MATHÉMATIQUES ET SCIENCES HUMAINES

B. LECLERC

Applications pratiques de lois de probabilité (7)

Mathématiques et sciences humaines, tome 30 (1970), p. 43-51
<http://www.numdam.org/item?id=MSH_1970__30__43_0>

M.S.H. (8e année, nº 30, 1970, p. 43-51)

APPLICATIONS PRATIQUES DE LOIS DE PROBABILITÉ (7)

par

B. LECLERC [1]

Les modèles étudiés dans les articles résumés ci-dessous ont un trait commun : il s'agit de processus de formation de groupes. On notera le rôle central, souvent sous-jacent, de la loi de Poisson et la diversité des applications de la loi binomiale négative. Finalement, et compte tenu de fiches déjà publiées sur des sujets analogues, on constate que le nombre de lois obtenues est inférieur à celui des modèles proposés : là aussi, il y a un phénomène de regroupement.

LOI DE YULE — *HISTOIRE SOCIOLOGIE*

HORVATH W. et FOSTER Caxton J. — « Stochastic models of war alliances», *J. of conflict resolution*, Vol. 7, nº 2, June 1963, p. 110-116.

Richardson a fait de nombreux travaux sur les statistiques des conflits. Il a notamment conjecturé l'existence d'une loi générale sur la formation des alliances constituées en vue de conflit, ayant observé qu'en Mandchourie en 1935, comme à Chicago dans les années 1920, le nombre d'attaques effectuées par n bandits est à peu près proportionnel à $n^{-2,3}$(loi de Pareto).

Rappoport a suggéré que ceci peut s'expliquer par un processus donnant une distribution de Yule, dont la loi ci-dessus serait une approximation. Les deux lois ne diffèrent notablement que pour les plus faibles valeurs de n. Il convient donc d'ajuster à la loi de Yule les données bien établies pour ces valeurs, ce qui n'est pas le cas pour les exemples précédents. Les auteurs reprennent d'autres travaux de Richardson sur les coalitions de guerre.

Ils supposent donc que la formation des alliances en vue de conflit reste de même nature, qu'il s'agisse d'individus ou de nations, alors que la formation de groupes pacifiques demanderait un autre modèle, comme celui de Coleman et James (fiche correspondante dans *M.S.H.* nº 24) qui aboutissent à une loi de Poisson tronquée.

Modèle.

Les auteurs proposent deux processus différents aboutissant à la loi de Yule.

1. Centre de Mathématique Sociale, E.P.H.E., 6e section.

Modèle ouvert : de nouvelles nations entrent constamment dans les conflits, ce phénomène étant uniforme. Une nation nouvelle a une probabilité constante α de lutter seule et une probabilité $(1 - \alpha)$ de contracter une alliance. Dans ce cas, la probabilité d'entrer dans un groupe de taille n est proportionnelle au nombre de nations appartenant déjà à un groupe de taille n. La probabilité de destruction d'un groupe est constante dans le temps et ne dépend pas de la taille du groupe. L'auteur détaille les calculs permettant, à partir de ces hypothèses, de donner la distribution de la taille des groupes. Si $F(n)$ est le nombre de groupes de taille n, et N le nombre total de groupes :

$$F(1) = \frac{N}{2 - \alpha}$$

$$F(n) = \frac{(1 - \alpha)(n - 1)}{1 + (1 - \alpha)n} F(n - 1)$$

Modèle fermé : on suppose ici qu'un ensemble fixé de M nations est continuellement concerné par les conflits, se partitionnant en alliances suivant les hypothèses du modèle ouvert. Les auteurs ne détaillent pas ce modèle, indiquant que l'on obtient :

$$F(1) = M(1 - \beta)$$

$$F(n) = \frac{\beta(n - 1)}{\beta n + 1} F(n - 1) \qquad \text{avec} \quad \beta < 1$$

Estimation

Pour le modèle ouvert, α est le rapport du nombre de groupes formés au nombre total de nations entrées dans le système.

Pour le modèle fermé, on a :

$$\beta = 1 - \frac{F(1)}{M}$$

Test du χ^2.

