MATHÉMATIQUES ET SCIENCES HUMAINES

Problèmes d'enseignement. Formule pour le calcul du nombre de partitions en x classes d'un ensemble de n éléments

Mathématiques et sciences humaines, tome 29 (1970), p. 41-51

<http://www.numdam.org/item?id=MSH_1970__29__41_0>

M.S.H. (9e année, n° 29, 1970, p. 41-43)

PROBLÈMES D'ENSEIGNEMENT

FORMULE POUR LE CALCUL DU NOMBRE DE PARTITIONS EN x CLASSES D'UN ENSEMBLE DE n ÉLÉMENTS

par

P. GORDON

La formule de *P. Gordon* fournissant le nombre de partitions en x classes d'un ensemble de n éléments que l'on trouvera ci-dessous est susceptible de plusieurs interprétations (et par suite, de plusieurs démonstrations) ; en voici une.

Désignons par S (n, x) $(= x\,!\ \mathrm{p}\ (n, x))$ le nombre des surjections d'un ensemble de n éléments sur un ensemble de x éléments ; la formule de P. Gordon s'écrit alors :

$$\mathrm{S}\,(n, x) = x^n - \binom{x}{x-1}(x-1)^{n-1} + \binom{x}{x-2}(x-2)^{n-2} \ldots$$

$\binom{x}{i}$ désigne ici, comme il est usuel, le nombre des parties ayant i éléments dans un ensemble de cardinal x. On peut encore l'écrire, en appelant a (n, x) $(= x^n)$ le nombre des applications d'un ensemble de cardinal n dans un ensemble de cardinal x :

$$\mathrm{S}\,(n, x) = \mathrm{a}\,(n, x) - \binom{x}{x-1}\mathrm{a}\,(n, x-1) + \binom{x}{x-2}\mathrm{a}\,(n, x-2) - \ldots \tag{1}$$

Et en effet, l'ensemble des surjections (de n sur x), c'est l'ensemble des applications moins l'ensemble des applications dont l'image est une partie propre de l'arrivée ; parmi celles-ci... (raisonnez comme pour la célèbre formule dite de G. Boole).

On rapprochera cette expression (1) de celle, classique, qui donne les nombres d'application en fonction des nombres de surjections, et qui est l'inverse de (1).

$$\mathrm{a}\,(n, x) = \mathrm{S}\,(n, x) + \binom{x}{x-1}\mathrm{S}\,(n, x-1) + \binom{x}{x-2}\mathrm{S}\,(n, x-2) + \ldots \tag{1'}$$

On se reportera d'ailleurs avec profit à l'article de G. Kreweras, « Une dualité élémentaire souvent utile dans les problèmes combinatoires » (*M.S.H.*, n° 3, 1963; réédité dans *Cahiers mathématiques III morceaux choisis d'algèbre et combinatoire pour les Sciences Humaines*, Mouton/Gauthier-Villars, 1970).

Si l'on note p (n, x) le nombre de partitions en x classes d'un ensemble de n éléments, on sait que les p (n, x) satisfont la relation de récurrence (valable pour $n <$ et $1 \leqslant x < n$)[1] :

$$p(n, x) = p(n-1, x-1) + x\, p(n-1, x) \tag{1}$$

(cf., par exemple, Mothes & Rosenstiehl, *Mathématiques de l'Action*, Dunod, p. 149).

1. On se propose d'abord de vérifier que l'expression :

$$\sum_{k=0}^{x} (-1)^k \frac{(x-k)^n}{(x-k)!\,k!} \tag{2}$$

satisfait la relation de récurrence (1), pour toutes les valeurs de n et x pour lesquelles celle-ci est définie. A cet effet, nous écrivons le deuxième membre de (1) au moyen de l'expression proposée, soit :

$$\sum_{k=0}^{x-1} (-1)^k \frac{(x-1-k)^{n-1}}{(x-1-k)!\,k!} + x \sum_{k'=0}^{x} (-1)^{k'} \frac{(x-k')^{n-1}}{(x-k')!\,k'!}$$

Posons $k + 1 = k''$ et remplaçons dans la première somme.

