

Les méthodes et les applications du credit-scoring. *(Methods and Applications of Credit Scoring).*

Jean-Marie Bouroche
COREF, Boulogne-Billancourt

Gilbert Saporta, CNAM
Département Mathématiques
Informatique, Paris

En 1976, G. Saporta proposait une méthode originale de discrimination sur variables qualitatives.

Cette technique avait été initialement conçue pour mettre au point des scores de risques bancaires.

Depuis lors, COREF a mis au point un processus opérationnel de calcul de scores et plus d'une centaine d'études ont été réalisées.

Dans notre exposé, nous dressons tout d'abord un bilan des différentes techniques statistiques utilisables et nous en évaluons l'intérêt respectif.

Dans une deuxième partie, nous dressons un bilan des travaux opérationnels. Nous montrons en particulier que de nombreuses difficultés apparaissent (échantillonnage, redressement, accès aux données, robustesse, fiabilité...) et que la réalisation d'un score opérationnel dépend non seulement du choix d'un bon algorithme, mais encore d'un savoir-faire acquis grâce à l'expérience. Des perfectionnements statistiques possibles sont également évoqués.

1. Méthodologie statistique

1.1 Position du problème

Schématiquement un problème de credit-scoring peut se décrire comme la recherche d'une procédure de séparation entre deux groupes: les "mauvais payeurs" et les "bons payeurs", connaissant un ensemble de descripteurs. Les variables explicatives sont presque toujours qualitatives (catégorie professionnelle, état matrimonial, "géotype"...) ou

rendues qualitatives par découpage en classes (ancienneté d'emploi par exemple) ce qui exclut un usage classique des procédures d'analyse discriminante.

Les données sont donc constituées de deux échantillons de taille généralement élevée (plusieurs centaines d'observations) décrits par p variables qualitatives X_1 à m_i modalités $i = 1, 2, \dots, p$. p est en général de l'ordre de plusieurs dizaines.

1.2 Techniques statistiques utilisables

L'objectif étant de prévoir la qualité d'un dossier à partir des variables explicatives, deux grandes catégories de procédures sont envisageables: celles fondées sur l'estimation de la probabilité d'appartenance à l'une des classes et celles fondées sur une règle numérique de classement: "le score".

1.2.1 Estimation de la probabilité

Le schéma multinomial est ici inutilisable car

il y a $2 \times \prod_{i=1}^p m_i$ cas possibles et ce nombre est en général

excessif. Il faut donc réduire le niveau des interactions utiles et prendre par exemple des modèles du type log-linéaire (voir Goldstein-Dillon, Daudin par exemple).

On peut également utiliser une méthode de segmentation.

Ces techniques sont cependant très lourdes à mettre en oeuvre et mal adaptées à la taille des fichiers que nous traitons.

1.2.2 Fonction de score

Cette approche est la plus usuelle: on attribue à chaque catégorie de chaque variable une valeur numérique. Le score est la somme des valeurs des catégories auxquelles appartient un individu et est la variable numérique de décision: on peut soit se fixer un seuil d'acceptation, soit décider en fonction des courbes donnant les pourcentages de la demande et le pourcentage de mauvais dossiers en fonction du score. La règle est extrêmement simple d'application et facile d'interprétation.

Le score n'est en réalité que la variable discriminante au sens de Fisher, combinaison linéaire des $m_1 + m_2 + \dots + m_p$ variables indicatrices des catégories des p variables.

La méthode du score est donc une analyse discriminante sur variables qualitatives, et une façon de quantifier les variables qualitatives.

L'ensemble des variables indicatrices n'est pas de plein rang car la somme des indicatrices d'une même variable qualitative vaut 1. Des solutions techniques élémentaires comme la suppression de la dernière catégorie de chaque variable, permettraient d'utiliser un programme ordinaire d'analyse discriminante. Telle n'est cependant pas la technique que nous

avons proposé (Saporta 1976) et qui est utilisée sous le nom de Disqual.

1.2.3 Disqual

Dans cette méthode nous procédons tout d'abord à une analyse factorielle (correspondances multiples) de l'ensemble des p variables qualitatives explicatives. L'ensemble de tous les facteurs engendre le même espace que l'ensemble des indicatrices.

Nous éliminons les facteurs de trop faible inertie et ceux ne séparant pas suffisamment les deux groupes. (Un des avantages de cette procédure est que les facteurs sont orthogonaux 2 à 2, ce qui permet d'utiliser une métrique diagonale. Les résultats de l'ACM sont par ailleurs très robustes ce qui est une sécurité.)

