MATHÉMATIQUES ET SCIENCES HUMAINES

B. LECLERC

Applications pratiques des lois de probabilité

Mathématiques et sciences humaines, tome 36 (1971), p. 79-89
<http://www.numdam.org/item?id=MSH_1971__36__79_0>

Math. Sci. hum. (9e année, n° 36, 1971, p. 79-89)

APPLICATIONS PRATIQUES DES LOIS DE PROBABILITÉ

par

B. LECLERC

LOIS DE PARETO
LOI LOGNORMALE
LOI COMPOSITE
LOI DOUBLE EXPONENTIELLE
LOI DE BENINI

GÉOGRAPHIE
ÉCONOMIE

QUANDT E. E., "Statistical discrimination among alternative hypotheses and some economic regularities", *Journal of regional science*, vol. 5, n° 2, 1964, pp. 1-23.

De nombreuses distributions de grandeurs socio-économiques positives possèdent une propriété commune : très dissymétriques, elles « s'étirent vers la droite ». C'est par exemple, le cas de la distribution des revenus dans une région donnée à une époque fixée, ou des tailles des villes dans une aire géographique donnée. Ces distributions ont été ajustées à diverses lois, les plus courantes étant les lois de Pareto et la loi lognormale, et des modèles ont été proposés pour rendre compte de leur forme générale.

Le propos de cet article est de comparer les ajustements de ces lois à certaines distributions observées, avec une préoccupation méthodologique : l'auteur s'intéresse aux méthodes d'ajustement et de test de convenance de celui-ci. Trois critères autres que la qualité de l'ajustement sont donnés pour le choix d'une fonction mathématique représentant une distribution observée : ce sont le réalisme des hypothèses aboutissant à la forme mathématique, la facilité des calculs analytiques et la signification économique des paramètres.

Modèles

Huit formes mathématiques sont proposées pour les fonctions de répartition $F(x)$ des distributions étudiées. $F(x)$ est la probabilité d'une observation inférieure à x, pour tout x réel positif.

a) Loi de Pareto de première espèce, de coefficient de Pareto $a \geqslant 1$.

$$F(x) = 1 - \frac{K}{x^a}$$

K est une constante, et la loi n'est vérifiée que pour x suffisamment grand.

b) Loi de Pareto de seconde espèce de paramètres K, c, a.

$$F(x) = 1 - \frac{K}{(x + c)^a}$$

c) Une forme simplifiée de la loi de Pareto de troisième espèce (que Pareto lui-même jugeait rarement meilleure que les précédentes), souvent nommée loi de Champernowne, de paramètres K, a, b.

$$F(x) = 1 - \frac{K\, e^{-b\, x}}{x^a}.$$

d) Loi exponentielle de moyenne a, prise comme exemple de distribution présumée ne s'ajustant pas bien au type de données étudiées.

$$F(x) = 1 - e^{-a\, x}.$$

e) Loi lognormale à trois paramètres t, μ, σ définie pour $x \geqslant t$.

$$F(x) = \int_t^x \frac{1}{(\xi - t)\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2}\left(\frac{\log(\xi - t) - \mu}{\sigma}\right)^2} d\,\xi$$

$\log(x - t)$ suit une loi normale $N(\mu, \sigma)$.

f) Loi de Benini, mentionnée par Pareto comme parfois satisfaisante, de paramètres a et b.

$$F(x) = 1 - a\, e^{-b(\log x)^3}.$$

g) Loi composite (dénomination de l'auteur) de paramètres m, K, a.

$$F(x) = e^{-m\left(\frac{K}{x}\right)^a}.$$

h) Loi double exponentielle (dénomination de l'auteur) de paramètres m, K, a, b.

$$F(x) = e^{-m\frac{K}{x^a}\, e^{-bx}}$$

Divers modèles ont été proposés, aboutissant à certaines de ces distributions. C'est ainsi que Champernowne (1953) a proposé, entre autres, un processus aboutissant, pour la distribution des revenus, à la loi de Pareto de première espèce. Les revenus sont répartis en classes, les bornes inférieures et supérieures de chaque classe étant liées par un facteur de proportionnalité constant pour les classes. Champernowne postule des probabilités de transition inter-classes constantes dans le temps, un individu ayant un revenu appartenant à une classe ne pouvant passer que dans la classe supérieure ou dans une classe inférieure suffisamment proche. Les transitions ne dépendent pas de la classe de départ, mais du nombre de classes franchies. Kalecki (1945) aboutit à la loi lognormale en supposant que le revenu est le produit d'une suite de variables aléatoires indépendantes et en appliquant les théorèmes limites centraux.

