



William Gosset

1876-1937

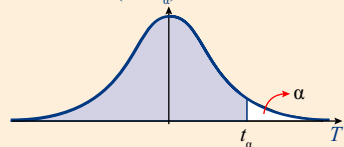
La statistique a connu de nombreux développements au XX<sup>e</sup> siècle et l'un de ces développements, le test de Student est due au brasseur et statisticien William Gosset, pionnier de l'analyse expérimentale sur de petits échantillons.

# William Sealy Gosset

## La loi de Student

La loi de Student est une loi de probabilité dont la courbe ressemble à celle de la loi normale, c'est-à-dire une cloche symétrique par rapport à la moyenne 0. Cependant, la courbe de la loi de Student est plus aplatie que la courbe normale.

$$\Pr(T \leq t_\alpha) = 1 - \alpha$$



En fait, il existe plusieurs distributions de Student et il faut choisir celle qui convient selon la taille de l'échantillon à l'aide de laquelle on détermine  $d$  le degré de liberté de la courbe,

$$d = n - 1,$$

où  $n$  est la taille de l'échantillon.

William Sealy Gosset est né à Canterbury, en Angleterre le 13 juin 1876. Il est le fils aîné d'Agnes Sealy Vidal et du colonel Frederic Gosset du Royal Engineers<sup>1</sup>. Gosset étudie d'abord au Winchester College<sup>2</sup> avant de s'inscrire en chimie et en mathématiques au New College d'Oxford<sup>3</sup>. Après avoir obtenu son diplôme en 1899, il rejoint la brasserie Arthur Guinness & Son à Dublin, en Irlande où il a fait une carrière de 38 ans.

Le travail de Gosset à la brasserie vise à uniformiser le goût de la bière. Pour ce faire, il développe de nouvelles méthodes statistiques portant sur de petits échantillons. Les méthodes développées par les statisticiens anglais étaient surtout appliquées en biométrie sur des échantillons comportant des centaines

1. Le *Corps of Royal Engineers* (Corps des ingénieurs royaux) est l'un des corps d'appui de l'armée britannique. Il s'occupe du génie de combat : génie militaire, ingénierie et support techniques aux forces armées britanniques.
2. Le Winchester College est une *public school* anglaise située à Winchester, dans le Hampshire. La vocation des *public schools* étant de promouvoir aux postes du clergé et de l'administration des enfants sans fortune mais doués pour les études, en échange de leurs prières pour l'âme du fondateur.
3. Le New College est l'un des établissements constitutifs de l'université d'Oxford, en Angleterre.

d'individus. Pour uniformiser le goût de la bière, Gosset doit prélever des échantillons à la brasserie mais également à la ferme pour déterminer les variétés d'orge qui donnent le meilleur rendement en terme de goût.

Au début du XX<sup>e</sup> siècle, l'analyse statistique est dominée par les concepts de populations et d'échantillons de très grande taille dont le principal acteur est Karl Pearson. Cette analyse se fonde la distribution normale. Obtenue pour la première fois par Abraham de Moivre en 1733 et éditée en 1738 elle est initialement utilisée pour prédire l'issue des jeux de hasard. Par la suite, elle sert à décrire la distribution des erreurs en astronomie, puis comme distribution de fréquence de probabilité avec Gauss.

Gosset n'a pas à sa disposition d'outil statistique applicable à de petits échantillons. Il procède d'abord par essais et erreurs et ses recherches sur la culture de l'orge l'amène à penser que les expériences devaient viser non seulement à améliorer le rendement moyen mais aussi à sélectionner des variétés dont le rendement était insensible aux variations du sol et du climat.

En 1906-1907, il passe deux trimestres au laboratoire de biométrie de Karl Pearson au University College de Londres. Gosset et Pearson développent alors une

bonne relation et Pearson aide Gosset à peaufiner les mathématiques de ses articles. Cependant, Pearson, habitué à travailler sur de grands échantillons en biométrie ne perçoit pas toute l'importance des travaux de Gosset.

Gosset ne peut publier ses travaux sous son nom. À l'époque, un autre chercheur de Guinness avait déjà publié un article contenant des secrets commerciaux de la brasserie Guinness. Le conseil d'administration de la Guinness a alors décidé d'autoriser ses scientifiques à publier des recherches, mais à condition de ne pas mentionner la bière ni leur nom de famille. Il doit donc se trouver un pseudonyme.

Inspiré par un cahier de notes de 1906-1907, *The Student's Science Notebook*, Gosset choisit le nom de plume « Student ». C'est le nom que porte sa réalisation la plus remarquable, la *loi de Student*. C'est une loi de probabilité, faisant intervenir le quotient entre une variable suivant une loi normale centrée réduite et la racine carrée d'une variable distribuée suivant la loi du  $\chi^2$ .

Gosset a publié la plupart de ses 21 articles universitaires, y compris *The probable error of a mean*, dans la revue *Biometrika* de Pearson sous son pseudonyme Student. Ce n'est cependant pas Pearson mais Ronald A. Fisher (1890-1962) qui a apprécié l'importance des travaux de Gosset sur de petits échantillons. En 1912, Fisher écrit à Gosset expliquant que la distribution  $t$  de Student devrait être divisée par degrés de liberté et non par la taille totale de l'échantillon. De 1912 à 1934, Gosset et Fisher échangèrent plus de 150 lettres. En 1924, Gosset écrit dans une lettre à Fisher : « Je vous envoie une copie des Tables de Student car vous êtes le seul qui soit susceptible de les utiliser ».

Fisher a contribué aux applications de la distribution  $t$  de Student à l'analyse de régression. En 1935, à l'âge de 59 ans, Gosset quitte Dublin pour occuper le

poste de brasseur en chef dans une nouvelle (et deuxième) brasserie Guinness au nord-ouest de Londres. En septembre 1937, il est promu brasseur en chef de toute la compagnie Guinness. Il décède un mois plus tard d'une crise cardiaque, à l'âge de 61 ans.

## Les lois normale et de Student

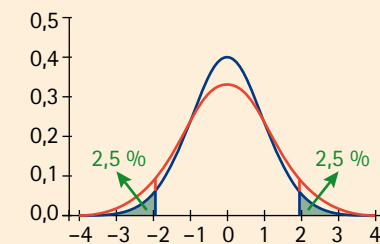
La loi normale s'écrit :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

où  $\mu$  est la moyenne et  $\sigma$  est l'écart type. En biologie, un grand nombre de paramètres sont décrits avec précision par une distribution normale dont le graphique a les caractéristiques suivantes :

- symétrique par rapport au centre, soit la valeur ayant la plus grande fréquence ;
- la moyenne, la médiane et le mode coïncident ;
- l'inflexion de la courbe est à  $\pm\sigma$  du centre  $\mu$  ;
- les extrémités sont asymptotiques à zéro.

La moyenne d'un échantillon de grande taille ( $n > 120$ ) issue d'une population décrite par une distribution normale, est elle aussi normalement distribuée. Ce n'est pas le cas pour les petits échantillons, ce qui a amené Gosset à établir la distribution de Student.



La courbe bleue est celle de la loi normale centrée réduite ( $\mu=0, \sigma=1$ ). La courbe rouge est la distribution de Student. La somme des aires délimitées par le trait vertical bleu aux extrémités de la courbe normale aux valeurs  $\pm 1,96\sigma$  représente 5% de l'aire sous la courbe normale.

En prolongeant ces traits verticaux jusqu'à la courbe de Student, la somme des aires de ces régions est plus grande que 5% de l'aire sous la courbe de Student.