deux premières. Dans les colonnes suivantes, on a calculé les produits deux à deux, avec leurs signes, des écarts de nuptialité et de natalité, d'abord en associant les chiffres des mêmes années, puis l'écart de nuptialité d'une année avec l'écart de natalité de la première année suivante, ensuite l'écart de nuptialité d'une année avec l'écart de natalité de la seconde année suivante et ainsi de suite.

ANNÉES	ÉCARTS DES COURBES VRAIES PAR RAPPORT aux courbes interpolées		CARRÉS DES ÉCARTS		PRODUITS DES ÉCARTS DE LA NUPTIALITÉ PAR CEUX DE LA NATALITÉ				
	Nuptialité	Natalité	Nuptialité	Natalité	la même année	un an après	deux ans après	trois ans après	quatre ans après
1851.	+ 1	+ 5	1	25	+ 5				
1852.	+ 4	+ 5	16	25	+ 20	+ 5			
1853.	+ 9	— 5	81	25	— 45	— 20	— 5		
1854.	+ 2	+ 2	4	4	+ 4	+ 18	+ 6	+ 2	
1855.	— 8	— 4	64	16	+ 32	— 8	— 36	— 16	— 4
1856.	— 3	+ 2	9	4	— 6	— 16	+ 4	+ 18	+ 8
1857.	— 5	0	25	0	0	0	0	0	0
1858.	— 9	— 8	81	64	+ 72	+ 40	+ 21	+ 64	— 16
1859.	+ 1	+ 4	1	16	+ 4	— 36	— 20	— 12	— 32
1860.	+ 2	— 3	4	9	— 6	— 3	+ 27	+ 15	+ 9
1861.	— 6	— 2	36	4	+ 12	— 4	— 2	+ 18	+ 10
1862.	— 8	+ 1	64	1	— 8	— 6	+ 2	+ 1	— 9
1863.	— 1	+ 3	1	9	— 3	— 24	— 18	+ 6	+ 3
1864.	+ 3	+ 3	9	9	+ 9	— 3	— 21	— 18	+ 6
1865.	+ 7	+ 2	49	4	+ 14	+ 6	— 2	— 16	— 12
1866. . . .	+ 8	— 1	61	1	— 8	— 7	— 3	+ 1	+ 8
1867.	— 2	0	4	0	0	0	0	0	0
1868.	— 6	+ 3	36	9	— 18	— 6	+ 24	+ 21	+ 9
1869.	— 8	— 8	61	64	+ 16	+ 48	+ 16	— 61	— 56
1870.	— 5	— 4	25	16	+ 20	+ 32	+ 24	+ 8	— 32
1871.	+ 1	— 7	1	49	— 7	+ 35	+ 56	+ 42	+ 14
1872.	+ 9	— 1	81	1	— 9	— 1	+ 5	+ 8	+ 6
1873	+ 11	— 3	121	9	— 33	— 27	— 3	+ 15	+ 24
1874.	+ 6	+ 3	36	9	+ 18	+ 33	+ 27	+ 3	— 15
1875	+ 5	— 2	25	4	— 10	— 12	— 22	— 18	— 2
1876.	+ 5	+ 8	25	64	+ 40	+ 40	+ 48	+ 88	+ 72
1877.	— 1	+ 7	1	49	— 7	+ 35	+ 35	+ 42	+ 77
1878.	— 5	+ 6	25	36	— 30	— 6	+ 30	+ 30	+ 36
1879.	— 11	0	121	0	0	0	0	0	0
1880.	— 5	— 3	25	9	+ 15	+ 33	+ 15	+ 3	— 15
1881.	0	— 3	0	9	0	+ 15	+ 33	+ 15	+ 3
1882.	+ 4	0	16	0	0	0	0	0	0
1883.	+ 5	+ 1	25	1	+ 5	+ 4	0	— 5	— 11
1884.	+ 2	+ 5	4	25	+ 10	+ 25	+ 20	0	— 25
1885.	— 4	+ 2	16	4	— 8	+ 4	+ 10	+ 8	0
1886.	— 6	+ 4	36	16	— 21	— 16	+ 8	+ 20	+ 16
1887.	— 5	— 2	25	4	+ 10	+ 12	+ 8	— 4	— 10
1888.	— 5	— 6	25	36	+ 30	+ 30	+ 36	+ 24	— 12
1889.	0	— 3	0	9	0	+ 15	+ 15	+ 18	+ 12
1890.	+ 5	— 9	25	81	— 45	0	+ 45	+ 45	+ 54
1891.	+ 5	+ 6	25	36	+ 30	+ 30	0	— 30	— 30
1892.	+ 2	— 1	4	1	— 2	— 5	— 5	0	+ 5
1893.	— 6	+ 5	36	25	— 30	+ 10	+ 25	+ 25	0
1894.	— 3	— 4	9	16	+ 12	+ 21	— 8	— 20	— 20
1895.	— 5	+ 4	25	16	— 20	— 12	— 24	+ 8	+ 8
1896.	+ 2	— 1	4	1	— 2	+ 5	+ 3	+ 6	— 2
1897.	+ 3	0	9	0	0	0	0	0	0
1898.	+ 4	0	16	0	0	0	0	0	0
1899.	+ 5	+ 4	25	16	+ 20	+ 16	+ 12	+ 8	— 20
1900.	0	— 2	0	4	0	— 10	— 8	— 6	— 3
1901.	— 2	— 3	4	9	+ 6	0	— 15	— 12	— 9
SOMMES .	0	0	1 428	844	+ 83	+ 298	+ 363	+ 341	+ 38

