

L'UTILISATION TRANSVERSALE DE L'ALGÈBRE LINÉAIRE DANS LES COURS DE STATISTIQUE ET DE PROBABILITÉS AU COURS DU PREMIER CYCLE UNIVERSITAIRE

Olivier Renault

Dpt STID UGA - olivier.renault@univ-grenoble-alpes.fr

Résumé. Le formalisme algébrique, introduit dans les cours d'algèbre linéaire en appui des cours de statistique et de probabilités du département de Statistique et d'Informatique Décisionnelle de Grenoble (DUT STID), permet de donner une vision synthétique et cohérente des méthodes de calculs et de programmation. Que ce soit pour le calcul d'une moyenne, de la covariance ou même pour redécouvrir la formule de Bayes, le modèle vectoriel offre un outil et un modèle de pensée structurants. La programmation sous R en tire également de sérieux avantages. Dans cet exposé nous décrirons différents aspects de la transversalité de l'algèbre linéaire, qui sont mis en pratique dans notre formation en statistique et probabilités, tout en revisitant certains concepts.

Mots-clés. Produit tensoriel, formule de Bayes, covariance, ACP, régression linéaire, diagonalisation, forme quadratique, programmation en langage R.

Abstract. The algebraic formalism introduced in the courses of linear algebra in support of statistical and probability courses in the department of statistics and decisional informatics of Grenoble (DUT STID), gives a synthetic and coherent view of the methods of calculation and programming. Whether for averaging, covariance or even rediscovering the Bayes formula, the vector model offers a structuring tool and a model of thought. Programming with this model in mind under R has also significant advantages. In this presentation we will describe different aspects of the transversality of linear algebra, which are put into practice in our formation in statistics and probabilities, while revisiting certain concepts.

Keywords. Outer product, Bayes formula, covariance, PCA, linear regression, diagonalization, quadratic form, R programming language.

1 Le contexte

L'enseignement de la statistique au niveau du premier cycle universitaire s'accompagne inévitablement d'une formation solide en probabilités. Mais quelle place donner à l'algèbre linéaire dans le cadre de cet enseignement? Nous nous plaçons dans le contexte des programmes de la formation STID[1], qui délivre un diplôme universitaire technologique en statistique et informatique décisionnelle. Les programmes du DUT STID donnent l'opportunité d'épauler, sous différents aspects, les cours de statistique et de probabilités par le biais d'un module d'algèbre linéaire en première année et un module d'algèbre

avancée en deuxième année. Nous nous proposons d'observer, au travers des activités mises en place au département STID de Grenoble, comment les outils de l'algèbre linéaire accompagnent ou supportent la construction et la compréhension de la statistique et des probabilités pour les étudiants. Nous souhaitons ainsi mettre en avant les outils vectoriels, porteurs d'une compréhension et d'une expression synthétique de la statistique et des probabilités, une intention que l'on retrouve d'ailleurs dans les logiciels de programmation à vocation statistique. Ces prérequis permettent en outre d'appréhender les modèles linéaires abordés au cours de la deuxième année.

2 Les objectifs

Le concept de vecteur ou de matrice conduit à plusieurs niveaux d'utilisations, allant du simple support informatique de données à la notion plus avancée de transformation. Le passage d'une notion ponctuelle de la donnée, comme le nuage de points, à celle de visualisation vectorielle, puis de transformation linéaire, est une des premières difficultés que rencontre l'étudiant en statistique. Nous soulignerons ce facteur bloquant et nous évoquerons des pistes de progrès en lien avec les programmes du secondaire.