Il vient :

$$-\sum_{k''=1}^{x} (-1)^{k''} \frac{(x-k'')^{n-1}}{(x-k'')!\,(k''-1)!} + \quad \text{— id. —}$$

Soit encore :

$$-\sum_{k''=1}^{x} (-1)^{k''} \frac{(x-k'')^{n-1}\,k''}{(x-k'')!\,k''!} + \quad \text{— id. —}$$

Isolant le terme $k' = 0$ dans la deuxième somme, et rapprochant les termes de même rang (1 à x) dans les deux sommes, il vient (k' et k'' pouvant désormais être notés tous deux k) :

$$x\frac{x^{n-1}}{x!} + \sum_{k=1}^{x} (-1)^k \left[\frac{x\,(x-k)^{n-1}}{(x-k)!\,k!} - \frac{(x-k)^{n-1}\,k}{(x-k)!\,k!}\right]$$

$$= \frac{x^n}{x!} + \sum_{k=1}^{x} (-1)^k \frac{(x-k)^n}{(x-k)!\,k!}$$

$$= \sum_{k=0}^{x} (-1)^k \frac{(x-k)^n}{(x-k)!\,k!}$$

C.Q.F.D.

2. Si l'on tient compte du fait que p (n, x) est défini pour $n \geqslant 1$ et pour $1 \leqslant x \leqslant n$, on remarque que la relation de récurrence (1) permet de calculer tous les p (n, x) (pour $n > 1$; $1 < x < n$), dès l'ins-

1. En posant p $(n, 0) = 0$.

tant que sont connus p (1,1) et plus généralement, les p $(n-1)$ $(n \geqslant 1)$, lesquels ne sauraient être calculés par la relation (1), puisque celle-ci ferait alors apparaître des termes non définis.

Or la valeur de ces termes est évidente :

$$\mathrm{p}\,(n, 1) = 1 \qquad \forall\, n \geqslant 1$$

car une seule partition répond à la question, à savoir la moins fine.

Il ne nous reste donc, pour démontrer que l'expression (2) est égale à p (n, x) pour tout $n \geqslant 1$ et pour tout x $(1 \leqslant x < n)$, qu'à nous assurer qu'elle donne bien 1 pour $x = 1$, pour tout $n \geqslant 1$.

Or, pour $x = 1$, k varie de 0 à 1, et (2) se réduit à :

$$(-1)^0 \frac{(1-0)^n}{(1-0)\,!\,0\,!} + (-1)^1 \frac{(1-1)^n}{(1-1)\,!\,1\,!} = 1 - 0 = 1.$$

3. Il est intéressant de remarquer que, de même que la relation de récurrence (1) n'est pas valable pour $x = n$, de même l'expression (2) ne donne pas en général pour $x = n$ la valeur exacte p $(n, n) = 1$ (laquelle traduit l'existence d'une seule partition en n classes d'un ensemble de n éléments : la partition la plus fine).

4. Conclusion : *le nombre de partitions en* x *classes d'un ensemble de* n *éléments* (n $\geqslant$ 1) *est donné par la formule :*

$$\boxed{\mathrm{p}\,(n, x) = \begin{cases} \displaystyle\sum_{k=0}^{x} (-1)^k \frac{(x-k)^n}{(x-k)\,!\,k\,!} & (\text{pour } 1 \leqslant x < n) \\ 1 & (\text{pour } x = n) \end{cases}}$$

M.S.H. (9e année, n° 29, 1970, p. 45-51)

APPLICATIONS PRATIQUES DE LOIS DE PROBABILITÉ

par

B. LECLERC

LOI BINOMIALE *GÉNÉTIQUE*

ROBERTSON A., « The analysis of heterogeneity in the binomial distribution », *Annals of Eugenics*, Vol. 16, n° 1, p. 1-14, 1951-1952.