Application

Application aux données de Richardson, qui a recensé 91 guerres ayant fait chacune plus de 3 160 morts entre 1820 et 1939, d'où la donnée de 182 alliances. Pour le modèle ouvert, la probabilité que la quantité-test soit dépassée par la variable du χ^2 correspondante est 0,7. L'ajustement est donc très bon. Il l'est moins pour le modèle fermé, où l'on ne trouve plus que 0,2. Les auteurs calculent ensuite les nombres théoriques de guerres opposant k nations à l autres, dans l'hypothèse d'indépendance de k et l. Les chiffres obtenus sont pratiquement identiques à ceux observés (tables incluses dans l'article). Il n'y a donc pas tendance à un équilibre ou à un déséquilibre du nombre de nations engagées de part et d'autre.

RÉFÉRENCES BIBLIOGRAPHIQUES

Coleman J. S., James J. — « The equilibrium size distribution of freely-forming groups », *Sociometry*, 24, 1961, p. 36-45.

Rappoport A., Lewis F. — « Richardson's mathematical theory of war », *Journal of conflict resolution*, 1, 1957, p. 249-99.

RICHARDSON L. F. — *Statistics of deadly quarrels*, Pittsburgh, Boxwood Press, 1960.
SIMON H. A. — « On a class of skew distributed functions», *Biometrika*, 42, 1955, p. 425-40.
WHITE H. — « Chance models of systems of casual groups», *Sociometry*, 25, 1962, p. 153-72.

LOI DE POISSON
DOUBLE LOI DE POISSON *ÉCONOMIE*
LOI BINOMIALE NÉGATIVE
LOI DE NEYMAN (TYPE A)

ROGERS Andrei. — « A stochastic analysis of the spatial clustering of retail establishments», *J. of the Am. Stat. Ass.*, 60, 1965, p. 1094-1103.

Les magasins proposant les mêmes produits dans une même ville ne sont pas habituellement dispersés sur toute la superficie de celle-ci, mais ont tendance au contraire à se concentrer en groupes. C'est ce phénomène qui est étudié ici. Pour ce faire, la ville est divisée en lots de superficie égale et le nombre de magasins par lot est observé.

Modèles

Soient les probabilités $p(n)$ d'observer n groupes de magasins dans un lot, $g(s)$ d'observer s magasins dans un groupe, $f(s)$ d'observer s magasins dans un lot.

On suppose que $p(n)$ suit la loi de Poisson de moyenne m

$$p(n) = e^{-m} \frac{m^n}{n\,!}$$

Plusieurs modèles sont proposés pour $g(s)$, et donc $f(s)$:

1) $g(s)$ suit la loi logarithmique (loi de Fisher)

$$g(s) = \frac{ax^s}{s} \qquad \text{avec} \quad a > 0, \quad 0 < x < 1$$

on trouve pour $f(s)$ la loi binomiale négative :

$$f(s) = \frac{(am + s - 1)\,!\, x^s\, (1 - x)^{am}}{s\,!\, (am - 1)\,!}$$

(où les factorielles sont généralisées par la fonction gamma si am n'est pas entier).

2) $g(s)$ suit la loi de Poisson de moyenne v :

$$g(s) = e^{-v} \frac{v^s}{s!}$$

alors $f(s)$ suit la loi de Neyman de type 1 (type A, pour d'autres auteurs) :

$$f(s) = e^{-m} \frac{v^s}{s!} \sum_{t=0}^{\infty} (m\, e^{-v})t \frac{t^s}{t!}$$

de moyenne mv et de variance $mv(v + 1)$.

3) Thomas, pour l'écologie, a suggéré que $s + 1$ suit la loi de Poisson :

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

on aboutit à la « double loi de Poisson » :

$$f(s) = \sum_{n=1}^{s} \left(e^{-m} \frac{m^n}{n\,!} e^{-nv} \frac{(nv)^{s-n}}{(s-n)!} \right)$$

Estimation

Estimation par la méthode des moments, non discutée dans l'article.

Test du χ^2 au seuil 5%.