Les objets vectoriels peuvent ensuite être soumis à de multiples opérateurs. Le produit tensoriel uv^T de deux vecteurs u et v , le produit scalaire, le produit matriciel, le produit de Hadamard, ou le produit par composantes sont autant d'outils mis en œuvre tant en statistique qu'en probabilités. Ceux-ci servent non seulement à traduire algébriquement et réaliser les calculs d'indicateurs mais également à donner des clés pour modéliser des problèmes, en donnant dès que possible un sens géométrique. Au-delà, l'enjeu est également de former l'esprit à penser vectoriellement pour exprimer de manière concise des calculs élaborés en évitant l'écueil des boucles informatiques. Les itérations d'une boucle pénalisent souvent le temps de calcul et le potentiel de parallélisation d'une procédure. Nous illustrerons ces outils appliqués en cours, en travaux pratiques et en projets tutorés, tant pour la statistique que pour les probabilités.

3 La démarche proposée

3.1 Modélisation vectorielle en travaux pratiques

Dans une première approche nous portons notre attention sur les structures bivariées rencontrées en probabilités et statistique. Nous insistons sur l'expression en langage informatique de celles-ci lors des séances de travaux pratiques, soit pour les calculs effectifs de probabilités soit en simulation pour des illustrations empiriques. Si nous utilisons le langage R, les activités peuvent être transposées en langages matlab ou scilab. Les lois

conditionnelles dans le cas discret et la notion de covariance sont deux situations que nous détaillerons et qui sont particulièrement appropriées au jeu de l'expression matricielle.

3.2 Création de liens en cours d'algèbre linéaire

Pour approfondir les liens étroits et transversaux de l'algèbre linéaire avec nos cours de probabilités et statistique nous exposons le canevas qui conduit à la construction des méthodes de régression linéaire et d'analyse en composantes principales. Nous partons de l'introduction de l'outil fondamental de produit tensoriel, qui permet non seulement de décrire autrement le produit matriciel, mais inversement de le reconstituer. Enfin, en quelques lignes de programmes, dans une intention pédagogique, nous obtenons le schéma "algébrique" simplifié et explicatif de ces deux méthodes.

3.3 Activité autour de la fonction de répartition empirique

Au travers d'une activité élaborée pour illustrer le cours de probabilités, nous exposons une méthode de construction de la fonction de répartition empirique pour un couple de variables aléatoires. Nous mettons en évidence l'intérêt de la compréhension vectorielle et soulignons à nouveau la pertinence du produit tensoriel pour l'élaboration de cette méthode.

3.4 Un projet tutoré

Les projets tutorés sont l'occasion pour les étudiants d'exprimer un certain degré d'initiative personnelle. Nous verrons, au travers d'un exemple de projet réalisé cette année, comment les étudiants ont abordé une problématique de calculs de probabilités conditionnelles, et la modélisation algébrique adoptée. Le projet a pour vocation de réaliser une interface interactive de calculs de probabilités recouvrant tous les cas de figures qui n'utilisent que deux événements A et B. Ce travail concrétise des résultats exposés progressivement aux 44e JDS 2012 [2] puis au CFIES 2012.

		$P_B(A)$	$P_{\bar{B}}(A)$		
		B	\bar{B}		
$P_A(B)$	A	$P(A \cap B)$	$P(A \cap \bar{B})$	$P(A)$	$P_A(\bar{B})$
$P_{\bar{A}}(B)$	\bar{A}	$P(\bar{A} \cap B)$	$P(\bar{A} \cap \bar{B})$	$P(\bar{A})$	$P_{\bar{A}}(\bar{B})$
		$P(B)$	$P(\bar{B})$		
		$P_B(\bar{A})$	$P_{\bar{B}}(\bar{A})$		

3.5 Redécouvrir la formule de Bayes

Pour finir, en illustrant une fois de plus l'importance de la modélisation vectorielle, nous exposerons une méthode inédite qui permet de redécouvrir la formule de Bayes comme résultat d'une diagonalisation de matrice.

Bibliographie

- [1] Programme pédagogique national du DUT STID (2013). <http://stid-france.fr>
- [2] O. Renault (2012). *Le principe de déconditionnement*. Colloque international : 44e Journées de Statistique, SFdS, 21-25/05/2012, Bruxelles (BE) .