Journal de la société statistique de Paris

Jean Dufrénoy

Représentation rectilinéaire de distributions logarithmiques

Journal de la société statistique de Paris, tome 88 (1947), p. 47-53

<http://www.numdam.org/item?id=JSFS_1947__88__47_0>

VI

VARIÉTÉ

Représentation rectilinéaire de distributions logarithmiques.

Pour représenter la répartition des longueurs (ou des surfaces, des volumes ou des masses) dans une population, on peut porter en abscisses les classes de dimensions mesurées et en ordonnées, soit les fréquences pour chaque classe (ce qui dessine la courbe de distribution de fréquences), soit les fractions du nombre total de fréquences, ayant une dimension inférieure à celle de la classe d'abscisse correspondante, ce qui donne la courbe cumulative, dite en S ou en ogive.

En fait, la plupart des courbes expérimentales de distribution de fréquences diffèrent de la « courbe normale de probabilité » par leur dissymétrie, due à ce que la branche descendante a une pente beaucoup moins forte que la branche montante.

La plupart de ces courbes expérimentales de distribution sont rendues « log. normales », c'est-à-dire symétriques par rapport à la moyenne géométrique, lorsqu'on inscrit les abscisses sur une échelle logarithmique.

Par exemple, selon les termes de d'Arcy Thompson, « si la richesse engendre la richesse, la distribution arithmétique des richesses dessine une courbe dissymétrique, mais, logarithmiquement, cette courbe devient normale ».

D'une façon générale, si la déviation varie en fonction de la grandeur des individus composant la population, il est évident que la distribution de fréquence sera dissymétrique par rapport à la moyenne arithmétique, mais symétrique par rapport à la moyenne géométrique.

Cette signification de la moyenne géométrique a, au cours de ces récentes années, acquis une importance particulière dans deux ordres de recherches en apparence bien différentes, mais mettant en œuvre, dans l'un et l'autre cas, une distribution étudiée dans un espace à une dimension : d'une part l'agrégation de syllabes en mots ou celle des mots en phrases; d'autre part, l'agrégation d'unités physiques ou « monomères » en polymères linéaires.

Distribution log. normale des fréquences de mots dans les phrases.

C. B. Williams a montré que la distribution des fréquences des phrases d'un texte par classe de 1, 2, mots entre deux signes de ponctuation consécutifs, est fortement dissymétrique lorsque les nombres de mots sont portés en abscisses sur échelle arithmétique, mais devient normale sur échelle logarithmique.

La distribution de log x étant normale, nous pouvons en conclure que la variabilité de la longueur d'une phrase est proportionnelle à sa longueur; un auteur dont la pensée tend à s'exprimer en phrases courtes, de 10 mots environ, écrira des phrases oscillant entre 8 et 12 mots; un auteur dont la pensée s'exprime par des phrases d'une centaine de mots, écrira des phrases oscillant entre 80 et 120 mots; les variations sont « proportionnelles », c'est-à-dire géométriques, et la moyenne géométrique étant prise comme centre de distribution, il y a autant de phrases contenant un nombre de mots compris entre cette moyenne géométrique et la moitié de sa valeur numérique qu'il y a de phrases ayant une longueur comprise entre cette moyenne géométrique et le double de sa valeur.

Équations des distributions normales et log. normales de probabilité

L'équation de la distribution normale de probabilité

$$n = \frac{s\,(n)}{\sigma\sqrt{2\,\pi}}\, e^{\frac{-(x-\mathrm{M})^2}{2\,\sigma^2}}$$

devient, lorsqu'on utilise une échelle logarithmique des abscisses

$$n = \frac{s\,(n)}{\log\,\sigma_g\,\sqrt{2\,\pi}}\, e^{\frac{-(\log.\,x-\mathrm{M}_g)^2}{2\,\log^2\sigma_g}}$$

où M_g est la moyenne géométrique des valeurs de x et σ_g la déviation standard géométrique.