Modèle.

On considère k groupes d'épreuves susceptibles de donner des « succès » ou des « échecs ». La probabilité p_i de succès est la même pour toutes les épreuves du i-ème groupe. Elle peut varier faiblement d'un groupe à l'autre. Si le i-ème groupe est de taille n_i, la probabilité de m_i succès est donc :

$$\mathrm{I}_{m_i n_i} = \mathrm{C}_{m_i}^{n_i} p_i^{m_i} q_i^{n_i - m_i} = \mathrm{I}_{m_i n_i}(p_i) \tag{1}$$

où :

$$q_i = 1 - p_i$$

Soit p la moyenne des p_i et σ_p^2 leur variance. Pour n épreuves quelconques, on a approximativement pour probabilité de m succès :

$$\mathrm{I}_{mn} = \mathrm{I}_{mn}(p)\,(1 + K_{mn}\,\sigma_p^2)$$

où :

$$K_{mn} = 2\,[m\,(m - 1) + (n \quad m)\,(n - m - 1) - 2\,m\,(n - m)]$$

On cherche un estimateur $\widehat{\sigma_p^2}$ de σ_p^2 pour conclure sur la variabilité de p.

Estimation.

1) Pour $n_i = n$ constant et k groupes, Lexis (1877) a donné :

$$\widehat{\sigma_p^2} = \frac{\Sigma\, m_i\,(m_i - 1)}{k\,n\,(n - 1)} - \left(\frac{\Sigma\, m_i}{k\,n}\right)^2$$

2) Pour n_i variant avec i, la méthode du maximum de vraisemblance à partir de (1) ne donne pas un résultat simple. On a une solution approchée :

$$\widehat{\sigma_p^2} = \frac{\sum\limits_n K_{mn}}{\sum\limits_n K_{mn}^2}$$

L'estimateur sans biais :

$$\overline{\sigma_p^2} = \frac{\sum\limits_n K_{mn}}{\sum\limits_n \dfrac{n\,(n-1)}{2\,p^2\,q^2}}$$

a pour σ_p^2 petit une efficacité voisine de celle de l'estimateur du maximum de vraisemblance. Sa variance est approximativement :

$$\frac{2\,p_2\,q_2}{\sum\limits_n n\,(n-1)}$$

3) A partir des k groupes d'observations, on peut faire un tableau croisé à 2 k cases et calculer le « χ^2 d'hétérogénéité » à $k - 1$ degrés de liberté.

$$\chi^2_{k-1} = \frac{\Sigma \dfrac{m^2}{n} - \left(\Sigma \dfrac{m}{n}\right)^2}{p\,q} = k - 1 + \frac{(k-1)\,(n-1)\,\widehat{\sigma_p^2}}{p\,q}$$

Applications.

1) Rapport des sexes pour 43 000 naissances de bêtes à cornes (Johanson, 1932), regroupées par mères communes. Ces données sont meilleures que celles d'humains pour lesquelles Fisher (1944) a montré que les vrais jumeaux sont des causes d'erreurs.

a) Utilisation de la première méthode pour les « familles » de même taille. On fait ensuite la moyenne pondérée et l'on trouve $\widehat{\sigma_p^2} \simeq 0{,}00112$ avec :

$$\sigma\,(\widehat{\sigma_p^2}) \simeq 0{,}00079.$$

On ne peut conclure à une variabilité de p.

b) Estimateur du maximum de vraisemblance simplifiée. On trouve $\sigma_p^2 \simeq 0{,}00108$ avec :

$$\sigma\,(\sigma_p^2) \simeq 0{,}00080.$$

2) Rapport des sexes chez les humains : données de Gessler, 1889, sur des familles de douze enfants. La méthode 1) donne :

$$\widehat{\sigma_p^2} = 0{,}0026 \text{ avec l'écart-type } 0{,}0003.$$

La déviation est très significative.