La somme algébrique d'une colonne de produits semblables forme le numérateur du coefficient de dépendance correspondant. Le dénominateur est égal à la moyenne géométrique des carrés des écarts associés dans les produits.

Par exemple, la somme des carrés qui figurent dans la colonne 4 étant égale à 1 428 et la somme des carrés qui figurent dans la colonne 5 étant égale à 844, les sommes de carrés intervenant dans le dénominateur du coefficient de dépendance des mariages avec les naissances survenues un an plus tard seront respectivement 1 428 — 4 = 1 424 et 844 — 25 = 819. On se rend compte de la simplicité des calculs.

Finalement les coefficients sont les suivants :

Mariages d'une année comparés aux naissances survenues :	la même année	$k_0 = \frac{83}{1\,098} = 0{,}076$
	la 1re année suivante	$k_1 = \frac{298}{1\,070} = 0{,}278$
	la 2e —	$k_2 = \frac{363}{1\,063} = 0{,}341$
	la 3e — —	$k_3 = \frac{341}{1\,037} = 0{,}329$
	la 4e — —	$k_4 = \frac{38}{926} = 0{,}041$

Ainsi la répercussion des mariages sur les naissances devient assez sensible lorsque l'on compare les mariages aux naissances survenues deux ou trois ans plus tard et que l'on a éliminé l'influence des changements à longue période.

Ayant été invité par la Commission extraparlementaire de la dépopulation à rechercher l'influence que pouvaient avoir les chômages sur la natalité, il m'a semblé que la comparaison directe de la courbe du chômage et de la courbe de la natalité ne montrait aucun rapport.

Cependant, en Angleterre, la courbe du chômage, que l'on peut construire depuis une cinquantaine d'années à l'aide des renseignements fournis par les *trade-unions*, présente un synchronisme inverse remarquable avec celle de la nuptialité : ce cas remarquable d'antiparallélisme a déjà été soigneusement mis en évidence dans le pays intéressé [1].

Calculé par la méthode de Fechner, l'indice i de dépendance entre ces deux courbes est égal à 0,6. Quant au coefficient k, les chiffres contenus dans les colonnes 13, 9 et 3 du tableau III permettent de le déterminer, toutes réserves étant faites sur l'exactitude des coefficients de chômage.

On trouve

$$k = -\frac{23\,623}{32\,461} = -0{,}73.$$

Ainsi la courbe du chômage et la courbe de la nuptialité dépendent l'une de l'autre par une relation inverse et assez étroite. En général, quand le chômage augmente, la nuptialité diminue ; quand le chômage diminue, les mariages deviennent plus nombreux. Puisque d'autre part nous avons observé une certaine répercussion des mariages sur la natalité, nous pouvons admettre maintenant que les fluctuations du chômage influent sur la natalité : c'est la conclusion que j'ai soumise à la Commission de la dépopulation.

V — Précision du coefficient de dépendance.

