

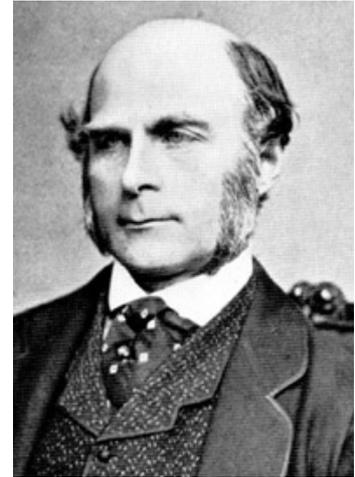
L'incroyable histoire de la statistique

Par Jean-Marie Gogue

Au 20^e siècle, sept mathématiciens inconnus du grand public sont à l'origine de profondes transformations : ils ont inventé la statistique moderne. Pas celle à laquelle on pense habituellement, celle des études économiques et des enquêtes d'opinion. Non. Ils ont inventé des méthodes d'analyse statistique, des algorithmes, comme on dit aujourd'hui, qui ont accéléré la recherche scientifique dans tous les domaines. Une accélération prodigieuse.

Francis Galton (1822-1911)

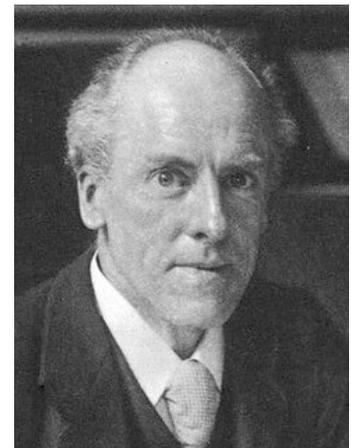
Francis Galton est né en Angleterre, dans les Midlands. Son père est banquier ; il est cousin de Charles Darwin. Il commence des études de médecine, à Birmingham, puis à Londres, mais la médecine ne l'intéresse pas. En 1840, il entre au *Trinity College* à Cambridge pour suivre des cours de mathématiques, mais il quitte l'université parce que l'enseignement ne lui convient pas. Il reprend alors des études de médecine, à Londres. À la mort de son père, en 1844, il hérite d'une grande fortune qui lui permettra de satisfaire sa passion : les voyages. Après un premier voyage en Egypte, le long du Nil, il dirige en 1852 une grande expédition en Afrique australe.



En 1859, la publication du livre de Darwin sur la sélection naturelle lui donne un nouveau sujet d'intérêt. Il abandonne alors ses idées de voyage et consacre l'essentiel de ses activités à l'étude des variations des caractères dans les populations humaines, la couleur des yeux par exemple. Pour cela, il faut trouver des méthodes afin de faire les mesures, de rassembler les données et de les analyser. C'est ainsi qu'il invente la régression et la corrélation. Ses travaux seront poursuivis par d'autres statisticiens.

Karl Pearson (1857-1936)

Karl Pearson est né dans le Yorkshire. Ses parents l'inscrivent au *University College* de Londres, mais il le quitte à seize ans en raison d'une santé fragile. Aidé par un précepteur, il réussit l'examen d'entrée à Cambridge en 1875 et entre au *King's College*. Il y étudie les mathématiques. En 1879, à sa sortie de l'université, il obtient une bourse qui lui permet de voyager. Il se rend en Allemagne pour étudier la littérature allemande, puis il revient en Angleterre pour suivre des cours de droit. En 1884, il est nommé professeur de mathématiques au *University College* de Londres. Il écrit de nombreux articles de mathématiques, de physique et d'histoire des sciences.



En 1891, il est nommé à la chaire de géométrie du *Gresham College* de Londres, où il fait la connaissance de Galton. Celui-ci vient de publier un livre : *L'héritage naturel*. Devenu son ami, il crée des méthodes statistiques qu'il appliquera à l'étude de la sélection naturelle. En 1900, il fonde avec Galton la revue *Biometrika* dans laquelle il publiera de nombreux articles. Il invente le « test du khi-deux » qu'il présente dans un article la même année. Ce test permet de savoir si le hasard est la seule cause des écarts observés au cours d'une expérience.

Karl Pearson a un fils, Egon, lui aussi un éminent statisticien. Il lui succédera en 1933 à la tête du Département de Statistiques Appliquées au *University College* de Londres.