Application

Application aux données de Artle ; la superficie du centre de Stockholm a été découpée en 210 carrés de 250 m de côté, sur lesquels ont été effectués des comptages de six catégories de détaillants : magasins d'antiquités, de vêtements féminins, de revêtements de sol, épiciers, marchands de tabacs et de liqueurs. Les résultats des cinq premières catégories sont inclus dans l'article sous forme de figures (courbes). Pour les trois premières séries (marchands de biens), la loi binomiale donne le meilleur ajustement, seul à ne pas être rejeté au seuil 5%. Pour les épiceries et tabacs, les trois ajustements sont acceptables, les meilleurs étant obtenus avec la loi de Neyman pour les premières, et avec la loi double de Poisson pour les secondes.

Pour l'auteur, il est malaisé d'interpréter le fait qu'une distribution s'ajuste bien à la loi binomiale négative : celle-ci peut découler de modèles divers.

RÉFÉRENCES BIBLIOGRAPHIQUES

Une trentaine de titres, dont :

ARTLE R. — *Studies in the structure of Stockholm economy*, Stockholm, The business Research Institute of the Stockholm School of Economics, 1959.

QUENOUILLE M. H. — « A relation between the logarithmic, Poisson and negative binomial series», *Biometrics*, 2, June 1949, p. 162-164.

NEYMAN J. — « On a new class of contagious distributions, applicable in entomology and bacteriology», *Ann. of Math. Stat.*, 10, 1939, p. 35-37.

THOMAS M. A. — « Generalisation of Poisson's binomial limit for use in ecology», *Biometrika*, 36, 1949, p. 18-25.

GRENWOOD M. et YULE G. U. — « An inquiry into the nature of frequency distributions of multiple happenings», *J. Roy. Stat. Soc.*, 83, 1920, p. 255-279.

FISHER R. A., CORBET A. S. et WILLIAMS C. B. — « The relation between the number of species and the number of individuals in a random sample of an animal population», *J. Anim. Ecol.*, 12, 1943, p. 42.

EVANS D. A. — « Experimental evidence concerning contagious distributions in ecology», *Biometrika*, 40, 1953, p. 186-211.

L'auteur étudie la distribution du nombre d'individus d'une espèce animale ou végétale par unité de volume ou de surface sur un grand nombre d'exemples afin de déterminer quels modèles usuels s'appliquent le mieux.

Modèles

Trois modèles typiques de distributions obtenues par un phénomène de contagion, proposés par de nombreux auteurs (Anscombe, 1950) sont confrontés. Tous trois ont pour paramètre m moyenne théorique et a tel que la variance théorique est $m(1 + a)$. Ils ont pour limite la loi de Poisson lorsque a tend vers 0. Si P_r est la probabilité d'observer r individus par unité de volume, on a :

Loi binomiale négative :

$$P_r = (1-a)^{-\frac{m}{a}} \left(\frac{a}{1+a}\right)^r \frac{\Gamma\left(\frac{m}{a}+r\right)}{r!\,\Gamma\left(\frac{m}{a}\right)}$$

où Γ est la fonction gamma.

Loi de Polya-Aeppli :

$$P_0 = e^{-\frac{2m}{2+a}}$$

$$P_r = P_0 \left(\frac{4m}{(2+a)^2}\right)^r \sum_{s=0}^{r-1} \frac{\binom{r-1}{s}\left(\frac{a(2+a)}{4m}\right)^s}{(r-s)\,!} \qquad \text{si} \quad r \geqslant 1$$

Loi de Neyman de type A :

$$P_r = \frac{a^r}{r\,!} e^{-\frac{m}{a}} \sum_{j=0}^{\infty} \frac{j^r}{j!} \left(\frac{m\,e^{-a}}{a}\right)^j$$

ou par récurrence :

$$P_{r+1} = \frac{m\,e^{-a}}{r+1} \sum_{u=0}^{r} \frac{a^u}{u!} P_{r-u}$$

Estimation

Deux méthodes sont utilisées pour chaque loi afin d'estimer le couple (m, a).

Méthode 1 : Identification de la moyenne et de la variance avec les données empiriques correspondantes (méthode des moments).

Méthode 2 : Identification de la moyenne et de P_0 avec les données empiriques correspondantes.