Les lois *g*) et *h*), appliquées à la taille de firmes industrielles par exemple, sont dérivées des lois *a*) et *c*) respectivement. Si on suppose qu'une activité industrielle composite compte r firmes, chacune exerçant une activité industrielle simple et étant la première dans cette activité, pour laquelle la taille des firmes est distribuée suivant la loi de Pareto (l'auteur reconnaît que ces hypothèses sont hautement artificielles), on aboutit à la distribution composite pour la taille des firmes dans l'industrie composite (en remplaçant la loi de Pareto par celle de Champernowne, on obtient la loi double exponentielle).

Estimation

Quatre méthodes sont d'abord utilisées, selon la loi ajustée.

A) Méthode des moments : on identifie les moments observés et les moments théoriques (il est nécessaire que ceux-ci existent) pris en nombre suffisant (égal au nombre de paramètres à estimer).

B) Méthodes des quantiles, analogue à la précédente en prenant des quantiles convenablement choisis (toujours définis) au lieu des moments.

C) Méthode du maximum de vraisemblance.

D) Ajustement d'une droite par la méthode des moindres carrés au tracé de la fonction de répartition observée avec des échelles de coordonnées convenables. L'auteur ne détaille généralement pas le choix de celles-ci. Les méthodes d'estimation pour chaque loi sont les suivantes :

a) Loi de Pareto : deux lois théoriques obtenues par les méthodes C et D. Pour celle-ci, on a :

$$\log (1 - F(x)) = \log K - a \log x.$$

b) Loi de Pareto de seconde espèce : mélange des méthodes A et B.

c) Loi de Champernowne : méthode D, à partir de :

$$\log (1 - F(x)) = \log K - b\,x - a \log x.$$

d) Loi exponentielle : méthode des moments, c'est-à-dire identification de a à la moyenne observée.

e) Méthodes mixtes non précisées dans l'article mais décrites par Aitchison et Brown (1963) pour la loi lognormale.

f) et *g*) La distribution composite et la double exponentielle sont ajustées par une méthode des moindres carrés.

Enfin, l'auteur propose une méthode d'estimation applicable à toutes les lois et sans biais. On cherche les valeurs des paramètres minimisant

$$S = \sum_{i=1}^{n+1} (F(x_i) - F(x_{i-1}) - \frac{1}{n+1})$$

où $x_1, \ldots, x_i, \ldots, x_n$ sont les observations, $F(x_0) = 0$ et $F(x_{n+1}) = 1$.

La variable $F(x)$ est en effet théoriquement uniformément répartie et l'espérance de $F(x_i) - F(x_{i+1})$ est $\frac{1}{n+1}$.

Cette estimation est faite par une méthode de gradient numérique. Les lois exponentielles et de Bénini ne sont pas étudiées de cette façon, et la loi lognormale est ici à deux paramètres : on prend $t = 0$. Les résultats en général ne sont pas exacts : les itérations sont arrêtées pour S suffisamment petit et on n'est pas sûr que le minimum de S ainsi approché est absolu.

Tests de convenance des modèles

L'utilisation du test du χ^2 est rejetée pour diverses raisons : le groupement des observations en classes est arbitraire, le jugement de l'ajustement n'est pas fait en fonction d'une ou de plusieurs alternatives précises à la loi testée, enfin et surtout on s'intéresse beaucoup ici à la distribution des plus grandes valeurs des variables étudiées. Pour celles-ci il y a relativement peu d'observations et le regroupement en classes d'effectif suffisant exigé par le test du χ^2 ne permet pas une bonne discrimination. L'auteur préfère donc le test de Kolmogorov-Smirnov, pour lequel la quantité test est :

$$D = \max_i | S(x_i) - F(x_i) |$$

où F est la fonction de répartition ajustée, S celle observée, $x_1, \ldots, x_i, \ldots, x_n$ sont les observations. Ce test n'est pas non plus pleinement satisfaisant. Ainsi, les valeurs critiques de la quantité-test (au delà desquelles l'ajustement doit être rejeté) sont inconnues lorsque, comme ce sera le cas ici, les paramètres sont estimés à partir des observations. La statistique D sera donc ici purement descriptive, sans interprétation probabiliste.

