

J.-P. BENZÉCRI

L'appréciation des demandes de crédit : un problème de discrimination et de régression

Les cahiers de l'analyse des données, tome 2, n° 4 (1977),
p. 412-414

http://www.numdam.org/item?id=CAD_1977__2_4_412_0

© Les cahiers de l'analyse des données, Dunod, 1977, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

L'APPRÉCIATION DES DEMANDES DE CRÉDIT : UN PROBLÈME DE DISCRIMINATION ET DE RÉGRESSION [SCORING BANCAIRE]

par J.-P. Benzécri ⁽¹⁾
d'après J. Dehédin

La vente d'une automobile met communément en jeu un crédit de 10.000 NF ou plus. Pour un million de ventes cela fait 10 milliards de NF. Sur cette somme, une perte ou un manque à gagner de quelques millièmes, correspond à des dizaines de millions de NF, ou si l'on préfère à des milliards de centimes! On conçoit donc que l'appréciation (en franglais "*scoring*") des dossiers de demandes de crédit soit de la part des organismes créditeurs, l'objet d'une attention singulièrement diligente. Les règles que les consortiums bancaires recommandent à leurs agents pour susciter, accepter, dissuader ou refuser les demandes, ne sont pas du domaine public. Mais nous savons, d'après ce qu'on chuchote plus encore que d'après ce qu'on dit..., que les praticiens ont perfectionné ces règles grâce à l'analyse statistique des amples données de leurs archives. Bien que ne disposant d'aucune de ces études multidimensionnelles, nous croyons bon d'exposer aux lecteurs des *Cahiers*, ce qui, selon nous, doit en être le principe : en effet, d'une part, il y a là pour quiconque matière à réfléchir ; et d'autre part, quelque spécialiste saisira peut-être cette bouteille que nous jetons à la mer, avec l'espoir de contribuer à l'assainissement du crédit, assainissement auquel est intéressé le corps social tout entier.

Notre exposé ne sera pas purement conjectural : il y a plus de cinq ans, Monsieur J. Dehédin, alors étudiant au D.E.A. de Statistique, nous avait remis une note qui bien que succincte et dépourvue de toute précision indiscrete, suffira pour poser avec exactitude notre problème.

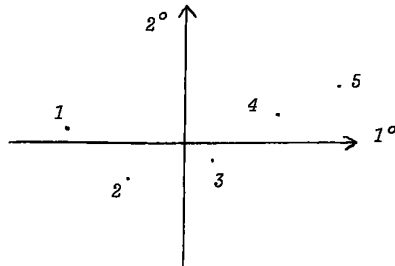
En bref chaque dossier se présente très classiquement comme un ensemble Q de questions q telles que : catégorie socioprofessionnelle, situation de famille, âge, revenu, logement...; pour chaque question q est prévu un ensemble J_q de modalités de réponse : en sorte qu'en notant $J = \{J_q | q \in Q\}$, un ensemble I de dossiers apparaît comme un tableau $I \times J$ sous forme disjonctive complète, avec en ligne les dossiers i , en colonnes les modalités j ; et selon l'usage $k(i,j) = 1$ si le sujet i a fourni la modalité de réponse j et zéro sinon.

A posteriori, l'appréciation d'un dossier est facile à faire : le cas idéal est celui du client qui paie aux dates prévues sans attendre de relance ; suit de près celui qui sur 20 paiements échelonnés n'a commis que des étourderies ; puis viennent les paresseux dont les comptes ne sont sans doute pas tout à fait sains, mais qui cèdent à la lettre recommandée ; puis ceux dont le créancier récupère finalement son dû, mais avec le secours des services spécialisés du contentieux ; enfin restent les insolubles proprement dits, dont après de coûteuses démarches le bien est éventuellement saisi, mais alors qu'il est déprécié par des mois d'usage. De ce point de vue, on peut ranger les dossiers en quelques classes, numérotées par exemple de 1 à 5, du meilleur au pire. Eventuellement, une étude plus fine permettrait de calculer (toujours *a posteriori*, cela s'entend) le pourcentage exact b (positif ou négatif) du bénéfice réalisé par l'organisme créditeur : en effet non seulement chaque lettre recommandée, chaque semonce par un agent spécialisé coûte, mais encore tout décalage en avance ou en retard dans le recouvrement d'une somme se traduit immédiatement en gain ou en perte selon le taux de production d'intérêt de l'argent : répétons-le, ici tout compte : 10% par an, cela fait 1% en trois jours ; et le millième de 10 milliards de NF c'est un milliard de centimes!

(1) Professeur de statistique. Université Pierre et Marie Curie

Le problème statistique est maintenant clair : étant donné un individu dossier i , décrit par une ligne $\{k(i, j) \mid j \in J\}$, de 1 et de zéros, affecter *a priori* cet individu à une classe c ($c \in C = \{1, 2, 3, 4, 5\}$) ou plus précisément estimer un taux de bénéfice b . C'est un problème de discrimination, ou de régression. Nous savons que ce problème se résout le mieux après une analyse de correspondance.