Jerzy Neyman (1894-1981)

Jerzy Neyman est né en 1894 dans une famille polonaise, en Roumanie orientale. Son père meurt alors qu'il a douze ans. Devenue veuve, sa mère déménage à Kharkov, où il fréquente le collège et l'université. Il parle couramment polonais, russe, allemand et français. En 1912, il décide d'étudier les mathématiques. Il obtient un diplôme de troisième cycle à Kharkov en 1917. L'année suivante, il est nommé professeur à l'Institut de Technologie de Kharkov.

Sa famille déménage en Pologne en 1921 à cause de la guerre russo-polonaise. Il obtient un doctorat en mathématiques, avant de partir à Londres en 1924 pour étudier sous l'autorité de Karl Pearson. Il entre en contact avec son fils, Egon.

En 1926, il obtient une bourse Rockefeller pour étudier à Paris. Il assiste à des conférences de Borel, Hadamard et Lebesgue au Collège de France. En 1927, il retourne en Pologne et devient chef de laboratoire dans un institut de Varsovie.

En 1934, Karl Pearson prend sa retraite. Son département du *University College* est coupé en deux ; une partie est dirigée par son fils, Egon, qui invite Neyman à venir travailler avec lui sur la théorie des tests d'hypothèse. Ils sont les premiers à reconnaître la nécessité d'émettre des hypothèses alternatives. Ils définissent les deux types d'erreurs connus aujourd'hui de la plupart des chercheurs : d'une part rejeter une hypothèse vraie ; d'autre part accepter une hypothèse fautive. Un progrès considérable.

Toujours en 1934, Neyman élabore la théorie de l'échantillonnage dans les sondages. Cette théorie répond à la demande américaine pour les enquêtes d'opinion et les sondages, une demande qui tourne parfois à l'obsession. C'est ainsi qu'il est invité à donner des conférences aux Etats-Unis.

En 1937, il entre à l'université de Berkeley pour créer un département de statistiques. Il donne des conférences à la *Graduate School* du ministère de l'Agriculture, à Washington. C'est la première fois que des statisticiens américains ont la possibilité de l'écouter.

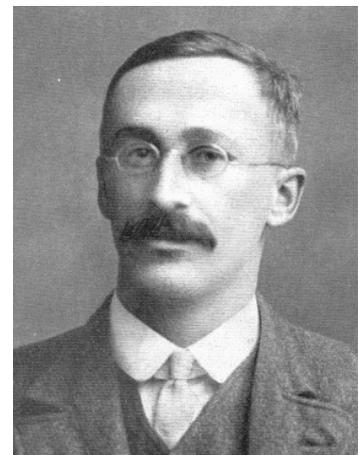
En 1948, il lance un symposium de statistique mathématique. Cette manifestation, qui s'est renouvelée tous les cinq ans jusqu'en 1972, a fortement influencé les statisticiens américains.



William Gosset (1876-1937)

William Gosset est né à Canterbury. Il fréquente le *Winchester College*, puis l'université d'Oxford où il étudie les mathématiques et les sciences naturelles. Diplômé de chimie en 1899, il est embauché par la brasserie Guinness, à Dublin en Irlande, en qualité de brasseur. Il s'intéresse aussitôt à la validité des résultats obtenus en laboratoire.

Il prend un congé en 1906 pour étudier la statistique sous l'autorité de Karl Pearson au *University College* de Londres. Celui-ci l'aide à écrire des articles, mais n'a pas compris leur importance. Gosset s'interrogeait sur la pertinence d'un calcul statistique portant sur de petits échantillons.



Pour empêcher les employés de divulguer des renseignements confidentiels, Guinness leur avait interdit de publier des documents, quel qu'en soit le contenu. Toutefois, après avoir expliqué à la direction du personnel que ses conclusions mathématiques et philosophiques ne pouvaient pas être utilisées par des brasseries concurrentes, Gosset fut autorisé à les publier, mais sous un pseudonyme. Les articles de Gosset ont donc été publiés dans *Biometrika* sous le pseudonyme de *Student*.

Guinness attachait une grande importance à la recherche scientifique pour améliorer ses produits. Comment augmenter la production en conservant la saveur de la bière ? Gosset devait répondre à cette question ; son équipe avait pour but d'améliorer le processus de sélection du houblon.