L'auteur compare également pour certaines lois, les coefficients de Lorenz théoriques et observés. La courbe de Lorenz est obtenue en portant F (x) en abcisses et :

$$F_1(x) = \frac{\int_K^x \xi \, \mathrm{d}\, F(\xi)}{\int_K^\infty x \, \mathrm{d}\, F(x)}$$

en ordonnées, où K est la borne inférieure de x. Le coefficient de Lorenz L est alors :

$$L = 1 - \int_K^\infty F_1(x) \, \mathrm{d}\, F(x).$$

Le coefficient de Lorenz varie entre 0,5 et 1, et est une mesure d'inégalité : L est d'autant plus grand qu'une masse plus importante de la variable est répartie sur un effectif plus restreint de la population.

Pour les distributions estimées par minimisation de la statistique S, des valeurs critiques de S ont été obtenues par des méthodes de simulation. Un second test est fondé sur l'idée que les résidus :

$$F(x_i) - F(x_{i-1}) - \frac{1}{n+1}$$

doivent être « au hasard ». On fait donc un test non-paramétrique classique en cherchant si le nombre de suites de résidus positifs ou négatifs (statistique Z) est acceptable. L'auteur fait également des « analyses spectrales » dont il ne détaille pas les calculs et l'interprétation.

Applications

Sept distributions sont étudiées.

1) Population des villes des États-Unis de plus de 50 000 habitants.

2) Mêmes populations diminuées de 50 000.

3) Distribution du nombre de firmes d'une valeur supérieure à un million de dollars par industrie (définie selon une classification standardisée), fondée sur les données de Dun et Bradstreet pour le "Million Dollar Directory" de 1963.

4) Nombre de firmes par industrie (même définition) manufacturière en 1958 (U.S. Bureau of Census).

5) et 6) Chiffres d'affaires des 500 plus grandes firmes des États-Unis en 1955 et 1960, donnés par la revue *Fortune*.

7) Population des pays du monde, d'après Hart et Prais (1956).

Deux tables donnent les résultats du test de Kolmogorov-Smirnov, pour les distributions entières d'abord, pour leur moitié droite ensuite. Les conclusions de l'auteur sont les suivantes :

Les rangements des neuf distributions théoriques (deux pour la loi de Pareto de type 1) par ordre de qualité de l'ajustement ne sont pas indépendants. Le test du coefficient de concordance W de Kendall est significatif au seuil 1%.

Quatre lois donnent le meilleur ajustement pour au moins un cas : la loi de Pareto de première espèce pour la distribution 1, celle de seconde espèce pour les distributions 2 et 7, la loi lognormale pour les distributions 3, 5, 6, la double exponentielle pour la distribution 4. Ces trois dernières lois donnent les meilleurs résultats d'ensemble. La loi exponentielle donne, comme prévu, de mauvais ajustements. Les différences entre le premier et le second tableau (demi-distributions) sont faibles (moins bons rangs pour la loi lognormale).

La loi de Pareto de seconde espèce donne des coefficients de Lorenz toujours plus proches de ceux observés que ceux obtenus à partir de la loi lognormale. La différence reste cependant généralement faible.

Une table donne ensuite les statistiques S. Ici le test W de Kendall n'est pas significatif. Aucune loi théorique étudiée n'est partout acceptable ou partout à rejeter, et toutes sont au moins une fois la meilleure. Pour départager ces lois, le recours à la statistique Z permet de mettre en tête la loi de Pareto de seconde espèce, puis la loi de Champernowne, la moins bonne étant pour ce test la loi lognormale. Les résultats de l'analyse spectrale donnée dans de nombreuses figures, donnent des résultats identiques.

La conclusion générale de l'auteur est que les lois étudiées sont assez voisines. Il observe ainsi que, lorsque l'une d'elles donne un bon ajustement, il en est de même de plusieurs autres.

BIBLIOGRAPHIE

AITCHISON J., et BROWN J. A. C., *The lognormal distribution*, Cambridge, Mass., Cambridge University Press, 1963.

BARTHOLOMEW J., *The meridian compact atlas*, Cleveland, Ohio, World Publishing, 1961,

CHAMPERNOWNE D. G., "A model of income distribution", *Economic Journal*, 63, 1953, pp. 318-351.

COCHRAN W. G., "Some methods for strenthening χ^2 tests", *Biometrics*, 10, 1954, pp. 417-421.