Données de Slater (1943), sur 8 433 naissances (psychopathes) sans jumeaux. La méthode du maximum de vraisemblance simplifiée donne :

$$\sigma_p^2 = 0{,}0032 \text{ avec l'écart-type } 0{,}0020,$$

ce qui n'est pas significatif, probablement à cause de l'insuffisance numérique de données.

Données de Wright sur le nombre de petits doigts dans une lignée consanguine de cochons d'Inde. L'existence de σ_p est évidente, d'aprés l'auteur, qui ne développe pas l'analyse de ces données.

RÉFÉRENCES BIBLIOGRAPHIQUES. Quinze titres dont :

FISHER, R. A., *Statistical methods of research Workers*, 10th ed. Edinburgh, Oliver & Boyd, 1944.

GESSLER, A. « Beiträge zur Frage des Geschlechtsverhältnisses des Geborenen », *Z. K. Sächs. Stat. Bur.*, p. 1-24, 1889.

JOHANSSEN, I., « The sex ratio and multiple births in cattle », *Z. Zücht* 24 B, p. 183, 1932.

LEXIS, W., *Zur Theorie der Massenerscheinungen in der menschlichen Gesellschaft*, Freiburg, 1877.

SLATER, E., « A demographic study of a psychopathic population », *Ann. Eugenics*, 12 London, p. 121, 1943.

WRIGHT, S., « An analysis of variability in number of digits in an inbred strain of Guinea pigs », *Genetics* 19, p. 506, 1934.

LOI DE POISSON — *SOCIOLOGIE*
LOI BINOMIALE NÉGATIVE — *PSYCHOLOGIE*

JEFFREYS Harold, « Some Applications of the Method of Minimum χ'^2 », *Annals of Eugenics*, Vol. 11, 1941-1942.

L'auteur s'intéresse à l'ajustement d'une distribution d'observations à une loi théorique par la méthode du χ'^2 minimal. On minimise :

$$\chi'^2 = \Sigma \frac{(m_r - n_r)^2}{n_r} + 2\,M$$

où n_r est le nombre d'observation dans la r-ième classe d'observations si celle-ci n'est pas vide. m_r est le nombre théorique correspondant. M est la somme des valeurs de m_r correspondant aux classes d'observations vides. On peut s'arranger pour supprimer celles-ci et l'on emploie alors la méthode des moindres carrés.

L'auteur considère cette méthode comme une bonne approximation de celle du maximum de vraisemblance et en donne quelques exemples.

1) Nombre de candidates dans un examen de mathématiques de 1910 à 1938 inclus (les guerres perturbent les données masculines).

Il résulte de travaux antérieurs (Jeffreys, 1939) que la distribution est poissonnienne. La moyenne empirique (estimateur du maximum de vraisemblance) est 1,52. La méthode du χ'^2 minimal donne 1,51 pour moyenne théorique.

2) Deux ensembles de données de Miss E. M. Newbold (1926, 1927) sur des accidents en usine. La distribution théorique correspondante est celle de la loi binomiale négative, dont les paramètres n'ont pas d'estimateurs efficaces. La probabilité d'observer r accidents est, pour tout r entier positif :

$$\mathrm{p}(r) = (1 - a)^{\mathrm{p}} \ \frac{1 - a}{a} \frac{a^r}{r\,!}\left(\frac{p}{a} - p\right)\left(\frac{p}{a} - p + 1\right) \dots \left(\frac{p}{a} - p + r - 1\right).$$

La moyenne théorique est p et la variance théorique $\frac{p}{1 - a}$. p et a sont les paramètres que l'on cherche à estimer.

L'auteur estime également les paramètres par la méthode des moments, identifiant p et $\frac{p}{1 - a}$ respectivement à la moyenne et à la variance des observations.