Toutes les formules et tables nécessaires sont données dans l'article. Pour la loi binomiale négative, il peut arriver qu'aucune des deux méthodes ne soit suffisamment efficace pour l'estimation de a. On construit un estimateur combiné $\hat{a}_w$:

$$\hat{a}_w = \hat{a}^{(1)} + w(\hat{a}^{(2)} - \hat{a}^{(1)})$$

où $\hat{a}^{(1)}$ et $\hat{a}^{(2)}$ sont respectivement les estimateurs de a des méthodes (1) et (2), et où w est choisi pour minimiser la variance de $\hat{a}_w$.

Tests

Utilisation de la statistique T, différence entre le troisième cumulant empirique et sa valeur théorique calculée à partir des estimateurs de la méthode 1 pour tester l'ajustement obtenu par celle-ci.

Utilisation de la statistique U, différence entre la variance empirique et la variance théorique pour tester l'ajustement obtenu par la méthode 2.

Dans ces deux cas, l'auteur calcule des estimations des écarts-types de T et U, permettant de considérer les valeurs obtenues comme acceptables ou non. Ces méthodes permettent mieux de faire la discrimination entre deux lois que le test du χ^2, qui est aussi utilisé.

Application

Le but de cette analyse étant de chercher à faire un choix entre les distributions théoriques envisagées, il est nécessaire d'avoir un important volume de données. L'auteur dispose des suivantes :

30 comptages d'espèces végétales dans 5 à 34 parcelles en huit points en Angleterre (données de Archibald et de Barnes et Stanbury).

Comptages de 38 espèces végétales dans 40 lots en Nebraska et au nord Dakota (données de Steiger et de Hanson).

11 populations d'insectes (Beall).

2 populations d'œufs d'insectes, comptés par plant de maïs (Marshall).

Toutes ces données, et les résultats de leur étude, sont fournies avec précision par l'auteur dans un grand nombre de tables. Les ajustements sont généralement acceptables. La conclusion générale de l'auteur est que, dans l'ensemble, les comptages de plantes s'ajustent mieux à la loi de Neyman tandis que pour les populations d'insectes, seule convient la loi binomiale négative.

RÉFÉRENCES BIBLIOGRAPHIQUES

ANSCOMBE F. J. — *Biometrika*, 37, 1950, p. 358.
ARCHIBALD E. E. A. — *Ann. Bot., Lond.*, 12, 1948, p. 221.
BARNES H., STANBURY F. A. — *J. Ecol.*, 39, 1951, p. 171.
BEALL G. — *Biometrika*, 30, 1938, p. 422.
BEALL G. — *Ecology* 21, 1940, p. 460.
CLAPHAM A. R. — *J. Ecol.*, 24, 1936, p. 232.
HANSON H. C. — *J. Agric. Res.*, 49, 1934, p. 815.
MARSHALL J. — *Ann. Appl. Biol.*, 23, 1936, p. 150.
MOLINA C. E. — *Poisson's experimental binomial limit*, New York, Van Nostrand, 1947.

RUSHTON S. — « Certain sequential tests of composite hypotheses », London University, Ph. D. Thesis, 1951.

SHENTON L. R. — *Biometrika*, 36, 1949, p. 450.

STEIGER T. L. — *Ecology*, 11, 1930, p. 170.

STIRRET G. M., BEALL G. et TIMONIN M. — *Sci. Agric.*, 17, 1937, p. 587.

LOI BINOMIALE NÉGATIVE LOI DE BATES ET NEYMAN — *DISTRIBUTION DES ACCIDENTS*

MELLINGER G. D., SYLVESTER D. L., GAFFEY W. R., MAHNEIMER D. I. — « A mathematical model with application to a study of accident repeatedness among children », *J. Am. Stat. Ass.*, 60, 1965, p. 1046-1059.

Modèle

Le modèle de Bates et Neyman pour la distribution des accidents repose sur les hypothèses suivantes :

a) Un individu a, dans une période donnée, une propension constante λ d'avoir un accident ;

b) La probabilité $p(k)$ qu'il ait k accidents durant cette période suit la loi de Poisson de moyenne λ

$$p(k) = e^{-\lambda} \frac{\lambda^k}{k\,!}$$

c) La valeur de λ varie d'une personne à l'autre. La distribution de λ dans la population est du type III de Pearson (loi gamma), de densité :

$$p(\lambda) = \frac{\beta^\alpha \lambda^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta\lambda}$$

où $\alpha \geqslant 1$, $\beta > 0$ et Γ est la fonction gamma.

d) Si λ_1 et λ_2 sont les propensions aux accidents d'un même individu sur deux périodes de temps disjointes, on a $\lambda_2 = \gamma\lambda_1$, où γ ne dépend pas de l'individu considéré.