DUESENBERRY J. S., *Income, saving and the theory of consumer behavior*, Cambridge, Mass., Harvard University Press, 1949.

FELLER W., "On the Kolmogorov-Smirnov limit theorems for empirical distribution", *Ann. math. Statist.*, 19, 1948, pp. 177-189.

FRIEDMAN M., *A theory of the consumption function*, Princeton, NJ, Princeton University Press, 1957.

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

GODFREY M. G., GRANGER C. W., et MORGENSTERN O., "The random walk hypothesis of stock market behavior", *Kyklos*, 17, 1964, pp. 1-31.

HART P. E., et PRAIS S. J., "The analysis of business concentration", *J. r. statist. Soc.*, series A, part 2, 1956, pp. 150-191.

JENKINS G. M., "General considerations in the analysis of spectra", *Technometrics*, 3, 1961, pp. 133-165.

KALECKI M., "On the Gibrat distribution", *Econometrica*, 15, 1945, pp. 161-170.

MALKIEL B. G., "Expectations, bond prices and the term structure of interest rates", *Quarterly journal of economics*, 66, 1962, pp. 197-218.

MANDELBROT B., "New methods in statistical economics", *Journal of political economy*, 71, 1963, pp. 421-440.

PARETO V., *Cours d'économie politique*, Genève, Droz, 1964.

QUANDT R. E., "Old and new methods of estimation and the Pareto distribution", *Research memorandum n° 68*, Econometric research program, Princeton University, Princeton, NJ, septembre 1969.

SIEGEL S., *Non parametric statistics*, New York, McGraw-Hill, 1956.

SIMON H. A., et BONINI C. P., "The size distribution of firms", *American economic review*, 48, 1958, pp. 607-617.

U.S. Bureau of Census, *Concentration ratios in manufacturing industries*, Washington, DC, US Government Printing Office, 1963.

ZIPF G. K., *Human behavior and the principle of least effort*, Reading, Mass., Addison-Weysley, 1949.

BUSSIÈRES R., *The spatial distribution of urban population*, International Federation for housing and planning, Centre de Recherche d'Urbanisme, Paris, 1970.

Modèle

On s'intéresse à la distribution spatiale de la population d'une concentration urbaine importante. Celle-ci est supposée avoir un centre ponctuel unique O. La densité de population D (r) à la distance r du centre O suit généralement dans ce cas approximativement une loi exponentielle (Clark, 1951).

$$\mathrm{D}\,(r) = A\, e^{-b\, r}.$$

A est la densité au centre et b un paramètre de décroissance.

A et b sont des constantes : on a alors pour la population totale d'habitation éloignée du centre d'une distance inférieure à r :

$$\mathrm{P}\,(r) = \frac{2\,\pi\, A}{b^2}\,[1 - e^{-b\, r}\,(1 + b\, r)].$$

Ceci équivaut au modèle probabiliste selon lequel la probabilité pour un individu de résider entre les distances r et $r + dr$ du centre est $\Phi\,(r)\,\mathrm{d}r$

$$\Phi\,(r) = b^2\, r\, e^{-b\, r}.$$

C'est la densité d'une loi gamma.

Trois autres modèles simples ont été rejetés :

1) $\mathrm{D}\,(r) = \dfrac{K}{r}$ d'où $\mathrm{P}\,(r) = 2\,\pi\, K\, r + C$

2) Pour $r \leqslant \dfrac{K}{A}$ $\mathrm{D}\,(r) = A - K\, r$ d'où $\mathrm{P}\,(r) = \pi\, r^2\,(A - \dfrac{2\,K\,r}{3})$

Pour $r > \dfrac{K}{A}$ $\mathrm{D}\,(r) = 0$ $\mathrm{P}\,(r) = \dfrac{\pi\, A^3}{3\, K^2}$

3) $\mathrm{D}\,(r) = K$ d'où $\mathrm{P}\,(r) = \pi\, K r^2$ pour $r \leqslant K\, R^2_{\mathrm{lim}}$

$\mathrm{D}\,(r) = 0$ et $\mathrm{P}\,(r) = \pi\, K\, R^2_{\mathrm{lim}}$ pour $r > R_{\mathrm{lim}}$.

Aucun de ces trois modèles ne rend compte des données observées.

Estimation

A et b sont calculés itérativement, pour obtenir le meilleur ajustement possible par la méthode des moindres carrés de la courbe théorique :

$$\mathrm{P}\,(r) = \frac{2\,\pi\, A}{b^2}\,[1 - e^{-br}\,(1 + b\, r)\,]$$

aux points observés.