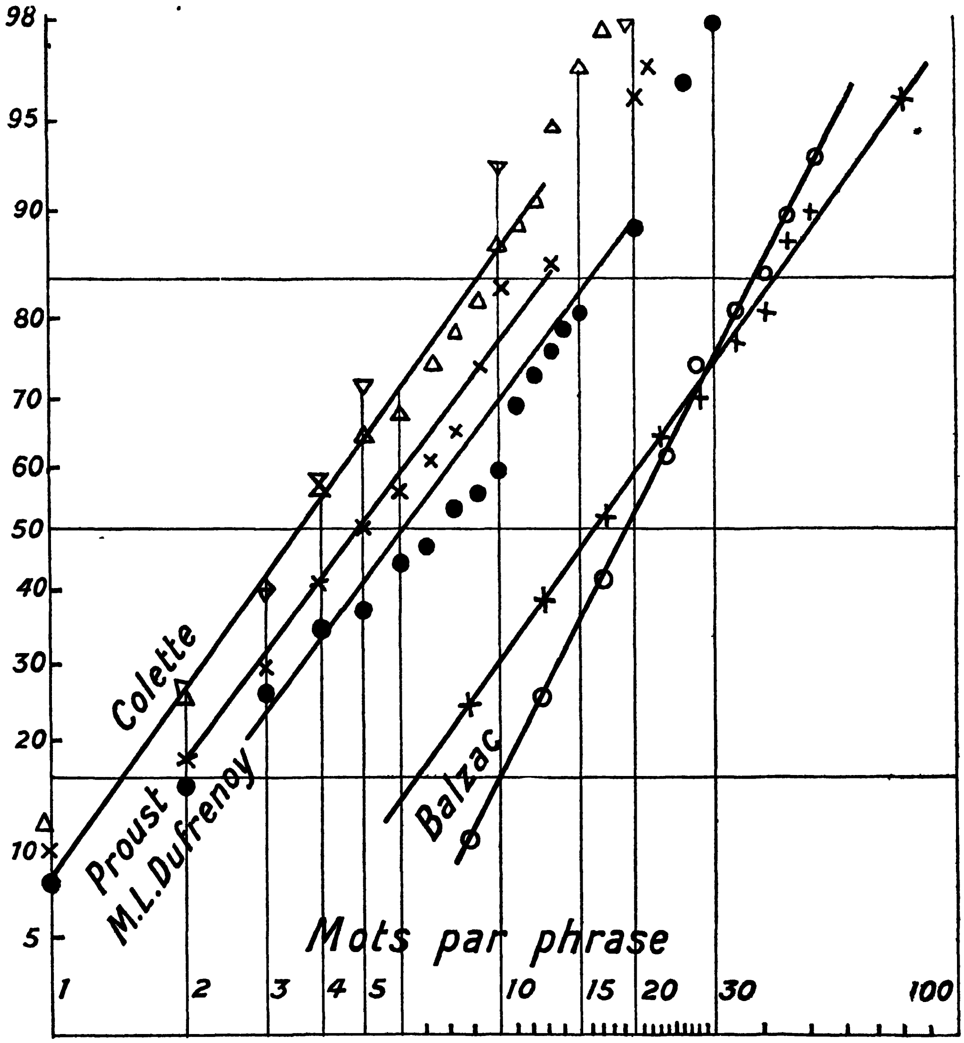
Les courbes d'intégration correspondant à la courbe de distribution normale (échelle arithmétique des abscisses) ou à la courbe de distribution log. normale (échelle logarithmique des abscisses) peuvent être transformées en droites par emploi en ordonnées d'un axe de pourcentage cumulatifs, c'est-à-dire d'une échelle d'unités espacées conformément à l'intégrale de la fonction de la distribution normale, et, en abscisses, soit de l'échelle arithmétique (papiers de probabilité), soit de l'échelle logarithmique (papiers de coordonnées log. probabilité).

Distribution de poids moléculaires.

Le papier log. probabilité convient particulièrement à la représentation rectilinéaire des longueurs de polymères linéaires comme à celle des longueurs de phrases dans un texte.

D'ailleurs la distribution des dimensions de polymères, dans un espace à deux ou à trois dimensions obéit à cette même distribution ainsi que les distributions de masses (exprimées par exemple sous forme de poids moléculaires) d'où les multiples applications industrielles de ces distributions.