Le test du χ'^2 mesure ensuite la qualité des ajustements. Pour la première série d'observations de Miss Newbold, l'ajustement par le χ'^2 minimal est le meilleur, quoique intrinsèquement peu satisfaisant (la quantité-test est 7,3 avec 4 degrés de liberté). Pour la seconde, les ajustements sont bons avec un léger avantage pour la méthode des moments.

3) L'auteur s'intéresse ensuite à un ensemble de données sur le nombre d'enfants de femmes australiennes, les ajustant à une densité « en ligne brisée ».

RÉFÉRENCES BIBLIOGRAPHIQUES.

CONNIE, L. J. & JEFFREYS, H., *Mon. Not. R. Astr. Soc. Geophysic Suppl.* 3, p. 10-13, 1932.

JEFFREYS, H., *Theory of probability*, Oxford, Clarendon Press, 1939. — *Proc. Comb. Phil. Soc.*, 34, p. 156-157, 1938. — *Proc. Comb. Phil. Soc.*, 33, p. 444-450, 1937.

NEWBOLD, E. M., « Industrial fatigue », *Research Board Report*, 34, 1926. — *J. of the Royal Statistical Society*, 90, p. 487-647, 1927.

POWYS, A. O., « Data for the Problem of Evolution in Man », *Biometrika*, 4, p. 268, 1905.

LOI GAMMA *PHYSIQUE*
LOI LOG-NORMALE

LAKME, C., « Analyse Statistique de la structure locale d'un écoulement diphasé. Description de l'arrivée des bulles », *Revue de Statistique Appliquée*, vol. 13, nº 1, 1965, p. 49-63.

Une émulsion liquide-gaz a été réalisée en injectant de l'air dans un tube en plexiglass où circule un courant d'eau. En un point situé en aval de l'injection, un dispositif permet de mesurer les intervalles entre les arrivées des bulles d'air.

Modèles.

Trois modèles sont successivement appliqués à la loi de ces intervalles.

1) Modèle poissonnien (ou d'Erlang de 1er ordre). La probabilité d'arrivée d'une bulle dans l'in-

tervalle de temps $[t, t + dt]$ est supposée constante. La loi des intervalles est alors exponentielle, de moyenne $m > 0$ et de densité :

$$\frac{1}{m} e^{-\frac{x}{m}}$$

2) Distribution log-normale (ou de Galton-Mac Allister) : les logarithmes des intervalles sont distribués suivant la loi normale de moyenne m et d'écart-type σ, de densité :

$$\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x - m}{\sigma}\right)^2}$$

3) Distribution d'Erlang de second ordre (appartenant à la famille des lois gamma) de moyenne m. La probabilité d'observer un intervalle de durée supérieure à t est :

$$p_t = \left(1 + 2\frac{t}{m}\right) e^{-2\frac{t}{m}}$$

Estimation.

Identification des moments théoriques et empiriques correspondants. Modèles d'Erlang de premier et second ordre : identification de m et de la moyenne des observations.

Modèle log-normal : identification de m et σ respectivement à la moyenne et à l'écart type des logarithmes des observations.

Test.

Test du χ^2 au seuil 0,05 sur 100 observations. Dans les trois cas, il y a 5 degrés de liberté. La valeur du χ^2 correspondant au seuil est 11,07. Les valeurs obtenues sont :

1) 14,37 pour le modèle poissonnien qui doit être rejeté ;

2) 2,47 pour le modèle log-normal qui suppose implicitement des interactions entre les bulles proportionnelles aux intervalles. L'ajustement est donc très bon. La constance du rapport $\frac{m}{\sigma}$ lorsque le débit d'air varie, conduit cependant l'auteur à essayer le troisième modèle, comme ne comportant qu'un paramètre.

3) 3,623 pour le modèle d'Erlang du second type. Cet ajustement est donc également très bon.