Test

La qualité des ajustements apparaît sur les graphiques joints à l'article. L'auteur calcule le coefficient de corrélation linéaire r entre les nombres d'habitants observés et ceux théoriques, correspondant aux mêmes distances du centre. r est proche de $+1$ si les deux séries de valeurs tendent à être égales.

Application

Sept agglomérations sont étudiées : Paris (7 600 000 habitants, 1962), Lyon (850 000 habitants, 1962), Marseille (800 000 habitants, 1962), Toronto (1 700 000 habitants, 1961), Toulouse (340 000 habitants, 1962) et Auxerre (28 000 habitants, 1962). Les données accessibles (recensements) sont fournies par zones (par exemple, quartiers et communes de banlieue pour Paris). Pour chaque zone est déterminé un centre ponctuel aussi proche que possible du centre de gravité de sa population. La population toute entière de la zone est alors supposée au centre pour les calculs.

Un centre de l'agglomération pour la loi exponentielle est déterminé, qui ne coïncide pas forcément avec le centre de gravité de la population. L'auteur observe que la densité de population moyenne $D_{O'}(\beta)$ à la distance β d'un point quelconque O′ situé à une distance α du centre O est maximale pour $\beta = \alpha$. En choisissant plusieurs points O′ successifs, on peut donc déterminer O par recoupements. Sur une carte, l'auteur montre les centres de Paris ainsi calculés en 1954 et 1962, ainsi que les centres de gravité aux mêmes dates. Tous ces points sont très voisins (300 m de différence pour une agglomération de 55 km de diamètre) et leur déplacement dans le temps faible. Une autre carte présente le centre calculé de Marseille.

Des graphiques montrent les ajustements obtenus. On a :

Pour Paris,	$A = 54\,892$ habitants/km²	$b = 0{,}211$	$r = 0{,}999$
Pour Lyon,	$A = 27\,613$ —	$b = 0{,}453$	$r = 0{,}996$
Pour Marseille,	$A = 42\,273$ —	$b = 0{,}578$	$r = 0{,}997$
Pour Toronto,	$A = 10\,523$ —	$b = 0{,}186$	$r = 0{,}999$
Pour Toulouse,	$A = 31\,481$ —	$b = 0{,}756$	$r = 0{,}999$
Pour Auxerre,	$A = 9\,176$ —	$b = 1{,}439$	$r = 0{,}989$

Un septième ajustement satisfaisant est obtenu pour la population active du secteur tertiaire de Paris, localisée cette fois en son lieu de travail, et non plus de résidence, en 1962. L'ajustement est encore bon et l'on trouve :

$$A = 46\,478 \text{ travailleurs} \qquad b = 0{,}448 \qquad r = 0{,}999.$$

Ceci suggère la recherche générale des variables géographiques qui suivent la loi exponentielle (on pourrait étudier les prix des terrains, par exemple).

L'auteur note que l'échantillon des villes étudiées est assez varié. Elles ont cependant toutes en commun une formation spontanée, « libre », non planifiée, ce qui situe le modèle exponentiel comme une loi du comportement social des masses. Il semble que cette loi apparaisse bien établie dans de nombreux cas, alors que jusqu'à une époque récente, on la considérait comme une première approximation très simplifiée et imprécise de la réalité.

BIBLIOGRAPHIE

BLUMENFELD H., "The concentric theory of urban growth", *Land economics*, mai 1949 ; "The tidal wave of metropolitan expansion", *J. of the American Institute of Planners*, hiver 1954.

CLARK C., *Population growth and land use*, London, MacMillan, 1967. "Urban populations densities", *J. r. stat. Soc.*, 114, part IV, 1951.

NEWLING B. E., "The mathematics of structure and process", *Annual meetings, Association of American geographers*, avril 12, 1967.

SEYFRIED W. R., "The centrality of urban land values», *Land economics*, 39, pp. 275-284.

STEWART J. Q., et WARNTZ W., "Physics of population distribution", *J. of regional science*, 1, 1958.

WARNTZ W. et NEFT D., "Contribution to a statistical methodology for areal distributions", *J. of regional science*, 2, nº 1, 1960.

LOI DE POISSON
LOI BINOMIALE NÉGATIVE *GÉOGRAPHIE*

KING L. J., *Statistical analysis in geography*, Englewood Cliffs, NJ, Prentice-Hall, 1969.