Considérons e.g. l'estimation de la classe c (l'estimation du bénéfice b étant tout analogue) : on construit un tableau de correspondance $k'_{C \times J}$, avec $k'(c, j) =$ nombre des dossiers (recensés e.g. dans un échantillon de 10.000 extrait des archives) qui possèdent la modalité descriptive j (e.g. plus de 4 enfants) et sont rentrés dans la classe de paiement c (dans le cas du bénéfice b , on prendrait à la place de C un ensemble B de niveaux de bénéfice ; avec peut-être 10 niveaux, numérotés du plus positif jusqu'au négatif). L'analyse des tableaux ainsi construits fournit en règle générale dans le plan 1×2 l'ensemble C égrené dans son ordre naturel en un chapelet dont la direction principale d'allongement est le premier axe ; et les modalités de



l'ensemble J se placent à proximité des classes c au pronostic desquelles elles contribuent positivement. Qu'un individu nouveau i se présente : on l'adjoint en ligne supplémentaire au tableau $k'_{C \times J}$: sa place dans le plan 1×2 , ou en termes numériques ses facteurs $F_1(i)$ et $F_2(i)$, sont la meilleure traduction du dossier en vue de son appréciation. Comme en définitive le créancier n'a le choix qu'entre deux réponses : accepter ou refuser le dossier, il suffirait de tracer dans le plan 1×2 une ligne frontière. Toutefois, bien que le calcul des coordonnées $F_1(i)$ et $F_2(i)$ d'après les réponses $k(i, j)$ soit très facile (il suffit d'appliquer la formule de transition dans le cas le plus simple où les $k(i, j)$ valent 1 ou zéro :

$$F_{\alpha}(i) = \lambda_{\alpha}^{-1/2} \sum \{ (k(i, j) / k(i)) G_{\alpha}(j) \mid j \in J \} ;$$

pour un exemple de tel calcul, cf C. Sabaton : *Pollution et Météorologie* ; § 2.2.1 ; Vol. 1 Cahier n° 2) on peut préférer utiliser indirectement les formules pour donner à l'agent qui étudie les dossiers des consignes verbales et non numériques. De plus, il n'est pas absolument exact que le créancier se borne à accepter ou à refuser les clients qui se présentent à lui : bien souvent il les sollicite en fonction de leur profession, du quartier qu'ils habitent, des journaux qu'ils lisent ; parfois il leur propose des conditions de faveur etc. L'outil statistique le mieux adapté à l'étude d'un dossier individuel, et plus encore à la réflexion qui précède l'élaboration d'une doctrine est donc le programme de recherche des plus proches voisins * dans le plan 1×2 (éventuellement l'espace $1 \times 2 \times 3$) : le point i étant placé en $(F_1(i), F_2(i))$, chercher, e.g. les cent dossiers (déjà traités, extraits des archives) qui se projettent le plus près de ce point : on calculera aussi, d'après les pourcentages des classes 1 à 5 dans ce voisinage, des probabilités estimées d'appartenance de i à celles-ci ; on pourra aller jusqu'à tracer dans le plan 1×2 des courbes de niveau ; ou seulement diviser e.g. le plan 1×2 en 25 carrés, et estimer le rendement de chaque carré d'après les dossiers d'archives qui s'y projettent. Ces estimations seront certes imparfaites ; mais répétons-le, sur l'énorme masse en jeu une amélioration minime du rendement produit finalement un bénéfice très important.

Les dossiers présumés très mauvais posent un problème particulier : sur eux, on

(*) cf M. O. Lebeaux, ce cahier, pp 467-481

dispose rarement d'informations *a posteriori* : car étant quasi-toujours refusés, ils n'ont pas la chance d'évoluer jusqu'au terme où l'on pourrait en calculer le coût. Mais, par l'inadvertance d'un agent, certains de ces dossiers passent parfois ; ils fourniront alors dans les zones du plan 1x2 les moins peuplées en individus i acceptés, vers la classe 5, de très utiles termes de comparaison. Sans doute faut-il même conseiller aux praticiens de prendre à l'essai quelques uns de ces dossiers marginaux, situés dans des zones suspectes mais mal connues du plan 1x2, où peut-être une clientèle particulière s'offre à qui saura la traiter avec discernement.

A cet audacieux conseil (peu recevable par qui ne prête qu'au riche) nous bornons nos conjectures, content d'avoir montré au moins dans les grandes lignes, que codage disjonctif, et régression ou discrimination après analyse factorielle peuvent être à l'occasion les outils du chercheur d'or...

P. S. Le manuscrit de cette note ayant circulé parmi les statisticiens qui fréquentent le laboratoire, a suscité des discussions dont voici un écho. Les mauvais payeurs ne sont pas tous des prodiges dépourvus des ressources nécessaires pour subvenir aux nombreuses dépenses où ils s'engagent à la légère. Les crédateurs doivent particulièrement redouter les emprunteurs cyniques qui n'ignorant rien de ce qu'il faut prendre et laisser des menaces imposantes d'une lettre de relance calculent exactement jusqu'où il leur est avantageux de différer un paiement... Les informations matérielles consignées dans un dossier - revenu, âge, charges familiales, logement loué ou acheté - peuvent ne rien laisser deviner de l'habileté dilatoire et de l'absence de scrupules d'un client. Pour apprécier celles-ci il serait bon d'imaginer une épreuve par un questionnaire (un "test"), mêlant des questions de fait (droit civil : un crédateur peut-il saisir...; que coûte un procès...) de psychologie individuelle (aimez-vous à passer une soirée parmi des personnes qui se connaissent entre elles et que vous ne connaissez pas?) de morale civique (est-il admissible d'accélérer quand le feu passe à l'orange?). Un tel questionnaire, même imparfaitement conçu, atteignant la personnalité du débiteur potentiel aiderait à prédire son comportement quand les faits matériels recueillis ne composent qu'un tableau ambigu.