Les k coordonnées factorielles retenues sont alors prises comme variables explicatives numériques pour une analyse discriminante. On calcule alors la fonction de Fisher, que l'on exprime in fine comme combinaison linéaire des indicatrices, ce qui donne la fonction de score.

Une variante de cette méthode, mais sans la sélection des facteurs, figure dans le logiciel SPAD (étape DIS2G).

On a en effet, si d est la fonction de Fisher = $\begin{bmatrix} d_1 \\ \vdots \\ d_k \end{bmatrix}$

$$d_j = \frac{(\bar{z}_1^j - \bar{z}_2^j)}{\sqrt{\lambda_j}}$$

où \bar{z}_i^j est la moyenne des coordonnées du groupe i sur l'axe factoriel $n^0 j$ de variance λ_j .

Le score s d'un individu s'obtient par:

$$s = \sum_{j=1}^k d_j z^j$$

Comme $z^j = \sum_{l=1}^n \alpha_l^j x_l$ on a $s = \sum_{j=1}^k \sum_{l=1}^n d_j \alpha_l^j x_l$

$$s = \sum_l \left(\sum_j d_j \alpha_l^j \right) x_l$$

où x_l est l'indicatrice de la catégorie l .

La valeur ou score de cette catégorie vaut donc $\sum_{j=1}^k d_j \alpha_l^j$.

Par construction de la fonction de score, celle-ci correspond à un modèle purement additif sans interaction. La prise en compte des interactions entre variables explicatives améliore très nettement les performances du score. Ceci peut s'effectuer de

la façon suivante : on détecte au moyen d'un modèle log-linéaire ou d'une segmentation que X_1 et X_2 interagissent: on remplace X_1 et X_2 par la variable $X_1 \times X_2$ à $m_1 \times m_2$ catégories et on applique la méthode Disqual. Souvent $m_1 \times m_2$ est trop grand et des regroupements de catégories sont possibles.

2. Les applications du crédit scoring : risques d'échec et conditions de réussite

Nous dressons ici un bilan des travaux opérationnels. Lors de la mise en oeuvre, de nombreuses difficultés apparaissent, liées soit à des problèmes méthodologiques, soit à des problèmes pratiques. Pour qu'un score soit opérationnel et utilisable, il faut non seulement choisir le bon algorithme, mais encore acquérir un grand "savoir faire".

2.1 Quelques problèmes méthodologiques

2.1.1 La population, l'échantillonnage et les redressements

Le premier problème est peut-être le plus délicat à résoudre. C'est celui qui entraîne le plus grand nombre d'échecs. Pour être opérationnel, un score doit s'appliquer à l'ensemble de la demande. Celle-ci peut être assimilée à une population, dénombrable (P). Pour construire un score, l'idéal serait d'observer le comportement payeur d'un échantillon aléatoire représentatif de P . Or, dans la pratique, une préselection est effectuée et certains dossiers sont refusés. Par conséquent, l'échantillon observé (E_1) est tiré dans $P_1 \subset P$, P_1 constituant la population des dossiers acceptables. Si l'on n'y prend garde, le score s'appliquera donc à P_1 et non à P .

Dans la mise en oeuvre, cela signifie que les dossiers devront être tout d'abord sélectionnés à l'aide des méthodes traditionnelles, et dans un deuxième temps, à l'aide du score. C'est bien entendu inacceptable en pratique.

Pour remédier à cet inconvénient, plusieurs approches sont possibles. On peut les classer en deux catégories faisant appel toutes deux à un échantillon (E_2) de dossiers refusés tirés dans $P_2 = P - P_1$. Dans le premier cas, on cherche à redresser E_1 pour lui donner la structure de P sur les principales variables de contrôle. Dans le deuxième cas, on simule le comportement payeur (inconnu) des observations appartenant à E_2 .

Un deuxième problème d'échantillonnage se pose en prenant en compte la dimension temporelle de la demande. On observe fréquemment des phénomènes saisonniers : par exemple Juin est plus mauvais que Octobre. De plus, l'observation du comportement payeur d'un dossier accepté doit être faite sur

une période fixe (e.g. 24 mois), en tenant compte des lois d'apparition des impayés. Les dossiers techniquement utilisables ont alors tous au moins 24 mois d'ancienneté. Entre-temps, la structure de la demande, la réglementation et les conditions d'acceptation ont pu évoluer et le score risque donc d'être obsolète avant d'avoir été utilisé. Un plan de sondage bien construit permet, d'une part, de créer un échantillon complet de dossiers arrivés à maturité et, d'autre part, un échantillon représentatif de la demande récente, sur laquelle on cherchera à s'ajuster.