La formation des fibres synthétiques, qui a acquis récemment un si grand



Graphique I.

intérêt industriel, résulte d'une polymérisation dans un espace à une dimension, c'est-à-dire de la répétition du même type de réaction élémentaire, intervenant aussi bien entre les unités (monomères ou m_1) qu'entre les chaînes provenant de l'agrégation, bout à bout, d'un certain nombre de ces monomères;

cette réaction peut s'écrire $m_i + m_j \longrightarrow m_{i+j}$ où i et j sont des nombres entiers.

Un monomère m_i peut s'intégrer dans un polymère en réagissant soit avec un autre monomère, soit avec une chaîne comprenant déjà x de ces monomères; l'analogie avec l'agrégation des mots en séries linéaires, ou phrases, est évidente. Cependant, le polymère linéaire peut être étudié non seulement quant à ses propriétés linéaires dans un espace à une dimension, mais quant à toutes ses autres propriétés physiques, dépendant de son degré de polymérisation, et en particulier quant aux distributions de poids moléculaire, de viscosité...

En général, les études expérimentales sur la distribution des poids moléculaires permettent, en portant en abscisses les degrés de polymérisation et en ordonnées les pourcentages cumulatifs correspondants, d'obtenir des courbes en ogive, que l'emploi de papiers à coordonnées log. probabilité permettent de transformer en droites.

De même, les études sur la distribution des longueurs de phrases d'un texte, permettent en portant en abscisses les log. du nombre des mots par phrase et en ordonnées les pourcentages cumulatifs, d'obtenir des courbes en ogive, transformables en droite par emploi de coordonnées log. probabilité (graphique I)

Modifications de l'échelle de probabilité.

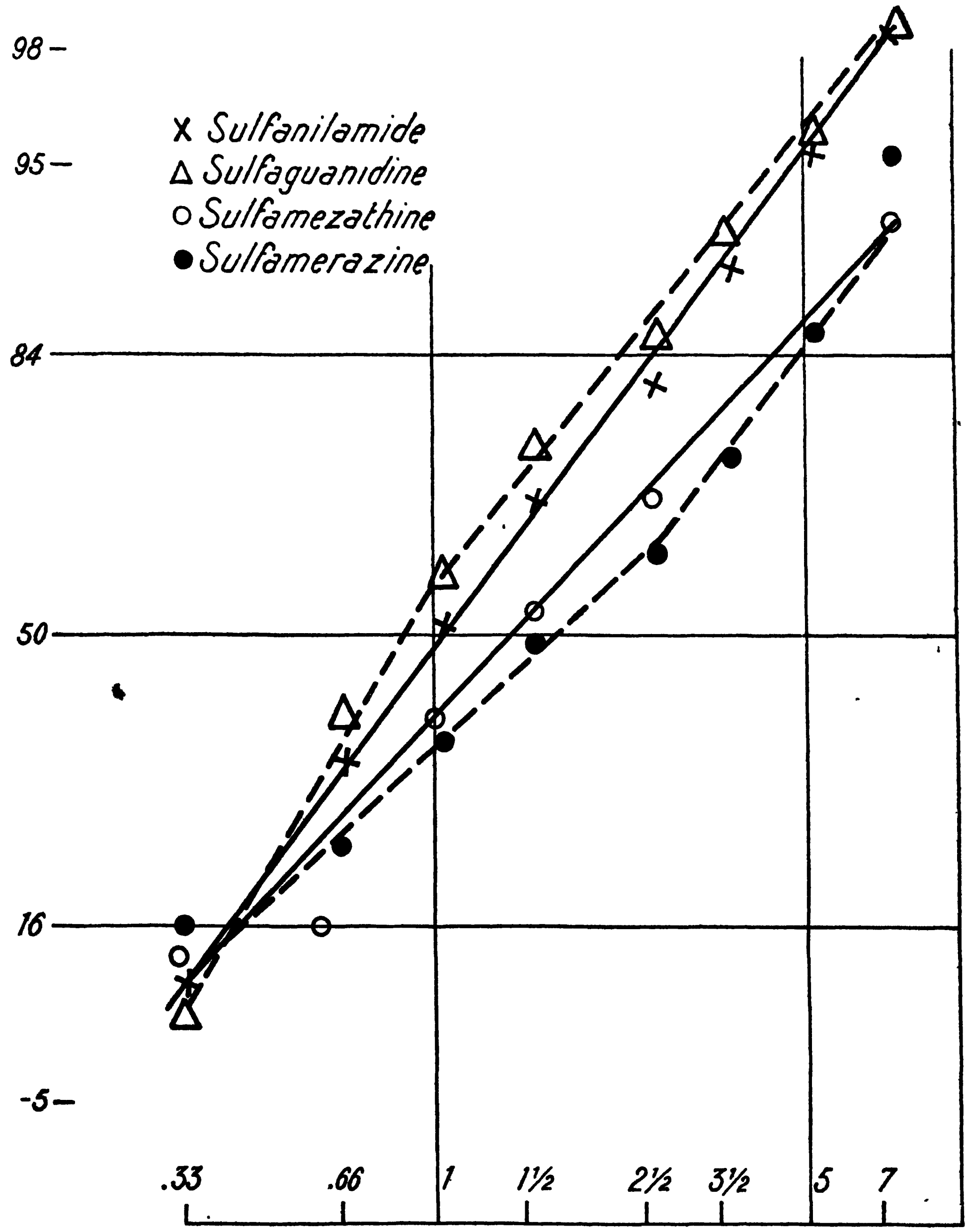
De récentes recherches expérimentales ont conduit à des distributions de fréquences de poids moléculaires qui ne pouvaient pas être rendues rectilinéaires par l'emploi des coordonnées log. probabilité, et pour lesquelles Boyer a construit une série d'échelles de pourcentages en introduisant dans l'équation

