EXTENSION DE L'ANALYSE STATISTIQUE IMPLICATIVE AU CAS DES VARIABLES ALETOIRES CONTINUES A VALEURS SUR [0; 1] ET DE LOI UNIFORME. PREMIERE APPROCHE

Jean-Claude RÉGNIER[1] et Régis GRAS[2]

TITLE

Extension of the Statistical Implicative Analysis for continuous random variables uniform in [0; 1]. First Approach

RÉSUMÉ

L'Analyse Statistique Implicative – ASI a été fondée à son origine sur l'étude de variables binaires réalisées sur des populations finies ou dénombrables pour définir une relation de quasi-implication ou implication statistique. La binarité visait à traduire l'appartenance ou non d'un individu à une catégorie définie par un caractère simple ou composite : 1 correspondant au fait que celui-ci possède ce caractère, 0 quand il ne le possède pas. Une première extension dans le cadre A.S.I. a été réalisée par M. Bailleul, dans sa thèse où il y définit le concept de variable modale en tant que variable ordonnée. Puis par J.B. Lagrange définissant la notion de variables numériques sur [0 ; 1] pour y introduire une relation de propension. Ensuite par R. Gras qui élargit aux notions de variable-intervalle, de variables sur intervalles et de variables floues. Une autre extension a été réalisée par J-C. Régnier et R. Gras à (2005) partir des variables de rang transformées affinement en variables modales introduisant la notion de relation de propension permutationnelle. Ici nous proposons une extension à des variables aléatoires continues à distribution uniforme sur [0 ; 1] dont l'ensemble est structuré par une relation de propension. Nous espérons que cette explicitation théorique incitera à des mises en pratique.

Mots-clés : *variable continue uniforme, variable modale continue.*

ABSTRACT

Keywords : *continuous random variables.*

1 Introduction

Nous avons traité le cas où l'espace des sujets E est continu muni d'une loi de répartition donnée. Nous abordons ici la situation où, cette fois, les variables elles-mêmes sont continues. On la rencontre, par exemple, dans le cas de données sensorielles (variables sensorielles continues). Mais aussi, comme dans (Gras et al, 2001), dans le recueil des notes attribuées à des élèves dans des disciplines variées et conduisant à des intervalles de R.

[1] Laboratoire UMR 5191 ICAR – Université Lumière de Lyon –Lyon2, 86 rue Pasteur, 69635 LYON Cedex 07, e-mail : jean-claude.regnier@univ-lyon2.fr

[2] Ecole Polytechnique de l'Université de Nantes, Équipe Connaissance et Décision, Laboratoire d'Informatique de Nantes-Atlantique (LINA), UMR 6241, Site de la Chantrerie, rue C.Pauc, BP 44306, Nantes cedex 3, e-mail : regisgra@club-internet.fr,

Force est de constater que les extensions de traitements de données dans une approche ASI, développées à partir des premiers travaux de R. Gras, par M. Bailleul (1994) et J.B. Lagrange (1998), considèrent les variables à valeurs sur [0 ; 1] comme des variables discrètes. Ce que nous souhaitons introduire maintenant, est la prise en considération de variables continues à valeurs sur l'intervalle [0 ; 1] ou s'y ramenant. Une extension plus générale conduirait à supposer que toutes les variables sont continues mais de lois différentes. Pour ce faire nous allons, dans un premier temps, étudier le cas des variables continues sur [0 ; 1] dont la distribution de probabilité est uniforme sur cet intervalle et identique pour toutes les variables.

2 Propriétés de la variable Z, produit de deux variables aléatoires continues sur [0 ; 1] de loi uniforme

Nous suivons la même procédure que celle suivie par J.-B. Lagrange à propos des variables modales discrètes. Soient X et Y variables aléatoires continues à valeurs dans [0 ; 1] de loi uniforme continue. Comme il est coutumier en Analyse Statistique Implicative, les deux variables X et Y sont supposées indépendantes. Nous savons alors que la densité de probabilité de X ou de Y est donnée par la fonction $f(u)=1_{[0\,;\,1]}(u)$ de laquelle nous pouvons calculer la fonction de répartition :

$$F(u) = \begin{cases} 1 & u > 1 \\ u & 0 \le u \le 1 \\ 0 & u < 0 \end{cases}$$

Par ailleurs sans difficulté on montre que les espérances respectives de X et Y valent $E(X) = E(Y) = \frac{1}{2}$ et les variances $V(X) = V(Y) = \frac{1}{12}$.

