MATHÉMATIQUES ET SCIENCES HUMAINES

Problèmes d'enseignement

Mathématiques et sciences humaines, tome 24 (1968), p. 51-57
<http://www.numdam.org/item?id=MSH_1968__24__51_0>

M.S.H. (7e année, nº 24, 1968, p. 51-57)

PROBLÈMES D'ENSEIGNEMENT

par

B. LECLERC

APPLICATIONS PRATIQUES DE LOIS DE PROBABILITÉ

Avec les quelques résumés d'articles qui suivent, est commencée la publication des produits d'une enquête en cours sur les cas où les distributions empiriques s'ajustent bien à des lois de probabilités, parfois conformément à un modèle mathématique sous-jacent (tel qu'un processus, comme pour le modèle de Coleman, cité ci-dessous).

Chacun des résumés que nous présentons comporte le maximum de renseignements parmi les suivants : description sommaire du modèle utilisé, méthode d'estimation des paramètres et tests de convenance du modèle, application à des distributions empiriques et résultats, bibliographie.

Les intentions de cette publication sont de présenter au lecteur des modèles éprouvés, portant sur des champs d'applications variés, de l'informer éventuellement sur l'existence dans des branches éloignées de la sienne de modèles susceptibles de l'intéresser, de contribuer peut-être à mettre un peu d'unité parmi les modèles existant, enfin de montrer la diversité des distributions de probabilité utilisées dans la pratique : nous en verrons de bien moins connues que les lois normale, poissonnienne, exponentielle, ou même parétienne, rencontrées dans ces premiers résumés publiés.

LOIS DE PARETO — *ÉCONOMIE*

H. T. DAVIS, "The nature of wealth and Income", ch. 9 de *The analysis of economic time-series*, Cowles commission monograph, Bloomington, Ind., The Principia Press, 1941.

I. — LA LOI DE PARETO.

Modèle.

L'auteur énonce ainsi la loi de Pareto :« En tous lieux et en tous temps, la distribution des revenus dans une économie stable, lorsque l'origine des mesures est prise à un niveau de revenus suffisamment élevé, sera donnée approximativement par la formule empirique :

$$y = ax^{-\nu}$$

où y est le nombre de gens ayant un revenu égal ou supérieur à x et où ν est approximativement égal à 1,5.»

Estimation.

Tracé sur papier logarithmique et ajustement linéaire de la branche droite de la distribution par la méthode des moindres carrés.

Application.

Application à la distribution des revenus aux Etats-Unis en 1918. Les données fournies de façon très détaillée par l'auteur, sont prises dans une étude du "National Bureau of Economic Research": estimations soignées des revenus de 37 569 060 personnes (environ 1/3 de la population totale), sans militaires. On trouve :

$$\log_{10} a = 9{,}28462 \qquad \nu = 1{,}69672$$

L'auteur termine cette partie en donnant les estimations d'un certain nombre de coefficients de Pareto : Angleterre (1843, 1879-80 et 1893-94), Prusse (6 années entre 1852 et 1894), Saxe (1880, 1806), villes d'Italie, Bâle, Paris, Augsbourg, Pérou à la fin du XVIIIème siècle, Etats-Unis, année par année de 1914 à 1919 et de 1924 à 1928. Tous ces coefficients se situent entre 1, 24 et 1, 79.

II. — FORMULES POUR LA DISTRIBUTION NON TRONQUÉE.

Modèles.

Pareto a proposé :

$$y = Ae^{-Bx}(x - c)^{-\nu}$$

Formule d'Amoroso (1925) :

$$y = Ae^{-\gamma(x-a)^{1/s}}(x - h)^{\frac{p-s}{s}}$$

Formule de D. G. Champernowne :

$$y = \frac{A}{B\mathrm{ch}\,(\log x - C) - D}$$

Formule de Gibrat :

$$y = \frac{N}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2}\left(\frac{a\,\mathrm{Log}\,(x-k)\,+\,b}{\sigma}\right)^2}$$

mais cette dernière ne satisfait pas la loi de Pareto.

L'auteur propose :

$$\frac{dy}{dx} = \frac{a}{Z^n}\,\frac{1}{(e^{b/z} - 1)},$$

où :

$$z = x - c$$

c est la valeur des plus faibles revenus observés.

Estimation.

a et n sont obtenus par la droite de Pareto; b en identifiant les modes empirique et théorique, puis par une formule donnée par l'auteur :

$$b = z_0 (n - p)$$

où z_0 est le mode,

p est la solution non évidente de l'équation :

$$pe^{-p} = ne^{-n}$$

Application.

Application aux mêmes données que précédemment.

III. — LOI DES INÉGALITÉS GÉNÉRALISÉES.

Cette loi, proposée par Carl Snyder (1936) énonce que sous les hypothèses :

1) La variable x est la mesure d'une capacité mesurable d'un individu d'un groupe de taille N_1.

2) Le domaine de répartition de la variable x peut être considéré comme infini d'un côté.

