

STATISTIQUE ET ANALYSE DES DONNÉES

JEAN-JACQUES DROESBEKE

PHILIPPE TASSI

**George Udny Yule ou comment (ne pas)
parler de corrélation**

Statistique et analyse des données, tome 15, n° 1 (1990), p. 25-43

http://www.numdam.org/item?id=SAD_1990__15_1_25_0

© Association pour la statistique et ses utilisations, 1990, tous droits réservés.

L'accès aux archives de la revue « Statistique et analyse des données » 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/>

GEORGE UDNY YULE

OU

COMMENT (NE PAS) PARLER DE CORRELATION

Jean-Jacques DROESBEKE

Philippe TASSI

Université Libre de Bruxelles
Avenue F. Roosevelt, 50
B - 1050 Bruxelles
BELGIQUE

Médiamétrie
Rue Anatole France, 55/63
92532 Levallois-Perret Cedex
FRANCE

Résumé

Cet article a pour but de montrer la part prise par G.U. Yule dans l'étude de la corrélation et celle des séries chronologiques. Il traite aussi de la controverse entre Yule et Pearson.

Mots-clefs : histoire de la statistique, corrélation, corrélogramme, modèle autorégressif.

Classification AMS : 6203 - 01 A 60

Abstract

In this paper, the authors present the work of G.U. Yule in the study of correlation and time series. The controversy between Yule and Pearson is also concerned.

Keywords : history of statistics, correlation, correlogram, autoregressive model.

*Nonne hoc mirificum ?
Sicut sanctum templa
Attrahunt statisticum
Parvula exempla*

G.U. Yule.

1. INTRODUCTION

Dans cet article, nous voulons atteindre un double objectif. Le premier, prépondérant, consiste à souligner l'importance accordée par Yule au concept de corrélation et les conséquences de ce fait, à la fois sur son œuvre et sur ses relations avec d'autres statisticiens, tout particulièrement Karl Pearson. Le second doit nous permettre de répondre à un souhait formulé par ailleurs, mais non réalisé. Dans notre *Histoire de la Statistique* [3], nous avons consacré le chapitre 9 à quelques "grands noms". En raison des contraintes liées au nombre de pages imposées par l'éditeur, nous avons restreint cette liste à huit personnages. Nous aurions pu y inclure le nom de Yule. Cet article est une façon de lui rendre hommage.

2. QUI EST GEORGE UDNY YULE ?

Il naît le 18 février 1871 à Beech Hill, près de Haddington, à l'est de l'Ecosse, au sein d'une famille qui avait fait sa place dans la société intellectuelle britannique sous le double sceau de la littérature et de l'administration. Ses études, entamées à Winchester, s'achèvent par un diplôme d'ingénieur délivré en 1892 par le University College de Londres. Attiré par la physique expérimentale, il séjourne un an à Bonn, se consacrant à l'étude des ondes électriques sous la direction de Heinrich Hertz, ce qui lui permet de rédiger ses premiers articles scientifiques sur des sujets liés à l'électromagnétisme. Mais ni son diplôme d'ingénieur, ni son passage à Bonn ne vont être déterminants. Dès son retour à Londres, en 1893, il accepte une place de "démonstrateur" que lui offre Karl Pearson, alors professeur de mathématiques appliquées au University College. Très vite, il va s'impliquer dans les travaux dirigés par Pearson. A cette époque, ce dernier est particulièrement intéressé par les courbes de fréquence et la corrélation. Le deuxième sujet passionne Yule qui rédige son premier article à caractère statistique : *On the correlation of total pauperism with proportion of out-relief, I : all ages* [31]. Sa

collaboration avec Karl Pearson est bénéfique pour l'ensemble du laboratoire, de par ses qualités pédagogiques et son attirance vers les sciences sociales et économiques qui font de Yule un complément privilégié de Pearson, plus axé sur la biologie et les sciences expérimentales. Et pourtant ces intérêts distincts les amèneront à s'affronter (par articles interposés) quelques années plus tard. Nous évoquerons ce problème dans la suite.

Tenté par des appointements plus consistants, Yule quitte le University College en 1899, pour le Département de Technologie du *City and Guilds of London Institute*. Mais ce départ est apparent car, de 1902 à 1909, il donne dans l'établissement qu'il vient de quitter des conférences dont le contenu sera rassemblé sous la forme d'un ouvrage qui va devenir une référence pour beaucoup : *An introduction to the theory of statistics*. Le succès de ce livre est souligné par les 14 éditions qu'il connaîtra, un complément significatif étant apporté par M.G. Kendall, co-auteur à partir de la onzième édition (1937).

1912 voit débiter une nouvelle tranche de la vie de Yule avec son entrée à l'Université de Cambridge. La période mouvementée de la première guerre mondiale ne l'empêche pas de présenter, en compagnie du major Greenwood, des travaux importants dans le domaine de la statistique médicale. De cette rencontre entre notre statisticien et cet épidémiologiste va naître une amitié profonde. Parmi leurs travaux communs, il faut relever l'article publié en 1917, intitulé *On the statistical study of some bacteriological methods used in water analysis* [10]. On y trouve les fondements théoriques de la distribution binomiale négative. Cet article contient aussi les prémices d'une formalisation de ce qui sera appelé plus tard processus stochastique.