$$W(P)\, dP = \frac{(1 - \alpha)^{b+2}}{(b+1)!} P^{b+1} \alpha^{P}\, dP$$

(où W (P) représente le poids de la fraction de polymères compris dans l'intervalle (d P) et P le degré de polymérisation) les valeurs numériques des paramètres α et b telles que la courbe d'intégration graphique de la fonction présente un accord satisfaisant avec la courbe expérimentale étudiée. Ayant ainsi obtenu la courbe en S des pourcentages cumulatifs pour les degrés successifs de polymérisation, il suffit, à intervalles égaux le long de l'échelle des pourcentages cumulatifs, en ordonnées, de tracer des droites parallèles coupant la courbe, puis de projeter les points d'intersection sur l'axe des abscisses, pour définir l'échelle de pourcentage à employer.

Parmi les échelles ainsi obtenues, nous utiliserons celle correspondant à $\alpha = 0{,}995$ et $b = 2$ pour la rectification d'une distribution logarithmique, empruntée aux résultats des Recherches effectuées aux Research Laboratories, Imperial Chemical Industries Ltd, England. Les auteurs, F.-L. Rose et A. Spinks, ont publié dans « The Journal of Pharmacology and Experimental Therapeutics » (86 : 264-272, 1946) une série de courbes, où figurent en ordonnées, les nombres de milligrammes de sulfonamide, pour 100 cc. de sang, chez des souris ayant absorbé 5 milligrammes de drogue par 20 grammes de poids vif, et en abscisses, le nombre d'heures écoulées depuis l'ingestion de la drogue.

Si nous portons en abscisses les temps sur échelle logarithmique et en ordonnées, sur échelle de probabilité normale, les pourcentages cumulatifs, nous obtenons une droite dans le cas de la sulfanilamide ou de la sulfamezathine mais une ligne brisée dans les cas de la sulfaguanidine ou de la sulfamerazine