Gosset cherchait à comprendre dans quelle mesure un petit échantillon est moins représentatif d'une distribution qu'un échantillon plus grand. C'est ainsi qu'il a élaboré la « distribution de Student » qui est à la base de la statistique moderne.

Un jeune mathématicien nommé Ronald Fisher fut enthousiasmé par les articles de Student. Il fut particulièrement séduit par l'idée selon laquelle la distribution de Student peut être utilisée pour connaître la probabilité d'un résultat. Nous verrons plus loin que l'intérêt de Fisher pour le travail de Gosset a influencé toute la science statistique.

Cependant, Gosset reste peu connu du public. C'est sans doute parce qu'il a travaillé dans une entreprise plutôt que dans le milieu universitaire. S'il avait été professeur d'université, il aurait pu utiliser son propre nom dans ses articles et publier un livre pour exposer ses méthodes et ses idées. Mais il n'aurait jamais eu l'occasion d'aborder les problèmes pratiques qui ont stimulé son imagination.

Ronald Fisher (1890-1962)

Ronald Fisher est né à Londres où son père dirige un cabinet commissaires-priseurs. Très jeune il est attiré par la physique. En 1912, il obtient une licence d'astronomie à l'université de Cambridge, puis il enseigne la physique dans des écoles privées. En 1919, il est embauché au centre de recherches agronomiques de Rothamsted. Il y restera jusqu'en 1933.



Fisher avait fait la connaissance de Gosset en 1912, mais ne l'avait jamais rencontré. Dix ans plus tard il l'invite à Rothamsted pour discuter de statistique mathématique. C'est pendant cette période qu'il élabore le concept d'analyse de la variance.

L'analyse de la variance consiste à comparer simultanément les résultats de mesure de plusieurs échantillons soumis à des traitements différents. C'est un grand progrès par rapport à la méthode qui consiste à comparer les résultats deux à deux.

En 1925, Fisher publie le livre *Statistical Methods for Research Workers* qui sera réédité plusieurs fois. Sa renommée grandit ; il commence à voyager et à donner des conférences. En 1931, il passe six semaines aux Etats-Unis dans le centre de statistique du *Iowa State College*, où il donne trois conférences par semaine et rencontre de nombreux statisticiens américains, notamment George Snedecor. Il y retournera en 1936.

En 1933, Karl Pearson part à la retraite. Une partie du département de statistique appliquée est confiée à Egon Pearson, l'autre à Ronald Fisher. Mais cette partie disparaît en 1940 et c'est seulement trois ans plus tard qu'il retrouvera un emploi : un poste de professeur de génétique à Cambridge, où il restera jusqu'en 1957.

Ronald Fisher est reconnu dans le monde entier comme un statisticien de génie, comme le principal pilier de la statistique mathématique. Ses concepts révolutionnaires ont permis aux laboratoires de recherche, dans toutes les disciplines, de faire des progrès considérables.

George Snedecor (1881-1974)

George Snedecor est né à Memphis, Tennessee. Il grandit en Floride et en Alabama, où son père avait déménagé avec femme et enfants pour accomplir une vocation religieuse.

Snedecor obtient sa licence en mathématiques et en physique à l'université d'Alabama en 1905, puis son master en physique à l'université du Michigan en 1913. La même année, il est nommé professeur adjoint de mathématiques au *Iowa State College* à Ames. Il est promu professeur associé en 1914 et finalement professeur en 1931. Il entretiendra une relation avec Fisher pendant de nombreuses années.



Dès son arrivée dans l'Etat de l'Iowa, Snedecor s'intéresse à la statistique, en particulier aux méthodes concernant la recherche agronomique. Il collabore avec Henry Wallace, un ancien étudiant de l'université qui avait donné en 1924 une série de conférences sur la statistique. Henry Wallace sera plus tard vice-président des Etats-Unis, entre 1941 et 1945, sous la présidence de Roosevelt. Ses conférences ont provoqué un grand intérêt pour la statistique dans tout l'Etat de l'Iowa.