La troisième période de sa carrière couvre les années 1920-1930. C'est la plus prolifique. Elle voit éclore les travaux de Yule dans le domaine des séries chronologiques (corrélogramme, processus autorégressifs). Elle voit aussi confier à notre personnage la présidence de la Royal Statistical Society (1924-1926).

Des problèmes de santé vont malheureusement ternir son existence entre 1931 et 1938. La refonte de son "*Introduction...*" en compagnie de M.G. Kendall en 1937 lui redonnera le goût de la publication, ce qui le conduira en 1944, à produire un dernier ouvrage intitulé *The statistics of literary vocabulary* (Cambridge University Press).

Mais sa santé continue à décliner. Il s'éteint à Cambridge, le 26 juin 1951, victime d'une crise cardiaque.

3. YULE ET LA CORRELATION

Afin de montrer comment et dans quelles conditions Yule s'est intéressé à la corrélation, il nous semble utile de situer historiquement ce concept.

Il est communément admis que la corrélation est née de la réflexion de Francis Galton. Mais plusieurs historiens des sciences ont posé la question de savoir si ce concept n'était pas apparu antérieurement, en particulier dans le contexte de la loi normale en tant que loi des erreurs (nous utiliserons le terme "normale", bien qu'il ne fût pas de mise au début du 19^{ème} siècle). Dans le cas particulier d'une distribution bivariée, cette loi se définit par la fonction de densité présentée ci-dessous, où nous avons pris pour x et y les variables centrées $X - \mu_1$ et $Y - \mu_2$:

$$f(x, y) = \frac{1}{2 \sigma_x \sigma_y \pi \sqrt{1 - \rho^2}} e^{-\frac{1}{(1 - \rho^2)} \left[\frac{x^2}{\sigma_x^2} - \frac{2\rho xy}{\sigma_x \sigma_y} + \frac{y^2}{\sigma_y^2} \right]}$$

Dans cette expression, le coefficient de corrélation apparaît, à la constante de normalisation près, dans le deuxième terme de l'exposant. Cette loi donna lieu aux travaux bien connus de Laplace et de Gauss¹. Le premier, dans un article sur les intégrales définies [16], étudia le comportement et les propriétés de

$$\iint \frac{k}{4 k'' \pi} \frac{1}{\sqrt{E}} \frac{du du'}{a^2} e^{-\frac{k(Fu^2 + 2Guu' + Hu'^2)}{4k'aE}}$$

Mais il n'accorda pas une attention particulière à l'interprétation du double produit intervenant dans l'exposant. Une conclusion analogue est à tirer des travaux de Gauss dans sa *Theoria Combinationis Observationum Erroribus Minimis Obnoxiae* (1821 et 1823) et son *supplément* de 1826 [9].

Un autre personnage allait alimenter la discussion. Il s'agit de Auguste Bravais (1811-1863), célèbre par ses nombreux mémoires dans les domaines de l'astronomie, la géodésie et la physique. Ses contacts avec la théorie des probabilités sont rares, à l'exception d'une correspondance échangée avec Adolphe Quételet et d'un article intitulé *Analyse mathématique sur les probabilités des erreurs de situation d'un point* [1]. Dans

¹ Le lecteur intéressé par la vie de ces auteurs peuvent se reporter à [3], [14], [15] [29] et [30].

cet article, Bravais développe une approche géométrique permettant de connaître les ordonnées d'une surface de densité "normale" :

$$z = \frac{K}{\pi} e^{-\left(ax^2 + 2cxy + by^2\right)}$$

Bravais souligne le fait que l'équation

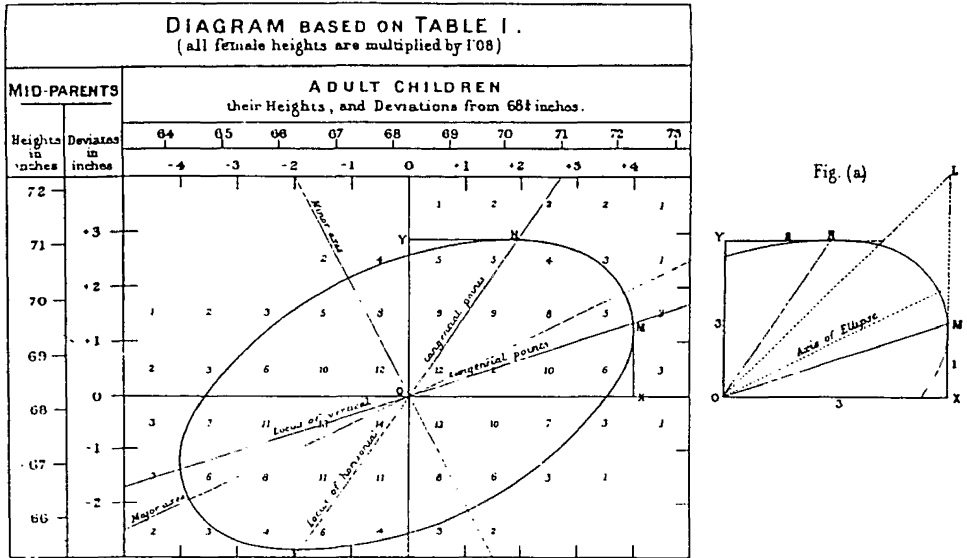
$$ax^2 + 2cxy + by^2 = k$$

représente une ellipse dont les axes principaux sont distincts des axes coordonnés. Il cite même le terme de *corrélacion* dans son texte : "*La co-existence des mêmes variables [...] dans les équations simultanées en x et y amène une corrélacion que les modules [...] cessent de représenter la possibilité des valeurs simultanées de (x, y) sous le vrai point de vue de la question* (p. 263)".