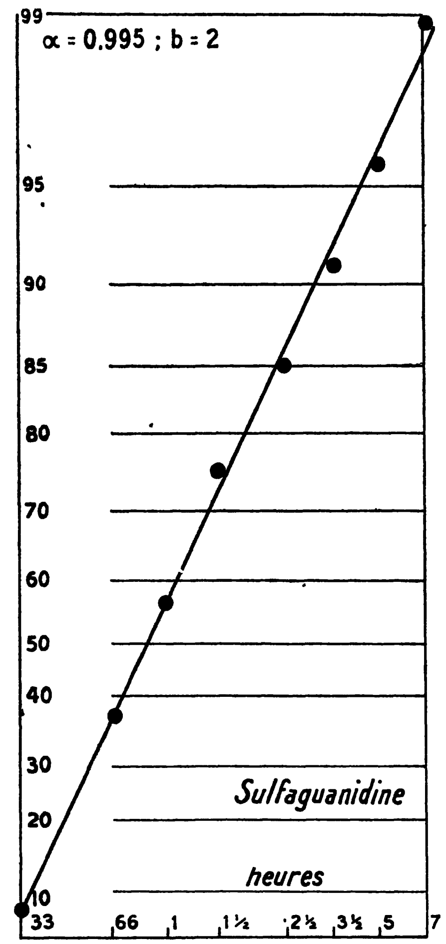


Graphique II.

(graphique II). Le ligne de régression pour la sulfaguanidine peut être rendue rectilinéaire par l'emploi de l'échelle de probabilité de Boyer où $\alpha = 0{,}995$ et $b = 2$ (graphique III).

Conclusions.

La plupart des courbes expérimentales de distribution de fréquences ne sont pas « normales » quand les classes de fréquences sont portées sur une échelle arithmétique, mais deviennent « normales » quand les classes de fréquences sont portées sur une échelle logarithmique. Ces distributions, « log. normales », sont symétriques par rapport à la moyenne géométrique.



Graphique III.

La courbe d'intégration correspondante est une courbe en S, qui peut se transformer en droite par transformation des fréquences cumulatives en pourcentages cumulatifs et inscription en ordonnée sur échelle de probabilité; l'abscisse correspondant au pont d'intersection de la droite et de l'horizontale d'ordonnée correspondant à 50 % détermine la moyenne géométrique tandis que les abscisses correspondant aux points d'intersection avec les horizontales correspondant à 16 et 84 % déterminent, de part et d'autre de la moyenne géométrique, les valeurs de la déviation standard géométrique.

Certaines courbes qui ne peuvent pas être rendues normales par l'emploi d'une échelle logarithmique des abcisses et qui, dès lors, ne peuvent pas se représenter par une droite sur échelle de pourcentage normale, peuvent se représenter par une droite grâce à l'emploi d'échelles convevables de pourcentage.

Jean Dufrénoy.

BIBLIOGRAPHIE

1916. Kaptein J.-C. « Skew frequency curves in Biology and Statistics. » *Rec. Trav. Botaniques Nederland* Groningen, 13 : 105-158, 1916.
1937. Dufrenoy J. et Vezian, « Représentation des phénomènes biochimiques et épidémiologiques par la courbe IV de Pearson. » *Rev. Microb. Appl.* 3 : 135-143, 1943.
1939. Austin J.-B. « Methods of representing Distribution of particle Size ». *Ind. Eng. Chem. Anal. Edit.* II : 334, 1939.
— Laskowski L. and Burk R.-E., *J. Chem. Phys.* 7 : 465, 1939.
1940. Williams C.-B., « A note on the statistical analysis of sentence-length as a criterion of literary style ». *Biometrika*, 31 : 356-61, 1940.
1942. Thompson D'Arcy, Growth and Form.

1943. TOBOLSKY A, POWELL R.-E. and EYRING H., « Elastic-viscous properties of Matter (The statistics of Long Chain Molecules, in *The Chemistry of Large Molecules : Frontiers in Chemistry*, vol. I, New-York, 1943).
1944. TRELOAR L.-R.-G., *Transact. Faraday Soc.* 40 : 109, 1944.
1946. BOYER R.-F. « Molecular weight Distribution Data on high Polymers. Graphical representation. *Ind. Eng. Chem. Anal. Edit.* 18 : 342, 1946.
— GUTH J. and MARK, *Kinetic theory of rubber elasticity*, *Advances in Colloidal Sc.*, II, 267, 1946.
— LAPPLE C. E., *Heating, Piping and Air conditioning*, 18 : 113, 1946.
— LUDWELL O.M., « A particle size distribution function for air born dusts », *Nature*, 158 : 61, 1946.
1946. TRELOAR, L.-R.-G., « The statistical length of longchain molecules ». *Transact. Faraday Soc.* 42 : 72-83, 1946; « The elasticity of a net work of long-chain molecules ». *Ibid.* 42 : 83-94, 1946.