L'auteur conclut en estimant que si la distribution log-normale se prête mieux à l'étude graphique, celle d'Erlang du second ordre conviendra mieux aux développements théoriques.

Tous les tableaux de calcul et de nombreux graphiques et figures sont inclus dans l'article qui ne comporte pas de références bibliographiques.

MARITZ, J. S., « On the validity of inferences drawn from the fitting of Poisson and negative binomial distributions to observed accident data ». *Psychological Bulletin*, Vol. 47, 1950, p. 434-443.

La distribution du nombre d'accidents par individus dans une population donnée pendant une période fixée, s'ajuste généralement à une loi de Poisson ou à une loi binomiale négative. Dans le premier cas, on considère que la propension aux accidents est la même pour tous les individus, hypothèse que l'on rejette dans le second cas. L'auteur se propose de montrer que ces conclusions peuvent être chaque fois erronées.

Modèle 1.

Le nombre d'accidents « sérieux » survenus à 122 cheminots est compté durant deux périodes consécutives de cinq mois chacune. La distribution à 2 dimensions obtenue est ajustée au modèle de la loi de Poisson à deux variables corrélées (Aitken) :

$$P(x, y) = e^{-\lambda_1 - \lambda_2 + \lambda_3} \frac{(\lambda_1 - \lambda_3)^x (\lambda_2 - \lambda_3)^y}{x!\, y!} \left(1 + \frac{x\, y}{(\lambda_1 - \lambda_3)(\lambda_2 - \lambda_3)} \frac{\lambda_3}{1!} + \frac{x(x-1)\, y(y-1)}{(\lambda_1 - \lambda_3)^2 (\lambda_1 - \lambda_3)^2} \frac{\lambda_3^2}{2!} + \ldots\right)$$

Cette expression donne pour x et y des lois marginales de Poisson, de paramètres respectifs λ_1 et λ_2.

Estimation.

Les paramètres λ_1 et λ_2 sont estimés d'après les distributions marginales. Puis l'identification du coefficient de corrélation empirique r avec le théorique ρ donne un estimateur de λ_3 par la formule :

$$\rho = \frac{\lambda_3}{\sqrt{\lambda_1 \lambda_2}}$$

Test.

Test du χ^2 pour les distributions marginales et pour la distribution à deux dimensions. L'ajustement est satisfaisant dans les trois cas. En particulier pour la distribution à deux dimensions, la quantité-test a une probabilité 0,49 d'être dépassée par une variable du χ^2.

Les données sont présentées intégralement dans l'article.

Les variables sont donc corrélées ($r = 0{,}29$), ce qui indique que la tendance aux accidents n'est pas la même pour tous les cheminots. Cependant, les distributions marginales sont poissonniennes.

Modèle 2.

Pour le second cas, l'auteur a construit une distribution d'accidents hypothétique pour illustrer sa thèse. Il montre que, bien que cette distribution durant la « période » totale soit binomiale négative, les distributions sur les demi-périodes ont un coefficient de corrélation très faible, en désaccord avec le modèle sous-jacent habituellement à cette distribution.

RÉFÉRENCES BIBLIOGRAPHIQUES.

AITKEN, A. M., *Statistical mathematics*, Edimburgh and London, Oliver and Boyd, 1944.

GREENWOOD, M., « Discussion of paper : Some statistical aspects of road safety research », by R. J. Smeed. *J. Roy. Stat. Soc.*, A, 1949, 112, 25.

MINTZ, A. & BLUM, M. L., « A reexamination of the accident proneness concept », *J. appl. Psychol.*, 1949, 33, 195-211.

NEWBOLD, E. M., « Practical applications of the statistics of repeated events, particulary to industrial accidents », *J. Roy. Stat. Soc.*, 1927, 90, 487.

PEARSON, K., « On certain types of compound frequency distributions in which the components can be individually described by binomial series », *Biometrika*, 1915, 11, 139-144.

(*A suivre.*)