Bravais s'intéresse cependant davantage aux surfaces des ellipses qu'aux relations pouvant exister entre les variables. Ici aussi, la corrélacion est implicite, ce qui fera dire à K. Pearson que Bravais est le père de ce concept (jusqu'à ce qu'il change d'avis dans la suite !).

Le rôle de Francis Galton va être décisif. Nous ne reviendrons pas sur la vie de ce personnage (cf., par exemple, [3], [29] et [30]). Influencé par l'*origine des espèces* que publie en 1859 son cousin Charles Darwin, Galton s'intéresse aux problèmes d'hérédité, encouragé par ses travaux d'anthropologie et ses voyages en Afrique. Il en résulte des ouvrages importants : *Hereditary genius* publié en 1869 et, surtout, *Natural inheritance* qui date de 1889. C'est dans ce dernier ouvrage qu'on peut apprécier le concept de régression introduit par Galton. Mais déjà en 1877, dans des articles intitulés *Typical laws of heredity* et publié à la fois dans *Nature* et dans les *Proceedings of the Royal Institution of Great Britain* [5], Galton émet des conclusions importantes. S'intéressant à la reproduction de pois (sweet peas) il souligne que "*the processus of heredity must work harmoniously with the law of deviation and be themselves in some sense conformable to it* (p. 512)". Se basant sur les groupes distincts qu'il avait constitués, il constate que les sept groupes de moyennes sont distribués normalement, mais pas de la même façon que leurs parents. Plus précisément, les distributions se répartissent autour de valeurs plus proches du poids moyen de la population que ce n'était le cas pour leurs parents. Il qualifie ce phénomène de *réversion* et choisit la loi de réversion la plus simple : celle d'une droite. En 1885, ayant abandonné le terme de réversion pour celui

de *régression*, il étudie une table - devenue un classique en la matière - portant sur les tailles de 928 couples (pères, enfants) - les seconds ayant atteint l'âge adulte -, ce qui le conduit à son célèbre graphique [6].



En 1888, alors que *Natural Inheritance* est sous presse, Galton présente un court article à la Royal Society de Londres intitulé *Co-relations and their measurement, chiefly from anthropometric data* [7]. Il y note que le problème de régression peut être vu de façon symétrique, ce qui conduit à la construction de deux droites de régression. Mais, ensuite, si les mêmes échelles statistiques (en *unités de leur erreurs probables*) sont utilisées pour les mesures, les deux droites ont la même pente r . Ce nombre exprime aussi l'intensité de leur co-relation ("closeness of co-relation").

Qui connaît la vie de Karl Pearson sait que la statistique ne fut pas son premier amour. Son existence fut d'abord partagée entre l'histoire, la philosophie, la physique, ... jusqu'en 1884, date à laquelle il entre au University College de Londres, en tant que professeur de mathématiques appliquées et de mécanique. Il est alors âgé de 27 ans. C'est au sein d'un Club mixte, participant à des débats portant sur le socialisme, le rôle de la femme, ... qu'il prend connaissance des œuvres de Galton. Mais les résultats obtenus par ce dernier ne le marqueront pas tout de suite. Il n'y fait même pas référence dans sa *Grammar of Science* [17], publiée en 1892. Et pourtant, cette même année, Pearson commence un cycle de conférences sur les applications des probabilités et dès 1893 publie

des articles de statistique à la Royal Society de Londres. Deux hommes ont influencé cette nouvelle passion : son ami, le zoologiste et biologiste Weldon - également professeur au University College - et Francis Ysidro Edgeworth, cousin éloigné de Galton.

Rendu célèbre par son ouvrage intitulé *Mathematical Psychics : an essay on the application on mathematics to the moral sciences* et publié en 1881 alors qu'il était âgé de 36 ans, Edgeworth se met à rédiger des articles à caractère statistique à partir de 1883, influencé en particulier par les travaux de Quetelet et de Galton. Il est certain que l'attirance de Pearson vers la statistique s'est déroulée sous l'influence de Edgeworth, avec lequel il a échangé une correspondance aujourd'hui disparue. Mais, bien vite, le torchon allait brûler entre ces deux hommes à propos des courbes de fréquence (de densité), en raison de la compétition qui se développe entre eux sur ce thème.

C'est dans ce contexte que G.U. Yule arrive au laboratoire de K. Pearson. Il joue rapidement un rôle dans le conflit opposant Pearson à Edgeworth puisqu'il présente les arguments de son "patron" dans son premier article sur le paupérisme.