3) Une unité de x est comparable avec toute autre unité (un dollar est un dollar, pour un bas revenu comme pour un revenu élevé).

Alors la distribution de x suivra approximativement la loi de Pareto pour x suffisamment grand.

L'auteur donne quelques exemples obéissant à cette loi :

— Distribution de scores au billard réalisés par des membres de l'université d'Indiana.

— Distribution du nombre de contributions à la littérature scientifique, par auteur (3 exemples sont donnés, extraits des travaux de Lotka, 1926).

— 278 auteurs en mathématiques.

Des nombres très supérieurs en chimie et en physique.

L'auteur conclut en supposant que la raison de la distribution de Pareto « doit être cherchée dans le mystérieux royaume de la psychologie humaine».

BIBLIOGRAPHIE. 42 titres, dont :

AMOROSO, Ricerche intorno alla curva dei redditi, *Annali di matematica*. Vol. 2, 4th series, 1924-1925, p. 123-160.

GIBRAT, *Les inégalités économiques*, Paris, 1931.

PARETO, *Cours d'économie politique*, Lausanne, 1897.

PARETO, *Manuel d'économie politique*, Paris, 1909.

LOTKA, The frequency distribution of scientific productivity. *J. of the Washington academy of Sciences*. Vol. 16, 1926, p. 317-323.

James S. Coleman & John James, "The equilibrium size distribution of freely-forming groups", *Sociometry*, Vol. 24, n° 1 (1961), p. 36-45.

Modèle.

Un système de groupes est supposé vérifier les hypothèses suivantes :

1) L'état d'un groupe est le nombre i d'individus qui le composent.

2) Pour chaque état, il y a une probabilité constante de perte d'une unité, proportionnelle au nombre de membres du groupe. A la limite le processus est continu par rapport au temps et la probabilité de perte d'un membre est :

$$dp\ (i, i - 1) = i\ \beta\ dt,$$

β étant le « coefficient de perte ».

3) Chaque groupe a une probabilité constante de gagner un membre, proportionnelle au nombre n_1 de groupes à l'état 1. A la limite, la probabilité pour un groupe de gagner un membre est :

$$dp\ (i, i + 1) = n_1\ \alpha\ dt,$$

où α est constant.

On en déduit, lorsque le système est à l'état stationnaire, que si p_i est la probabilité qu'un groupe soit à l'état i :

$$p_i = \frac{k^i}{i\,!\ (e^k - 1)} \qquad \text{où} \qquad k = \mathrm{Log}\left(\frac{n\,\alpha}{\beta} + 1\right)$$

n est le nombre total d'individus.

C'est une loi de Poisson tronquée, privée de la valeur 0.

Estimation.

On a $\mu = p_1 + k$, où μ est la moyenne théorique. En identifiant μ et p_1 avec les observations empiriques correspondantes, on obtient une estimation de k.

Test du χ^2 à un seuil non précisé.

Application à 23 ensembles de données, dont 5 sont précisés dans l'article, par des tableaux numériques.

1) Piétons, un matin de printemps, à Eugène (Oregon).
2) Groupes d'acheteurs, printemps, Eugène, grand magasin et marché public.
3) Groupes de joueurs, printemps, Eugène, terrain de jeu public A.
4) *Id.*, terrain de jeu public D.
5) Attroupements publics, printemps, Portland (Oregon), plage publique de l'étang de natation.

Sur 4 des 23 exemples, dont l'exemple 5 ci-dessus, la quantité-test est suffisamment grande pour dire que les différences entre valeurs théoriques et observées peuvent difficilement être dues au hasard. Les probabilités des 19 autres quantités-tests, d'être excédées par les variables du χ^2 correspondantes, sont distribuées à peu près uniformément sur le segment [0,1].

RÉFÉRENCES BIBLIOGRAPHIQUES.

W. FELLER, *An introduction to probability and its applications* 2nd edition, New York, Wiley 1957.

John JAMES, "The distribution of free-forming small groups size", *American sociological review*, 1953, Vol. 18, p. 569-570.

H. A. SIMONS, "On a class of skew distribution functions", *Biometrika*, Vol. 42, 1955, p. 425-440.

Cet article est commenté, et les hypothèses aboutissant au modèle de la loi de Poisson tronquée sont généralisées dans :

H. WHITE, "Chance models of systems of casual groups", *Sociemetry*, Vol. 25, 1962, p. 153-176.

LOI DE LAPLACE-GAUSS | *RECHERCHE OPÉRATIONNELLE*

LOI DE POISSON | *TRAFIC URBAIN*

Leslie C. EDIE, "Traffic Delays at Toll Booths", *Operations Research*, Vol. 2, 1954. p. 107-138.

I. — ANALYSE DES ARRIVÉES.

Description du modèle.

Le nombre x de véhicules arrivant à un poste de péage dans l'agglomération de New York est aléatoire. Des comptages du nombre de véhicules arrivant durant un intervalle de 30 secondes, ont été faits au tunnel Lincoln et au pont Georges-Washington. Les courbes de distribution des fréquences sont tracées, chacune correspondant à un volume horaire d'arrivées.