De manière évidente, la variable $\overline{Y} = 1 - Y$ est encore une variable aléatoire à valeur dans [0 ; 1] de loi uniforme continue dont l'espérance et la variance sont celles de Y. Poursuivant le chemin réalisé dans la construction de l'indice de propension établi avec des variables modales, nous posons Z = X(1-Y). Z est encore une variable aléatoire continue à valeurs dans [0; 1] mais qui n'admet plus une densité uniforme. On montre que la densité de probabilité de Z est $g(u) = -\ln(u)\, 1_{]0\,;\,1]}(u)$ ou encore que sa fonction de répartition est :

$$G(u) = \begin{cases} 1 & u > 1 \\ -u\ln(u) + u & 0 < u \le 1 \\ 0 & u \le 0 \end{cases}$$

À partir de ces informations, nous pouvons calculer l'espérance puis la variance de la variable aléatoire continue à valeurs sur [0; 1] de la manière suivante:

$E(Z) = \int_{0+}^{1} uf(u)du = \left[-\frac{1}{2}u^2 \ln(u) + \frac{1}{4}u^2 \right]_{0+}^{1} = \frac{1}{4}$ mais comme Z est le produit de deux variables indépendantes, nous pouvons obtenir directement : $E(Z) = E(X)\, E(1\text{-}Y) = \frac{1}{4}$

Pour ce qui est de la variance, nous savons que :

$$V(Z) = E[(Z-E(Z))^2] = E(Z^2) - [E(Z)]^2$$

Or $E(Z^2) = \int_{0+}^{1} u^2 f(u)du = \left[-\frac{1}{3}u^3 \ln(u) + \frac{1}{9}u^3\right]_{0+}^{1} = \frac{1}{9}$ donc la variance de Z vaut $V(Z) = \left(\frac{\sqrt{7}}{12}\right)^2$

3 Conclusion provisoire

Comme en 2010, lors du bilan du colloque ASI 5 à Palerme, nous adoptons comme défi pour ASI 7, de mener des travaux pour tenter de résoudre, dans le cadre de l'ASI, le problème posé par les variables continues. Il est probable que leur spécificité conduise à utiliser la structure des données en forme d'intervalles de R sur lesquels seraient données des lois de distribution continues des variables, lois peut-être distinctes d'une variable à l'autre. Le travail ci-dessus défriche le terrain puisqu'il conduit à exprimer la loi du produit de variables aléatoires qui est nécessaire pour comparer la contingence et la loi théorique des contre-exemples.

Références

[1] Bailleul, M., (1994). *Analyse statistique implicative : variables modales et contribution des sujets. Application à la modélisation de l'enseignant dans le système didactique*. Thèse Université de Rennes 1

[2] Gras R., Diday E., Kuntz P et Couturier R. (2001), Variables sur intervalles et variables-intervalles en analyse statistique implicative, *Actes du 8ème Congrès de la Société Francophone de Classification, Université des Antilles-Guyane,* 166-173

[3] Gras, R., Régnier, J.-C., & Guillet, F. (Eds) (2009) *Analyse Statistique Implicative. Une méthode d'analyse de données pour la recherche de causalités*. RNTI-E-16. Toulouse : Cépaduès Éditions

[4] Lagrange, J. B. (1998). Analyse implicative d'un ensemble de variables numériques ; application au traitement d'un questionnaire à réponses modales ordonnées. *Revue de Statistique Appliquée*. 46(1) p.71-93

[5] Régnier, J.-C., & Gras, R. (2005) Statistique de rangs et Analyse Statistique Implicative. *Revue de Statistique Appliquée*. 53(1) p.5-38