Mais revenons à la corrélation. On sait que la contribution première de Pearson à ce thème est contenue dans les articles intitulés *Contributions to the mathematical theory of evolution* adressés entre 1894 et 1896 à la Royal Society ([18], [19], [20]). A cette occasion, il propose de prendre comme "meilleure valeur" du coefficient de corrélation l'expression

$$S(x y) / N \sigma_x \sigma_y$$

où $S(x y)$ est la somme des produits des N déviations aux moyennes en x et y . Il obtient en outre l'erreur probable de r et propose deux formules permettant de calculer ce coefficient. Ces résultats sont cependant obtenus sous l'hypothèse de normalité des déviations. En janvier 1896, Yule suggère à Pearson que l'hypothèse de normalité peut être affaiblie en ne préconisant que la linéarité de la régression. Pearson ne peut accepter le point de vue de Yule arguant notamment du fait que la vision de ce dernier ne se conçoit pas en biologie. Cette différence de conception va s'accroître, comme nous le verrons ci-dessous.

4. LA CONTROVERSE YULE - PEARSON

Les travaux de Pearson et de Yule sur la corrélation se poursuivent séparément, mais sur des sujets communs. Ainsi, Pearson introduit les concepts de corrélation multiple et partielle, toujours dans le cadre de distributions normales. Yule fait de même, sans recourir à l'hypothèse de normalité et en utilisant parfois des termes distincts (ainsi, Yule utilise le *net coefficient of correlation* là où Pearson parle de *partial correlation*).

Par ailleurs, des mesures de corrélation autres que le coefficient r vont engendrer des travaux intéressants. Ainsi, en 1900, s'inspirant d'une "fonction de Quetelet" établie en 1832 dans une lettre adressée par le statisticien belge à M. Villermé, de l'Institut de France, "sur la possibilité de mesurer l'influence des causes qui modifient les éléments sociaux", Yule propose [32] un *coefficient d'association*, noté Q en hommage à Quetelet et défini pour une table de contingence 2×2 :

	1	2
1	P_{11}	P_{12}
2	P_{21}	P_{22}

$$Q = \frac{P_{11} P_{22} - P_{12} P_{21}}{P_{11} P_{22} + P_{12} P_{21}}$$

La même année, Pearson présente le test d'ajustement du χ^2 [21] qui allait le conduire en 1904 au coefficient de contingence [24]. En 1901, il publie un article [22] contenant notamment les bases de ce qui deviendra le *coefficient de corrélation tétrachorique*, se référant à des résultats obtenus par Sheppard en 1899 [27]. Il propose aussi quatre expressions notées Q_1, Q_3, Q_4 et Q_5 destinées à fournir des alternatives au coefficient d'association Q_2 de Yule, ajoutant perfidement "... notwithstanding the extreme elegance and simplicity of Mr. Yule's coefficient of association Q_2 , the coefficients Q_3, Q_4 and Q_5 , which also satisfy his requirements, are much nearer to the value assumed by the correlation."

Une suite d'articles importants sur le problème de la corrélation allaient suivre parmi lesquels nous mentionnons ceux rédigés par Yule, en 1903 [33], Pearson [23] et Spearman [28] - qui propose "son" coefficient de corrélation de rang - en 1904, Yule en 1906 [34] et la réplique de Pearson en 1907 [25], ...

Le ton allait se durcir l'année où Yule s'établit à Cambridge. En mai 1912, paraît en effet un article de Yule dans le *Journal of Royal Statistical Society*, intitulé *On the methods of measuring association between two attributes* [35]. Cet article a pour ambition de donner un aperçu des principales mesures d'associations d'une table de

contingence, introduisant même à cette occasion un *coefficient de colligation*. Ce dernier est défini pour la table de contingence 2 x 2 présentée plus haut :

$$Y = \frac{\sqrt{P_{11} P_{22}} - \sqrt{P_{12} P_{21}}}{\sqrt{P_{11} P_{22}} + \sqrt{P_{12} P_{21}}}$$

A cette occasion, Yule revient sur l'hypothèse de normalité sous-jacente dans les travaux de Pearson et souligne les limitations de cette approche. Il recourt dans ce but à plusieurs exemples concrets. Il en profite pour répondre aux critiques formulées à l'égard de son oeuvre par D. Heron, élève de Pearson, dans un article publié en 1911 [12]. Celui-ci lui reprochait, entre autres, d'avoir *"failed to apply the most obvious test of their validity, i.e., to compare the results obtained to the two methods when applied to the same data"*.

Yule dénie tout fondement à cette critique, arguant du fait que, par essence même, les deux coefficients possèdent des propriétés différentes. Il conteste le qualificatif de test attribué à la procédure de Heron, et montre que ce soi-disant *"test de validité"* conduit à des aberrations : *"The geometric mean is therefore invalid because it differs from the arithmetic mean, and the arithmetic mean, I suppose, is invalid because it differs from the geometric mean. The median is often, indeed usually, different from both, and therefore it is also invalid - and so on for the mode, the harmonic mean, and every other conceivable form of average."*

Tentant de défendre son point de vue avec fermeté mais aussi retenue, Yule reconnaît la remarquable qualité mathématique du travail de Pearson, mais ses résultats, sur le plan de l'interprétation statistique, dépendent totalement des hypothèses faites. *"Those assumptions were never adequately tested, even in cases in which such a test was called for and easy of application, and the few tests which I applied some years since in connection with other work sufficed to show that they were, to say the least, of exceedingly doubtful validity."*