L'auteur compare dans divers cas, deux modèles :

Distribution de Poisson :

$$P(x) = \frac{e^{-m} m^x}{x!} \qquad x \text{ entier positif ou nul.}$$

Distribution normale :

$$F(x) = \frac{e^{-\frac{(x-m)^2}{2S^2}}}{S\sqrt{2\pi}}, \quad x \text{ réel,}$$

où « $F(x)$ est la probabilité de l'arrivée de x véhicules dans un intervalle ».

L'auteur remarque que ces distributions sont toutes deux des approximations de la distribution binomiale :

$$P(x) = C_x^n \, p^x \, q^{n-x}$$

où n, nombre de véhicules, est grand et p est petit. Cependant on s'intéresse aux lois de Poisson et normale, car elles se manient plus aisément. L'approximation de Poisson doit alors être la meilleure pour les faibles trafics horaires et la normale pour les plus forts.

Estimation des paramètres, ou ajustement.

Non précisé.

Test de convenance du modèle.

Test du χ^2 au risque 0,05 proposé. L'auteur présente par ailleurs des graphiques dans des cas où la loi de Poisson semble l'emporter sur la loi normale, puis dans des cas contraires. Avec des exceptions, l'approximation normale l'emporte sur celle de Poisson pour les fortes intensités de trafic. Cependant pour les forts volumes de circulation, l'ajustement se détériore pour les deux lois, tout en restant acceptable au seuil de 0,05. La branche droite de la distribution se trouve alors au-dessus des valeurs théoriques.

II. — ANALYSE DE LA QUEUE ATTENDANT AUX GUICHETS.

Mesurée toutes les 30 secondes pendant 20 minutes. On se retrouve devant les modèles de la partie 1. Ici la loi de Poisson l'emporte nettement.

RÉFÉRENCES BIBLIOGRAPHIQUES.

BROCKMEYER, HOLSTROM & JENSEN, The life and Work of *A. K. Erland Copenhagen Telephone C°*, Copenhagen, Denmark, 1948.

G. S. BERKELEY, *Traffic and Trunking principles in automatic Telephony*, London, Ernest Benn Ltd, 1949.

C. D. CROMMELIN, P. O., *Elec. Eng., J.*, 24 Pt 4, Janv. 1934.

B. D. GREENSHIELDS, D. SHAPIRO & E. L. ERICKSON, *Traffic performance at Urban street Intersections*, Yale University, 1947.

LOI EXPONENTIELLE — *HYDROLOGIE* / *CALCUL ÉCONOMIQUE*

D. Van DANTZIG, "Economic decisions for flood preventions", *Econometrica*, Vol. 24, n° 3, July 1956, p. 276-287.

Modèle.

Wemelsfelder a étudié la distribution des crues marines à Hoek-van-Holland et plus haut dans les estuaires zélandais durant la période 1888-1937. Pour le premier cas, le tracé sur papier semi-logarithmique semble indiquer que :

$$p(h) = c\, e^{-h}$$

où $p(h)$ est la probabilité que la hauteur d'une crue soit supérieure à h.

Estimation.

La méthode n'est pas rapportée ici. La droite de Wemelsfelder est approximativement :

$$h = 4{,}40 - 0{,}80 \log_{10} p$$

(où h est mesuré en mètres), d'après la figure qui la représente.

L'article ne comporte pas de *test* de convenance du modèle, mais précise que l'hypothèse de Wemelsfelder a été soigneusement analysée depuis 1953, sans que l'on ait trouvé de déviation significative. Cependant, les plus fortes crues ont une tendance persistante à être déviées vers la droite de la droite de Wemelsfelder, ce qui peut faire penser qu'elles sont dues à des tempêtes d'un type particulier.

On a essayé également l'ajustement à une loi log-normale et aussi à une distribution de Gumbel (non précisée) des « années maxima ». Le premier a été bien moins bon que celui de la loi exponentielle. Quant à la loi de Gumbel, elle est ajustée sur un nombre inférieur de données et ne peut certainement pas donner une meilleure approximation que l'exponentielle.

RÉFÉRENCES BIBLIOGRAPHIQUES. Quinze titres dont :

D. Van Dantzig, *Voolopige oplossing van het investerings decisieprobleem van de Delta-commissie*, Repart 1953-32 (3) of the Stat. Department of the Math. Centre, Amsterdam.

E. D. Gumbel, "Statistical theory of Extreme Values and some Practical Applications», U.S. Department of Commerce, Nat. Bureau of Standards, *Applied Math.*, Series nº 33.

Wemelsfelder, "Wetmatigheden in het optreden van stormwloeden", *De ingenieur* 3, Maart 1939.

Wemelsfelder, "Frequentielijnen van hoogwater intul Nederlandse kustgebied Rijkswaterstaat", *Directie Algemene Dienst-Hydrometische Afdeling*, 1954.