Une demande croissante de conseils en méthodes statistiques, notamment par les agriculteurs, incite l'Etat de l'Iowa à créer en 1927 dans le département de mathématiques de l'université une section de statistique et à en confier la charge à Snedecor. En 1933, il crée un laboratoire de statistique. Sous son impulsion, ce laboratoire devient l'un des plus importants centres de statistique des Etats-Unis. Il abandonne ses fonctions administratives en 1947, mais il continue à enseigner et à faire de la recherche à plein temps jusqu'en 1952.

En 1958, Snedecor est embauché comme consultant par le laboratoire d'électronique de la Navy, à San Diego en Californie.

Au cours de sa carrière, Snedecor a écrit de nombreux articles et plusieurs livres de statistique dont la lecture est moins difficile que celle des ouvrages de Fisher. Il a publié notamment en 1937 le livre *Statistical Methods* qui a été réédité six fois jusqu'en 1967 et traduit en plusieurs langues, dont le français.

Walter Shewhart (1891-1967)

Walter Shewhart est né à New Canton, Illinois. Après de brillantes études à l'université de l'Illinois, il obtient en 1917 un doctorat de physique à Berkeley, université de Californie. Il est d'abord nommé assistant de physique à l'université de l'Illinois, puis chef du département de physique dans un collège du Wisconsin.

En 1918, il est embauché par la société *Western Electric*, qui lui donne la mission d'améliorer la production d'une grande usine de téléphonie près de Chicago. En 1925, il est muté aux *Bell Telephone Laboratories* à New York, la maison mère, où il restera jusqu'à sa retraite.



En 1932, Shewhart fait un voyage en Angleterre où il rencontre Egon Pearson et William Gosset, ce qui le fait reconnaître par la communauté scientifique. L'année précédente, il avait publié un gros ouvrage sur son expérience aux *Bell Telephone Laboratories*, mais il n'avait pas retenu l'attention du public américain. L'accueil de Pearson à Londres est une consécration. En 1935, il fonde l'Institut de Statistique Mathématique dont il assure la présidence. De 1935 à 1944, le ministère de la Défense lui donne une mission de conseil. En 1943, il fonde la collection de statistique mathématique chez l'éditeur *John Wiley and Sons*. Il la dirigera jusqu'en 1965.

En 1938, Deming invite Shewhart à donner des conférences à la *Graduate School* du ministère de l'Agriculture. Le texte de ces conférences est publié l'année suivante. Ce sera la base de l'enseignement de Deming au Japon après la seconde guerre mondiale.

Dans l'usine de *Western Electric*, Shewhart disposait de moyens importants qui lui permettaient de faire des expériences avec des données réelles. Il a compris que tous les problèmes liés à la qualité étaient provoqués par les variations des caractéristiques des produits. Par conséquent, l'entreprise ne pouvait pas améliorer la qualité sans connaître les variations des caractéristiques et sans en chercher les causes. C'était un nouveau concept.

À la base de ce concept se trouve la notion de contrôle statistique d'un processus. Shewhart a montré que, lorsqu'on étudie un processus, il faut d'abord déterminer s'il est sous contrôle ou hors contrôle. C'est ainsi qu'il a élaboré la théorie des variations et inventé un outil statistique : le graphique de contrôle.

Pour fonder la théorie des variations, Shewhart s'est servi du concept de « système de causes aléatoires » qu'un mathématicien français, Pierre-Simon de Laplace, avait déjà évoqué en 1815 dans sa « Théorie analytique des probabilités ». Le graphique de contrôle est un outil qui permet de repérer parmi des causes aléatoires la présence de causes significatives.

Shewhart était un homme modeste et désintéressé. Il concevait le rôle du statisticien dans une entreprise comme celui de coordinateur de plusieurs disciplines intervenant dans un processus d'apprentissage collectif.

Conclusion

Voici donc les sept mathématiciens : deux Américains, quatre Britanniques et un Polonais, qui ont inventé les méthodes statistiques modernes, des méthodes utilisées aujourd'hui dans les laboratoires du monde entier.

Plusieurs d'entre eux se connaissaient ; ils travaillaient en commun pour créer de nouvelles théories et de nouvelles méthodes. Ceux qui étaient employés dans des entreprises : Gosset et Shewhart, correspondaient régulièrement avec leurs collègues universitaires.

Grâce à eux, la science et l'industrie ont fait d'immenses progrès dans tous les domaines, des progrès que personne ne pouvait imaginer au début du vingtième siècle.

Août 2017