En conclusion, deux types de redressements s'imposent. Le premier permet de corriger la structure du sous-échantillon observé pour lui redonner la structure de la demande, le deuxième permet de corriger les effets de vieillissement de l'échantillon pour lui redonner la structure de la demande récente.

2.1.2 Le choix des variables et des catégories

D'autres problèmes se posent lorsqu'il s'agit de choisir les variables à introduire dans le modèle et leur découpage en catégories.

Un algorithme efficace de sélection reste encore à mettre au point qui ne nécessiterait pas de réaliser l'analyse discriminante à chaque pas. La méthode que nous avons proposée à l'aide des coefficients de Tschuprow partiels n'a pas apporté les résultats souhaités. Une difficulté provient du fait que les variables indicatrices d'une variable qualitative ne peuvent être dissociées et qu'il faut donc procéder par blocs de variables.

Cependant, les méthodes automatiques de sélection du type "stepwise" conduisent souvent à des solutions inacceptables par l'utilisateur. Une variable "classique", le taux d'endettement par exemple, peut être rejetée car son pouvoir explicatif est légèrement moins bon qu'une variable "exotique" et peu familière à l'utilisateur. De la même manière, une méthode de découpage optimal en catégories ou de regroupement optimal peut conduire à la construction de classes qui surprennent l'utilisateur. De plus certains choix "optimaux" sur un échantillon peuvent être trompeur. A ce niveau, le véritable savoir-faire ne réside pas dans le choix des techniques d'optimisation, mais essentiellement dans la connaissance du domaine et l'accumulation d'expériences. Les méthodes d'optimisation fournissent une première solution "standardisée", il est ensuite nécessaire de rechercher les solutions opérationnelles.

Il en est de même dans le choix de variables croisées. L'expérience nous a montré que, bien souvent, il était préférable de remplacer deux variables qualitatives par une seule variable obtenue par produit cartésien. Dans ce cas, il est d'ailleurs nécessaire de regrouper entre-elles certaines catégories de la variable croisée. La recherche des meilleurs croisements à réaliser peut être effectuée à l'aide de modèles tels que l'analyse de la variance qualitative ou le modèle log-linéaire. Là encore, ce type de modèle peut conduire vers des solutions artificielles et, bien souvent, l'expérience guidera l'analyse dans ses choix a priori de croisements à tester.

2.1.3 Stabilité et interprétation des coefficients

Les techniques de scoring s'apparentent à celles des modèles linéaires généralisés : il est tentant d'augmenter artificiellement le nombre de variables et le nombre de catégories introduites dans le modèle. Cela permet d'améliorer apparemment la qualité de la discrimination. Le problème de la stabilité des estimations des scores se pose alors de façon cruciale. De plus, on ne dispose pas, dans ce cas, de l'arsenal des tests utilisés dans le cas du modèle linéaire.

Les méthodes de rééchantillonnage du type bootstrap devraient permettre de déterminer des intervalles de confiance pour les scores des catégories et d'éprouver des hypothèses les concernant, en l'absence de toute hypothèse distributionnelle. En pratique, il est souhaitable de pouvoir "interpréter" le poids des différents scores, en liaison avec le taux d'impayé dans la catégorie. Une inversion, même justifiée statistiquement, est toujours difficile à expliquer à un utilisateur. La recherche du modèle le plus "économique" en variables et des regroupements judicieux de catégories permet généralement d'éliminer de telles inversions.

2.1.4 Estimation des probabilités à postériori

Si les distributions conditionnelles des coordonnées factorielles sont normales on peut utiliser les résultats usuels de l'analyse discriminante. Cette hypothèse n'est cependant pas toujours vérifiée et une estimation précise des probabilités peut s'effectuer au moyen d'une méthode non paramétrique du type noyau de Parzen [Gautier Saporta 1984]. L'intégration dans un logiciel de type Disquel reste à faire.

2.1.5 Ecart entre prévisions et réalisations

Le modèle étant élaboré, son utilisation est double. Bien entendu, les politiques de sélection fondées sur la grille de score sont mises en place. D'autre part, il est utilisé pour réaliser des prévisions d'impayés, en nombre et en volume, compte-tenu des politiques de sélection choisies. Très rapidement, des écarts entre prévisions et réalisations peuvent apparaître. L'analyse de ces écarts est un point crucial et permet d'orienter les actions. Ils peuvent provenir d'une évolution de l'environnement : contexte réglementaire ou politique commerciale de l'organisme. Ils peuvent également provenir d'une évolution de la demande, ce qui se détecte facilement en suivant le profil de la clientèle sur les variables du score et en le comparant à celui de l'échantillon. Ils peuvent enfin provenir de biais non contrôlés lors de la construction de l'échantillon.