Dès l'introduction de leur article intitulé *On theories of association* [26], Heron et Pearson s'insurgent contre le ton, jugé par eux inexplicable et inhabituel, utilisé par Yule. Ils affirment également que, si l'optique de Yule est acceptée par la classe statistique, elle causera un dommage irréparable au développement de la théorie statistique moderne. Le long extrait qui suit nous semble indispensable pour dépeindre le ton de l'article de Heron et Pearson :

"Mr Yule has invented a series of statistical methods which are in no case based on a reasoned theory, but which possess the dangerous fascination of very easy and ready application, and therefore are at once seized upon as applicable to all sorts of problems by those who are without adequate training in statistical theory, or without the mathematical

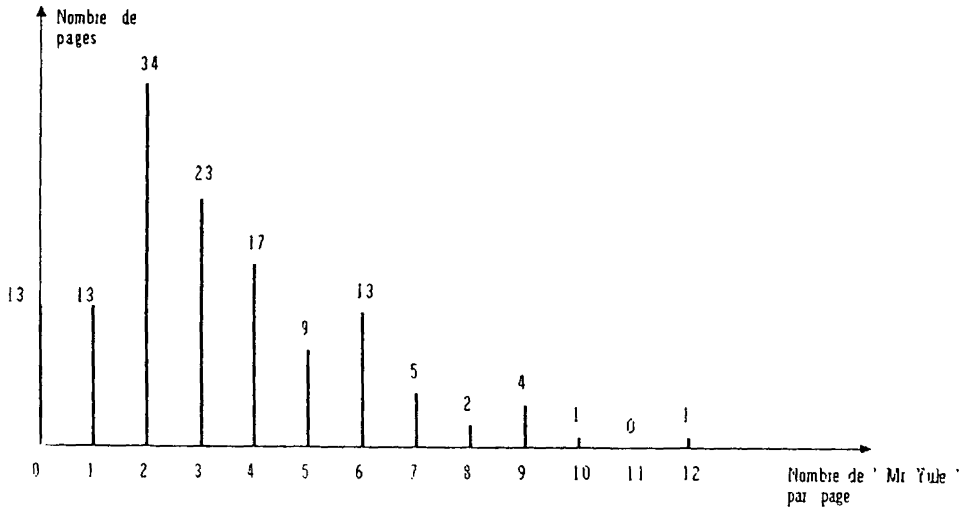
knowledge requisite to weigh cautiously their logical basis. The methods to which we refer are these :

- (i) the use of the so-called "coefficient of association" to measure the relationship of two attributes;*
- (ii) the use of a new coefficient, which Mr Yule terms a "coefficient of colligation", apparently to be used in like cases with the coefficient of association;*
- (iii) the development of a method which first appeared in a paper by the late Mr John Gray ; in this method each group of a contingency table is considered as a cell of unit subrange for both variates. This assumption being made, Mr Yule calculates the coefficient of correlation by the product-moment method, and on the basis of this procedure terms his coefficient the coefficient of correlation and uses the customary letter r for it.*

Such a terminology is absolutely unjustifiable and can only confuse the uninstructed and undiscerning reader."

Pendant 143 pages, Pearson et Heron font un procès virulent à Yule. Ce texte, riche en références et en analyses, souligne sans contexte l'érudition et l'habileté de ces auteurs à rechercher tous les moyens pour tenter d'établir leur droit mais démontre aussi une hargne peu digne d'un grand scientifique. Usant à la fois d'arguments pertinents et d'autres quelque peu tendancieux, ils établissent un panorama élogieux pour leurs travaux et injurieux pour Yule. Il est vrai que ce dernier n'était pas "neutre" dans son article de 1912, mais il s'agissait (déjà) d'une réponse à Heron.

A titre anecdotique, nous nous sommes penchés sur le nombre de fois que l'on trouve dans cet article une phrase contenant une expression faisant allusion aux "travaux" de "Mr Yule ..." - suivi souvent d'un verbe ou d'un objectif peu flatteur - en ne considérant pas les pages qui ne contiennent que des tableaux ou des graphiques. Le résultat de cette petite analyse est édifiant, ainsi qu'on peut le constater aisément sur le graphique suivant :



“Mr Yule ...!” est cité en moyenne plus de 3 fois par page et celles qui ne contiennent pas de référence à notre homme contiennent peu de phrases. Il est en outre significatif que sur les dix premières pages de l’article, le *Mr Yule* apparaît déjà 71 fois.

Nous ne voulons pas nier le fait que certains arguments de Pearson et Heron étaient pertinents. Il n’en reste pas moins vrai qu’un texte où l’on emploie des expressions agressives du type *Mr Yule never states..., the fictitious character of Mr Yule’s idea..., the fallacy of Mr Yule’s selection..., It is a pity Mr Yule has not the courage of his opinion..., Mr Yule erroneously terms..., With his usual ingenuity..., Mr Yule obscurely seems to indicate..., Mr Yule’s failure...*, ne constitue pas une référence sur les qualités humaines de Pearson.

Mais nous savons que l’orgueil de ce dernier l’amènera à se brouiller avec d’autres statisticiens (nous avons déjà parlé d’Edgeworth, pensons aussi à Fisher). Comme Yule était, reprenant l’appréciation de M.G. Kendall, *one of the ablest, kindest and most lovable of men*, on peut imaginer le choc ressenti lors de la parution de cet article.