Dans cet ouvrage, largement consacré par ailleurs aux méthodes d'analyse des données (analyse factorielle, analyse discriminante, etc.), l'auteur présente de nombreux exemples de l'emploi des lois de probabilité dans la recherche géographique. Nous reprenons ici ceux qui sont le plus complètement traités.

Loi de Poisson

La probabilité de l'entier positif x est :

$$\mathrm{p}\,(x) = \frac{e^{-\lambda}\,\lambda^{x}}{x!}, \qquad \lambda > 0.$$

La moyenne *observée m* permet d'estimer λ. La moyenne et la variance théorique sont égales à λ.

En géographie, on est amené à s'intéresser à des distributions spatiales bidimensionnelles. Dacey (1964) a donné quatre hypothèses conduisant à une loi de Poisson pour le nombre de points aléatoires sur une surface donnée.

i) La probabilité qu'une sous-surface contienne x points ne dépend pas de sa place dans l'espace considéré.

ii) La probabilité d'un point dans un élément de surface d A est λ d A, la probabilité de plus d'un point est d'un ordre inférieur.

iii) La probabilité p (x, A) de x points sur une surface égale à A est continument dérivable par rapport à A.

iv) p (0, 0) = 1 ; p $(x, 0) = 0$; p $(x, A) = 0$ pour $x < 0$.

L'auteur mentionne les travaux de Getis (1964) sur la localisation des magasins d'alimentation générale à East Lansing. Entre une période où ces magasins ont tendance à se grouper (peu de supermarchés) et une période où leur distribution tend à être uniforme (beaucoup de supermarchés), l'auteur postule qu'ils peuvent se répartir au hasard, suivant une loi de Poisson. Pour tester cette hypothèse, la cité est divisée en carrés de même surface (quadrats) et le comptage des magasins dans chaque carré est effectué. Les distributions obtenues pour les années 1900, 1910 et 1960 sont « très semblables par leur forme » à celles du modèle poissonnien.

Ce modèle « au hasard » est évidemment une simplification à l'extrême de la réalité. On peut penser que d'autres modèles pourront aussi bien rendre compte des faits observés.

Loi binomiale négative

Pour cette loi, la probabilité p (x) de l'entier positif x est selon Feller (1950) :

$$\mathrm{p}\,(x) = \binom{r + x - 1}{x} p^r (1 - p)^x$$

r est un paramètre entier positif et p est compris entre 0 et 1.

L'estimation de ces paramètres n'est pas simple. Williamson et Bretherton (1963) en ont fait la discussion. Le plus simple est d'utiliser la moyenne et la variance théorique m et s^2 (méthode des moments). On a :

$$p = \frac{m}{s^2} \qquad r = \frac{m\,p}{1 - p}.$$

Si r n'est pas entier (les factorielles sont étendues à ce cas par la fonction eulérienne gamma) d'autres estimateurs sont utilisés, notamment par Anscombe (1950). Ces estimateurs ne sont pas précisés par l'auteur.

Il fait d'abord état des travaux de Rogers (1965) sur la répartition de cinq sortes de magasins dans Stockholm (voir fiche correspondante). Le modèle sous-jacent est celui d'une répartition poissonnienne des « grappes » de magasins et pour le nombre de magasins par « grappe » d'une répartition logarithmique de Fisher, pour laquelle on a :

$$\mathrm{p}\,(x) = \frac{a\,b^x}{x} \qquad \text{avec} \qquad a > 0,\ 0 < b < 1.$$

Harvey (1966), étudiant des phénomènes de contagion dans l'espace a repris ce modèle sur deux séries de données de Hägerstrand (1953). Dans des quadrats sont mesurés dans la première le nombre d'accords d'amélioration des pâturages entre 1928 et 1933, dans la seconde le nombre d'usagers de machines à traire (1944). Les lieux de ces études ne sont pas précisés. L'auteur donne les tableaux des nombres observés et de ceux théoriques correspondant à une estimation par la méthode d'Anscombe.

Le test du χ^2 permet d'accepter l'ajustement, avec un niveau de signification de 0,30 pour la première distribution et de 0,80 pour la seconde. Deux autres distributions de Hägerstrand obtenues par des méthodes de simulation sont également étudiées.