La détection de tous ces écarts nécessite bien entendu la mise en place d'outils permanents de contrôle.

Comment réagir ? Bien souvent, un simple ajustement des seuils d'acceptation permet de corriger l'écart. Si cela ne suffit pas, et si les causes de dérive sont structurelles, la seule solution est d'ajuster un nouveau modèle sur un échantillon actualisé.

2.2 Quelques problèmes pratiques

2.2.1 L'accès aux données pertinentes

Il est rare que les informations contenues dans les dossiers acceptés et refusés soient stockées sur support informatique. Souvent, les dossiers refusés ne sont même pas conservés. Les statistiques de demande et d'acceptation permettant d'élaborer le plan de sondage n'existent pas toujours. Il est donc nécessaire de réaliser une phase préalable de comptage, de codification et de saisie. Par la suite, il faut mettre en place des procédures de saisies systématiques, permettant d'alimenter une base de données. Cette base servira à réaliser les travaux de suivi des écarts et à tirer les échantillons des futures études de scoring.

Un autre problème se présente fréquemment : l'historique n'est pas utilisable soit parce qu'il s'agit d'une activité nouvelle ou récente, soit parce que les règles de sélection étaient tellement sévères qu'aucun dossier n'est arrivé en impayé. Il est alors possible de mettre en oeuvre des techniques totalement différentes, fondées sur l'exploitation de simulations de la demande, de l'acceptation et du comportement.

2.2.2 La modification des procédures

D'autres problèmes sont encore plus délicats ; ils sont liés aux modifications des procédures administratives et informatiques. Pour les résoudre, il est nécessaire de faire évoluer les applications informatiques et de mettre en oeuvre des actions de formation et de motivation des utilisateurs. Nous ne comptons pas développer cet aspect des choses pour le moment mais, simplement, attirer l'attention sur ces difficultés dont la sous-évaluation conduit la plupart du temps à un échec.

Conclusion

Malgré toutes les difficultés méthodologiques et pratiques et les risques d'échec qui en résultent, dix années de conseil et de réalisation nous ont permis de mettre en évidence les avantages décisifs du crédit-scoring. Comparé aux procédures traditionnelles, le score possède trois qualités fondamentales:

- la simplicité :

Le recueil du nombre minimum de renseignements permet de prendre une décision rapide.

- l'homogénéité :

la politique d'acceptation est appliquée uniformément dans le réseau.

- la souplesse :

la politique de sélection est assise sur des bases objectives. Toute inflexion peut être décidée rapidement, en mesurant les conséquences des choix nouveaux.

Outre ces avantages, spécifiques de l'outil, citons les principales conséquences d'un score :

- la diminution des impayés :

Une sélection optimisée permet de réduire le volume des impayés.

- la prévision des impayés :

Le modèle permet de simuler avec précision les impayés futurs et donc d'optimiser le calcul des provisions.

- la productivité :

Seuls les dossiers à risque font l'objet d'un examen attentif. Dans la majorité des cas, la décision peut être déléguée et l'analyse accélérée.

Pour obtenir tous ces avantages, les banques et les organismes financiers doivent être conscients que la mise en place d'un score nécessite un ensemble de compétences statistiques et pratiques et que par conséquent, il s'agit d'une décision stratégique dont la mise en oeuvre nécessite temps et investissement.

Références

Bouroche J.M., Saporta G., Tenenhaus M. (1977) , Some methods of qualitative data analysis in Recent Developments in Statistics, J.R. Barra Ed., pp749-755, North Holland, Amsterdam

Daudin J.J. (1980), Régression qualitative : choix de l'espace prédicteur in Data Analysis and Informatics, E. Diday Ed., pp 329-345, North Holland, Amsterdam

Gautier J.M. , Saporta G. (1984), Méthodes non paramétriques en analyse discriminante ; quelques propositions nouvelles in Data Analysis and Informatics III, E. Diday and al. (editors), pp 591-605, North Holland, Amsterdam

Goldstein M., Dillon W.R. (1978), Discrete Discriminant Analysis, J. Wiley, New York

Saporta G. (1976), Discriminant Analysis when the variables are nominal. Spring Meeting of the Psychometric Society. Murray Hill (N.J.)

Saporta G. (1977), Une méthode et un programme d'analyse discriminante pas à pas sur variables qualitatives. 1ères Journées Analyse des Données et Informatique, INRIA, pp201-210, Paris

Sireyjol N. (1987), les apports du crédit-scoring, Banque, n° 475, pp788-794.