Si l'on admet que l'observation du nombre des concordances fournit la proba-

1. Voir notamment un article de M. Wood dans le *Journal of the Royal statistical Society*, numéro de décembre 1899.

bilité p de l'arrivée d'une concordance quelconque, si q est la probabilité d'une discordance, on peut écrire les égalités

$$\begin{aligned} p-q&=k \\ p+q&=1 \end{aligned} \quad \text{d'où} \quad p=\frac{1+k}{2} \qquad q=\frac{1-k}{2}.$$

Par suite l'erreur probable de p a pour expression

$$\pm\frac{0,67}{2}\sqrt{\frac{1-k^2}{n}}$$

n étant le nombre des couples.

VI — Extension de la méthode a la comparaison des courbes de succession et a celle des cartogrammes

La méthode qui vient d'être appliquée à la comparaison des courbes de succession peut être appliquée aussi à la comparaison des courbes de distribution de faits simultanés. Toutefois, dans ce cas, les variations successives des éléments comparés n'offrent pas d'intérêt. Ce qui importe le plus ordinairement, c'est l'allure générale de chaque courbe, en sorte qu'il est plus avantageux de faire porter la comparaison, soit sur des courbes simples interpolées, soit sur des quantités *caractéristiques* déterminées, pour chaque courbe séparément, en fonction des éléments de cette courbe.

On calcule par exemple, pour chaque courbe, soit l'écart moyen, soit la moyenne des carrés des écarts autour de la valeur moyenne des quantités observées (1).

L'application aux représentations par cartogrammes peut procéder de l'un ou de l'autre de ces systèmes et nous allons donner un exemple de calcul du coefficient de dépendance.

Supposons par exemple que l'on se propose de rechercher si les départements de la France où la nuptialité est le plus élevée sont généralement ceux où la natalité est aussi le plus élevée. A cet effet, on construira deux cartogrammes représentant, l'un la distribution des départements suivant le taux de la nuptialité, l'autre leur distribution suivant le taux de la natalité. Ces deux cartogrammes étant construits d'après les règles que nous avons exposées de manière à assurer leur comparabilité, leur juxtaposition donne l'impression que les départements où l'on se marie le plus ne sont pas ceux où la natalité est la plus forte ; on a plutôt l'impression qu'à une forte nuptialité correspond souvent une faible natalité et inversement.

Cette impression peut être précisée au moyen de la détermination de l'indice de dépendance ou du coefficient de dépendance dont nous avons indiqué la formule.

Sur le tableau IV annexé se trouvent réunis les éléments du calcul, d'abord la série des écarts autour de la valeur moyenne tant pour la nuptialité que pour la natalité, puis les produits deux à deux de ces écarts. En notant le nombre des concordances et le nombre des discordances, c'est-à-dire le nombre des produits positifs et le nombre des produits négatifs, on détermine l'indice de dépendance

$$i=-\frac{62-25}{87}=-0,425.$$

1. Par exemple, courbes de distribution des salaires : Voir *Journal de la Société de statistique de Paris,* numeros de juin 1898 et juillet 1902.

Si l'on forme de plus les sommes des carrés des écarts, on obtient les éléments du coefficient de dépendance dont la valeur est

$$k = -0{,}583.$$

i et k sont négatifs ; donc, en général, une forte nuptialité marche de pair avec une faible natalité et inversement.

Lorsque le cartogramme représente des faits ordonnés suivant deux directions principales du plan et se distribuant autour d'une moyenne générale, on peut employer une méthode de comparaison analogue à celle qui est la plus usitée lorsqu'il s'agit des courbes de distribution, c'est-à-dire que l'on peut se servir d'une *caractéristique* de la distribution des faits dans le plan, caractéristique dont la forme est encore celle du coefficient de dépendance dont nous nous sommes servi.

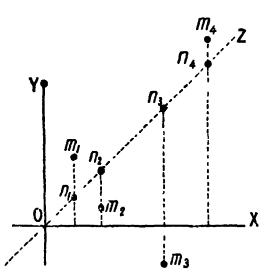
Tel est le cas des tables à double entrée, par exemple de la table des couples mariés suivant l'âge des époux [1]. Supposons cette table construite de la façon suivante. On prend deux axes rectangulaires Ox, Oy issus d'un centre O correspondant à l'âge moyen des époux et à l'âge moyen des épouses. Sur l'un des axes Ox on porte les âges des époux, comptés à partir de leur âge moyen.

Pour un couple donné, on obtient ainsi un point de Ox correspondant à l'âge du mari. En ce point on trace une parallèle à Oy de longueur correspondant à l'âge de la femme compté à partir de l'âge moyen des femmes. L'extrémité de cette ligne représente le couple observé.