Nous savons que Yule ne publiera pratiquement plus de travaux sur le sujet par la suite, hormis concernant l’usage de la corrélation dans l’étude des séries chronologiques, comme nous le verrons ci-dessous. On peut supposer que les commentaires de Pearson à son égard ne sont pas étrangers à cette situation.

5. YULE ET LES SERIES CHRONOLOGIQUES

Après la première guerre mondiale, Yule va entrer dans une nouvelle phase d'intérêt consacrée aux séries chronologiques. Trois articles témoignent essentiellement de ses qualités en lui permettant d'introduire le *corrélogramme*, d'une part, les fondements des *processus autorégressifs* d'autre part. Le premier, intitulé *On the time-correlation problem, with special reference to the variate-difference correlation method* [36], date de 1921. Il permet à Yule de faire une synthèse des travaux existant dans l'analyse des composantes d'une série chronologique sur base des travaux de Poynting, Hooker, March, "Student", Anderson, Pearson et Persons. Comme il le souligne dans sa conclusion, "*the problem of time-correlation is then, in my view, the problem of isolating, for the purpose of discussing the relations between them, oscillations of different durations - such oscillations being in all probability, not strictly periodic but up-and-down movements of greater or less rapidity*".

Le deuxième article de Yule a été rédigé en 1926. Intitulé *Why do we sometimes get non-sense-correlations between time-series ? - A study in sampling and the nature of time-series* [37], ce travail est digne d'intérêt, tant par sa clarté d'expression que par l'importance des problèmes abordés.

Yule introduit la notion de "*nonsense correlations*" à partir de l'exemple du taux de mortalité en Angleterre et au pays de Galles et de la proportion des mariages selon le rite anglican (cf. graphique ci-après).

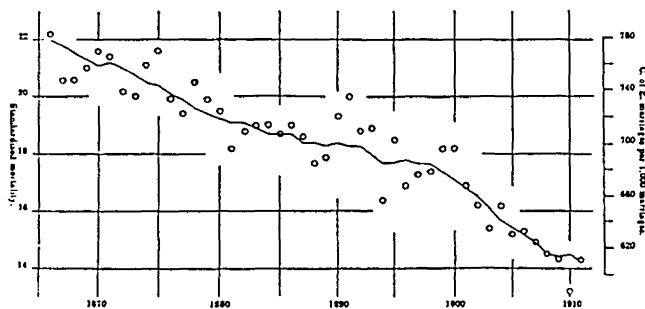
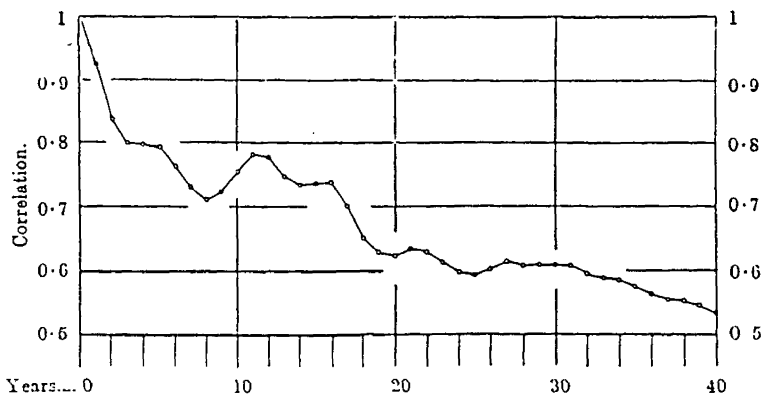


FIG. 1.—Correlation between standardized mortality per 1,000 persons in England and Wales (circles), and the proportion of Church of England marriages per 1,000 of all marriages (line), 1888–1911. $r = +0.6512$.

Argumentant simplement, Yule conclut à l'absurdité d'une quelconque relation de causalité, même indirecte, entre ces deux variables.

Nous reproduisons ci-après un paragraphe entier de l'article de 1926 dans lequel la clarté et la logique du raisonnement de Yule apparaissent : *“Starting from a question that may have seemed to some silly and unnecessary, we were led to investigate the correlations between samples of two simple mathematical functions of the time. It appeared that small samples (in time) of such functions tended to give us correlations departing as far as possible from the truth, the correlations tending to approach ± 1 if the time for which we had experience was very small compared with the time necessary to give the true correlation. Asking ourselves, then, what types of statistical series might be expected to give results analogous to those given by the mathematical function considered, we were led to a classification of series by their serial correlations $r_1, r_2, r_3, \dots, r_k, r_k$ being the correlation between terms s and $s + k$. The important matter in classification was the form of function relating r_k to k , which indicated the nature of the serial correlations between differences of the time-series. If this function is linear, the time-series has random differences; if it gives a graph concave downwards the difference correlations are positive. We concluded that it was series of the latter type (positively correlated series with positively correlated differences, or conjunct series with conjunct differences to use my suggested term) that formed the dangerous class of series, correlations between short samples tending towards unity. Experimental investigation completely confirmed this suggestion. Samples from conjunct series with random differences gave a widely dispersed distribution of correlations, samples from conjunct series with conjunct differences gave a completely U-shaped distribution, with over one-third of the correlations exceeding ± 0.9 .”*

Yule a ainsi l'occasion d'introduire le corrélogramme construit à partir de la célèbre série de l'indice du prix du blé de Beveridge, tous deux devenus des “classiques” en leur domaine.