On obtient alors une loi à deux dimensions pour le couple (k, m), où k et m sont les nombres d'accidents par individu observés durant les deux périodes étudiées.

$$p(k, m) = \frac{\Gamma^{(\alpha+k+m)}}{k!m!\Gamma(\alpha)} \beta^\alpha \gamma^m (1 + \beta + \gamma)^{-(\alpha+k+m)}$$

Les hypothèses *a*), *b*) et *c*) ont été proposées par Greenwood et Yule et appliquées avec succès par divers auteurs. Elles aboutissent à la loi binomiale négative pour les distributions marginales de k et m. L'hypothèse *d*) a été émise par Bates et Neyman.

Estimation.

Bates et Neyman ont donné des estimateurs $\hat{\alpha}$, $\hat{\beta}$ et $\hat{\gamma}$ de α, β et γ, avec des indications pour résoudre les équations correspondantes. Si $\overline{X}$ et $\overline{Y}$ sont les moyennes empiriques, on a :

$$\hat{\gamma} = \frac{\overline{X}}{\overline{Y}}$$

$$\log_{10}\left(1 + \frac{\overline{X} + \overline{Y}}{\hat{\alpha}}\right) = \log_{10} e . \sum_{t=0}^{\infty} \frac{1 - \sum_{r=0}^{t} q(r)}{\hat{\alpha} + t}$$

$$\hat{\beta} = \frac{\hat{\alpha}}{\overline{X}}$$

où $q(r)$ est le nombre d'individus ayant eu r accidents au total.

Test du χ^2

Application

Application à quatre groupes d'enfants vivant à Oakland ou à Berkeley (Californie), sur des données médicales (accidents ayant nécessité une intervention médicale).

I — 621 garçons blancs de 4 à 7 ans, puis de 8 à 11 ans
II — 422 garçons blancs de 4 à 7 ans, puis de 12 à 15 ans
III — 537 filles blanches de 4 à 7 ans, puis de 8 à 11 ans
IV — 125 garçons noirs de 4 à 7 ans, puis de 8 à 11 ans.

Les données détaillées sont fournies dans l'article pour le groupe I.

Les probabilités obtenues pour des variables du χ^2 de dépasser les quantités-test (niveaux de signification) sont respectivement 88%, 77%, 45% et 73%. Ces ajustements sont donc bons.

Les auteurs notent qu'il n'est pas possible de distinguer le modèle de Bates et Neyman de certains modèles de contagion (où un accident modifie la propension de l'accidenté). Ils espèrent d'autre part que le succès de l'ajustement de Bates et Neyman pourra servir à construire une bonne manière d'évaluer la dépendance entre les accidents et d'autres variables, cette dépendance étant fréquemment sous-évaluée par le coefficient de corrélation linéaire usuel.

RÉFÉRENCES BIBLIOGRAPHIQUES

Quinze titres dont :

Anscombe F. J. —« Sampling theory of the negative binomial and log-series distributions», *Biometrika*, 37, p. 358-382, 1950.

Arbous A. G. et Kenich J. E. —« Accidents statistics and the concept of accident proneness», *Biometrika*, 7, pp. 340-390, 1951.

BATES G. E. et NEYMAN J. — « Contribution to the theory of accident proneness», *University of California Publications in statistics*, 1, p. 215-275, 1952.

COBB P. W. — « The limit of usefulness of accident rate as a measure of accident proneness», *J. of applied psychology*, 24, p. 154-159, 1940.

GREENWOOD M. et YULE G. U. — « An inquiry into the nature of frequency distributions representative of multiple happenings», *J. Roy. Stat. Soc.*, 83, p. 255-279, 1920.

NEWBOLD E. M. — « A contribution to the study of the human factor in the causation of accidents», *Industrial Health Research Board*, Report 34, London, H. M. Stationery Office, 1926.