Dacey (1967) a fait le même travail pour 21 échantillons du nombre de maisons par quadrats à Porto Rico, et a obtenu de très bons ajustements, même en regroupant les quadrats en unités plus grandes. Cependant, dans ce cas, le paramètre r devrait croître proportionnellement aux éléments de surface, et p devrait être inchangé dans l'hypothèse d'indépendance entre les quadrats qui est contenue dans les modèles classiques de la loi binomiale négative. Ce n'est pas le cas et Dacey pense qu'un modèle vraiment réaliste est encore à trouver.

Lois en relation avec la loi de Poisson

Dacey (1964) a considéré d'autres modèles tirés de celui de Poisson pour la distribution de points sur un espace bidimensionnel. Tout d'abord, on peut supposer qu'une vaste région homogène est divisée en c comtés de même taille. Un comté a la probabilité p d'avoir un chef-lieu et $1 - p$ de n'en pas avoir.

Le nombre de centres urbains autres que les chefs-lieux que contient chaque comté est distribué suivant la loi de Poisson de moyenne m. On obtient donc comme probabilité pour un comté de contenir x centres urbains :

$$p(x) = (1 - p)\, m^x \frac{e^{-m}}{x!} + p\, m^{x-1} \frac{e^{-m}}{(x - 1)!}$$

La moyenne et la variance théoriques sont $m + p$ et $m + p - p^2$, d'où des estimateurs m et p de m et p par la méthode des moments.

$$\hat{p} = \sqrt{\bar{x} - s^2} \qquad \hat{m} = \bar{x} - \hat{p}$$

où $\bar{x}$ et s^2 sont la moyenne et la variance observées.

Pour $p = 0$, on retrouve le modèle purement poissonnien. Lorsque p croît, la répartition des centres urbains tend à être plus régulière. Pour $p = 1$, $x - 1$ est une variable de Poisson.

Ce modèle a été appliqué au nombre de centres urbains par comté en Iowa, pour les années 1840, 1870, 1900, 1930 et 1960.

Un tableau donne les résultats. Le dernier ajustement (1960) semble moins bon que les précédents.

D'autres modèles théoriques tirés de la loi de Poisson par Dacey sont ensuite mentionnés, sans ajustement à des distributions observées.

BIBLIOGRAPHIE

ANSCOMBE F. J., "Sampling theory of the negative binomial and logarithmic series distributions", *Biometrika*, 37, pp. 358-382, 1950.

CLARK P. J., et EVANS F. C., "Distance to nearest neighbour as a measure of spatial relationships in populations", *Ecology*, 35, pp. 445-453.

CURRY L., "The geography of service centers within towns : The elements of an operational approach", *Lund studies in geography, Series B*, 24, pp. 31-53, 1962.

CURRY L., "The random spatial economy : An exploration in settlement theory", *Annals*, Association of American Geographers, 54, pp. 138-146, 1964.

DACEY M. F., "Two-dimensional random points patterns : A review and an interpretation", *Papers*, The Regional Science Association, 13 pp. 41-55, 1964, a.

— "Modified Poisson probability law for point pattern more regular than random", *Annals*, Association of the American Geographers, 54, pp. 559-565, 1964, b.

— "Order distance in an inhomogeneous random point pattern", *Canadian geographer*, 9, pp. 144-153, 1965.

— "A compound probability law for a pattern more dispersed than random and with areal inhomogeneity", *Economic geography*, 42, pp. 172-179, 1966.

— *An empirical study of the areal distribution of houses in Puerto Rico*, Evanston, Ill., North-western University, 1967 (mimeo).

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

GETIS A., "Temporal land-use patterns analysis with the use of nearest neighbour and quadrat methods", *Annals*, Association of the American Geographers, 54, 391-399, 1964.

HÄGERSTRAND T., *Innovationsfönloppet ur Korologisk synpunkt*, Lund, Gleerups Förlag, 1954.

HARVEY D. W., "Geographical processes and the analysis of point patterns", *Transactions and papers*, Institute of British Geographers, 40, pp. 81-95, 1966.

King L. J., "A quantitative expression of the pattern of urban settlements in selected areas of the United States", *Tijdschrift voor economische en sociale geografie*, 53, pp. 1-7, 1962.

Parzen E., *Modern probability theory and its applications*, New York, Wiley, 1960.

Rogers A., "A stochastic analysis of the spatial clustering of retail establishments", *J. Am. statist. Ass.*, 60, pp. 1094-1103, 1965.

Williamson E., et Bretherton M. H., *Tables of the negative binomial distribution*, New York, Wiley, 1963.