Si l'on suppose le temps divisé en intervalles infiniment petits, on peut admettre qu'il n'existe ni deux époux ni deux épouses exactement de même âge, ce qui permet de représenter chaque couple par un point différent du plan.

Considérons maintenant la figure formée par les différents points représentatifs m_1, m_2, etc., des couples mariés observés et par leurs distances à Ox ; nous avons appelé cette figure une courbe statistique, c'est la courbe des âges des femmes suivant les âges de leurs maris.

Menons par le centre O une droite OZ inclinée à 45° sur Ox. Cette ligne peut être regardée comme représentant une distribution théorique de couples mariés pour lesquels les âges des femmes seraient constamment égaux à ceux de leurs maris. Elle coupe les parallèles à Oy menées par m_1, m_2, m_3, etc., en des points n_1, n_2, n_3, etc., dont les distances à Ox sont égales respectivement aux écartements de ces parallèles à partir de Oy, c'est-à-dire qu'elles représentent les âges des maris, tandis que les distances à Ox des points m_1, m_2, etc., représentent les âges des femmes respectives de ces maris.



D'après ce que nous avons vu précédemment, le coefficient de dépendance de la courbe des points m_1, m_2, m_3, etc., et de la ligne O Z des points n_1, n_2, etc., a pour expression

$$k = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \, \Sigma y^2}}.$$

1. Voir *Journal de la Société de statistique de Paris*, numéro de janvier 1904, page 27.

Si l'on prend, pour unité de mesure des x, $\sqrt{\Sigma x^2}$, et, pour unité de mesure des y, $\sqrt{\Sigma y^2}$, cette expression prend la forme simple

$$k = \Sigma xy = \frac{1}{2}[\Sigma x^2 + \Sigma y^2 - \Sigma(x-y)^2] = \frac{2 - \Sigma(x-y)^2}{2}.$$

De là résulte que l'on peut écrire

$$\Sigma(x-y)^2 = 2(1-k).$$

Si k est égal à 1, tous les x sont nécessairement égaux aux y correspondants, c'est-à-dire que les âges des maris sont égaux aux âges des femmes, au facteur constant près $\frac{\Sigma x^2}{\Sigma y^2}$. Un lien invariable unit l'âge du mari et l'âge de la femme.

A mesure que k diminue, le lien entre l'âge du mari et l'âge de la femme se relâche; lorsque $k = 0$, $\Sigma xy = 0$, c'est-à-dire que, en considérant par exemple des couples dont le mari est plus âgé que la moyenne (x positif), la femme est d'âge tantôt supérieur, tantôt inférieur à la moyenne; on peut donc dire que l'âge de la femme est indépendant de celui du mari. Si k devenait négatif, c'est que, généralement, des hommes d'âge supérieur à la moyenne auraient des femmes d'âge inférieur à la moyenne, et inversement : en fait, k est positif et voisin de 1; quand l'âge du mari augmente, en général, l'âge de la femme augmente pareillement.

Le coefficient k peut donc être regardé comme une caractéristique de la distribution représentée par le cartogramme tabulaire. Il indique la dépendance, d'une part, de la distribution des faits observés dans une direction, et d'autre part, de la distribution des mêmes faits dans une autre direction. A cet égard, il peut servir de terme de comparaison de diverses distributions dans le plan.

C'est pourquoi ce coefficient intervient dans la théorie des erreurs de situation d'un point dans un plan donnée par Bravais en 1837; c'est pourquoi il a été introduit par Galton et Pearson dans l'étude de l'hérédité des caractères biologiques. Karl Pearson, qui a donné une théorie générale des distributions statistiques, a indiqué (1) comme expression de l'erreur probable de k, dans le cas de distribution normale, la fraction

$$0{,}67\frac{1-k^2}{\sqrt{n(1+k^2)}},$$

n étant le nombre des couples.

Lorsque l'on a mesuré les x avec l'unité $\sqrt{\Sigma x^2}$ et les y avec l'unité $\sqrt{\Sigma y^2}$, la ligne droite $y = kx$ indique la direction dans laquelle les points représentatifs sont le plus denses; elle s'identifie avec ce que l'on appelle en dynamique l'un des axes principaux d'inertie.

(*A suivre.*) Lucien MARCH.

1. *Philos. trans. of the Royal Society*, vol. 187 A, p. 266.