Enfin, en 1927, paraît *On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers* [38]. Là aussi, Yule fonde ses développements sur une série qui deviendra un autre "classique" dans l'étude des séries chronologiques, celle du nombre de taches solaires de Wolfer.

Yule propose de considérer cette série comme analogue à une suite d'observations obtenues à partir d'un mouvement périodique perturbé, tel un pendule soumis à une succession de petits chocs aléatoires. Il constate que le graphe d'un tel mouvement ressemble fortement à celui de la série de Wolfer, et suggère que, dans une situation semblable, l'application du "périodogramme" fournit des résultats peu fiables. *"The problem of determining the period and the disturbances, in the case of the sunspot numbers, was attacked in the first instance by finding the best (least square) linear equation relating $u_x + u_{x-2}$ to u_{x-1} , this giving the form of difference equation required for a simple harmonic function. The equation gave a period which was obviously too low. It is suggested that this results is due to the presence of superposed fluctuations in addition to disturbances, a suggestion borne out by applying the same method to graduated values of the numbers. This yielded a much closer approximation to the period suggested by the graph"*.

Yule applique ensuite la même méthode pour rechercher la présence éventuelle d'une autre période superposée à la fondamentale. Son analyse l'amène alors à énoncer : *"The better and more general method was then applied of determining the regression equation for u_x on u_{x-1} and u_{x-2} , and solving this as a finite difference equation. The solution is a rapidly damped harmonic function ... The correlations equally fail to suggest the presence of any period other than the fundamental."*

Toute la démarche de Yule n'est pas parfaite (il en critiquera d'ailleurs certains aspects lui-même dans la suite), mais nous y trouvons les bases de la théorie qui conduira l'étude des séries chronologiques à l'utilisation des modèles linéaires stochastiques de type autorégressifs.

Le dernier article de Yule portant sur les outils d'analyse des séries temporelles date de 1945 : *On a method of studying time-series based on their internal correlations* [39]. Il lui permet, dans l'étude de la variance des moyennes de n observations consécutives, d'introduire un nouveau diagramme devenu le *Lambdagramme*. Celui-ci aura moins de succès que le premier !

CONCLUSION

Yule a, sans contestation possible, participé activement à l'essor de la statistique. Ne lui doit-on pas des résultats importants pour la compréhension de la notion de corrélation et une optique - probabiliste - d'analyse et de modélisation des séries chronologiques qui va s'avérer fructueuse dans les décennies suivantes ? Ses travaux se sont placés dans un contexte d'ouverture et de réserve. Ne disait-il pas lui-même, commentant son article de 1945, "*I do not put forward the methods used in the present paper as necessarily the best, nor even in all cases applicable. I was attacking a problem which, to me at least, was a new one, and used the methods that seemed best at the moment; but experience may suggest better methods*".

Autant que l'on puisse en juger par ses écrits, sa personnalité, en tant qu'homme, fût aussi attachante. Il ne fait aucun doute que Yule fut marqué par sa querelle avec Pearson, son ancien maître, et il est probable qu'il cessa ses travaux sur la corrélation pour ne pas entretenir un inutile combat. Il est révélateur, à ce propos, que toute animosité éventuelle disparut. Ainsi, la mort de Pearson en 1938 l'affecta-t-elle profondément. On peut le constater dans les éloges officiels qu'il fit de ce dernier ou encore dans ses lettres comme celle dont l'extrait suivant est cité par M.G. Kendall : "*I feel as though the Karlovingian era has come to an end, and the Piscatorial era which succeeds it is one in which I can play no part*" [13].

Une telle attitude n'est pas si fréquente. Les deux autres grands noms de la statistique que sont Fisher et Neyman ne l'ont jamais eue.

REFERENCES

- [1] Bravais, A. (1846), Analyse mathématique sur les probabilités des erreurs de situation d'un point, *Mémoires de l'Institut de France*, 9, pp. 255-332.
- [2] Droesbeke, J.J., Fichet, B. et Tassi, Ph. Eds. (1989), *Séries chronologiques - Théorie et pratique des modèles ARIMA*, Economica, Paris.
- [3] Droesbeke, J.J. et Tassi, Ph. (1990), *Histoire de la statistique*, Presses Universitaires de France, Coll. Que Sais-je ? (à paraître) Paris.
- [4] Galton, F. (1869), *Hereditary Genius*, Macmillan, London.

- [5] Galton, F. (1877), Typical laws of heredity, *Natura*, 15, pp.492-495, 512-514, 532-533 (publié aussi dans les *Proceedings of the Royal Institution of Great Britain*, 8, pp. 982-301).
- [6] Galton, F. (1885), Presidential Adress before the Anthropological Section of the British Association for the Advancement of Science, *British Association Reports for 1885*.
- [7] Galton, F. (1888), Co-relations and their measurement, chiefly from anthropometric data, *Proceedings of the Royal Society of London*, 45, pp. 135-145.
- [8] Galton, F. (1889), *Natural Inheritance*, Macmillan, London.
- [9] Gauss, C.F. (1821, 1823 et 1826), Theoria Combinationis Observationum Erroribus Minimis Obnoxiae, Part. 1, 2 et suppl., *Werke*, pp. 1-108.
- [10] Greenwood, Major and Yule, G.U. (1917), On the statistical study of some bacteriological methods used in water analysis, *Journal of Hygiene*, 16, pp. 36-54.
- [11] Harris, J.A. (1912) On the calculation of intra-class and inter-class coefficients of correlation from class moments when the number of possible combinations is large, *Biometrika*, 9, pp. 446-472.
- [12] Heron, D. (1911), The danger of certain formula suggested as substitutes for the correlation coefficient, *Biometrika*, 8, pp. 109-122.
- [13] Kendall, M.G. (1952), George Udny Yule, 1871-1951, *Journal of the Royal Statistical Society (A)*, 115, pp. 156-161.
- [14] Kotz, S. and Johnson, N.L. Eds (1982 à 1988), *Encyclopedia of Statistical Sciences*, 9 vol., J. Wiley, New York.
- [15] Kruskal, W.A. and Tanur, J., Eds (1978), *International Encyclopedia of Statistics*, 2 vol., Free Press, New York.

- [16] Laplace, P.S. (1811), Mémoires sur les intégrales définies et leur application aux probabilités, *Mémoires de l'Institut Impérial de France*, Année 1810, pp. 279-347.
- [17] Pearson, K (1892), *The Grammar of Science*, Dutton, New York.
- [18] Pearson, K. (1894), Contributions to the Mathematical Theory of Evolution - I, pp. 1-40 in Karl Pearson, *Karl Pearson's Early Statistical Papers*, Cambridge Univ. Press (First published as "On the Dissection of Asymmetrical Frequency Curves" in Volume 185 of the *Philosophical Transactions* of the Royal Society of London, Series A).
- [19] Pearson, K. (1895), Contributions to the Mathematical Theory of Evolution - II : Skew Variation in Homogeneous Material, pp. 41-112 in Karl Pearson, *Karl Pearson's Early Statistical Papers*, Cambridge Univ. Press (First published in Volume 186 of the *Philosophical Transactions* of the Royal Society of London, Series A).
- [20] Pearson, K. (1896), Mathematical Contributions to the Theory of Evolution - III : Regression, Heredity and Panmixia, pp. 113-178 in Karl Pearson, *Karl Pearson's Early Statistical Papers*, Cambridge Univ. Press (First published in Volume 187 of the *Philosophical Transactions* of the Royal Society of London, Series A).
- [21] Pearson, K. (1900), On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, *Philosophical Magazine*, 5th series L, pp. 157-175.
- [22] Pearson, K. (1901), On the correlation of characters not quantitatively measurable, *Philosophical Transactions*, A, 195, pp. 1-47.
- [23] Pearson, K. (1904), On a generalised theory of alternative inheritance, with special reference to Mendel's laws, *Philosophical Transactions*, A, 203.

- [24] Pearson, K. (1904), On the theory of contingency and its relation to association and normal correlation, *Drapers' Company Research Memoirs, Biometric Series*, I.
- [25] Pearson, K. (1907), Reply to certain criticism of Mr. G.U. Yule, *Biometrika*, 5, pp. 470-476.
- [26] Pearson, K. and Heron, D. (1913), On theories of association, *Biometrika*, 9, pp. 159-315.
- [27] Sheppard, W.F. (1899), On the application of the theory of error to cases of normal distribution and correlation, *Philosophical Transactions*, A, 192, pp. 101-167.
- [28] Spearman, C. (1904), The proof and measurement of association between two things, *American Journal of Psychology*, 15, pp. 72-101.
- [29] Stigler, S. M. (1986), *The history of statistics - the measurement of uncertainty before 1900*, The Belknap Press of Harvard University Press, Cambridge, Massachussets.
- [30] Walker, H.M. (1929), *Studies in the history of statistical method*, Williams and Wilkins, Baltimore (réimprimé en 1975, chez Arno Press à New York).
- [31] Yule, G.U. (1895), On the correlation of total pauperism with proportion of outrelief, I : all ages, *Economic Journal*, 5, pp. 603-611.
- [32] Yule, G.U. (1900), On the association of attributes in statistics : with illustrations from the material from the childhood society, *Philosophical Transactions*, A, 194, pp. 257-319.
- [33] Yule, G.U. (1903), Notes on the theory of association of attributes in statistics, *Biometrika*, 2, pp. 121-134.

- [34] Yule, G.U. (1906), On a property which holds good for all groupings of a normed distribution of frequency for two variables, with applications to the study of contingency tables for the inheritance of unmeasured qualities, *Proceedings of the Royal Society*, A, 77.
- [35] Yule, G.U. (1912), On the methods of measuring association between two attributes, *Journal of the Royal Statistical Society*, 75, pp. 579-652.
- [36] Yule, G.U. (1921), On the time-correlation problem, with especial reference to the variate-difference correlation method, *Journal of the Royal Statistical Society*, 84, pp. 497-526.
- [37] Yule, G.U. (1926), Why do we sometimes get nonsense-correlations between time-series ? - a study in sampling and the nature of time-series, *Journal of the Royal Statistical Society*, 89, pp. 1-64.
- [38] Yule, G.U. (1927), On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers, *Philosophical Transactions*, A, 226, pp. 267-298.
- [39] Yule, G.U. (1945), On a method of studying time-series based on their internal correlations, *Journal of the Royal Statistical Society*, 108, pp. 208-230.