ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 28 — 1992

Courbes de performance, de sélection et de discrimination

Christian GOURIEROUX *

RÉSUMÉ. – Dans cet article nous introduisons un ordre sur les fonctions scores habituellement utilisées en crédit et en assurance et reflétant l'idée d'une sélection plus précise de la clientèle. Ces ordres apparaissent naturellement définis par comparaison de courbes dites de performance, de sélection et de discrimination. Divers résultats sont décrits concernant ces ordres et les façons d'utiliser les trois types de courbes en pratique.

Performance, Selection and Discrimination Curves

ABSTRACT. – We introduce an ordering on the scoring functions used in credit and insurance problems; this ordering corresponds to the idea of a more precise selection of the customers. It is naturally based on the comparison of some curves, called performance, selection and discrimination curves. Various results are described concerning this ordering and the pratical use of the associated curves.

* C. GOURIEROUX CREST, 15, Boulevard G. Peri, 92245 Malakoff Cedex, France.

1 Introduction

Dans divers problèmes de finance, de crédit ou d'assurance, on est souvent amené à sélectionner les individus (titres, entreprises, personnes) les moins risqués possibles. C'est par exemple le cas d'une société d'assurance automobile qui, à prime égale, préfère assurer les conducteurs susceptibles d'avoir le moins d'accidents possible, d'une institution de crédit, qui préfére accorder des prêts aux demandeurs ayant les plus fortes chances de rembourser, d'une compagnie aérienne qui veut détecter les clients susceptibles de ne pas se présenter au moment du vol, afin d'effectuer au mieux le surenregistrement. Dans ce type de problèmes, on peut en première approche considérer que la population d'individus peut être partitionnée en deux sous-populations : celle des non défaillants et celles des défaillants. Ces deux sous-populations seront dans la suite repérées au moyen des valeurs 1 ou 0 prises par une caractéristique Y, le fait que cette variable soit nulle indiquant la défaillance de l'individu correspondant.

La sélection *a priori*, c'est-à-dire avant observation de Y, d'individus peu risqués est généralement effectuée à partir d'une fonction score. Celle-ci est une variable scalaire, notée S, fonction de caractéristiques X observables de l'individu. Elle est utilisée pour classer entre eux les divers clients potentiels. En pratique on se donne un seuil s_0 et on retient les individus pour lesquels la valeur du score dépasse ce seuil. La proportion d'individus retenus est alors $P[S \geqslant s_0]$ et la proportion de défaillants parmi ceux-ci est $P[Y=0/S \geqslant s_0]$. Le niveau s_0 pouvant être fixé à diverses valeurs en fonction des coûts, des parts de marché souhaitées..., il est naturel de considérer qu'à un score donné est attachée une infinité de sélections $S \geqslant s_0$, où s_0 varie.

Il existe deux approches duales pour construire des scores, approches qui possèdent des propriétés d'optimalité pour certaines formes de fonctions de coût (voir e. g. Bardos [1989]) : la première consiste à étudier la loi conditionnelle du risque de défaillance, c'est-à-dire de Y, sachant des caractéristiques observables X de l'individu et à retenir comme score une fonction croissante de la probabilité de non défaillance conditionnelle aux caractéristiques $P[Y=1/X]$. De telles études fondées sur des modèles conditionnels logit ou probit conduisent à des scores fonctions linéaires des caractéristiques introduites dans la formulation. La seconde approche consiste à se donner les formes des distributions des caractéristiques X dans la sous-population des défaillants et dans celle des non défaillants, et à retenir comme score une fonction croissante du rapport des deux densités correspondantes (rapport de discrimination). Dans le cas où ces deux lois sont normales, ceci revient à rechercher la combinaison des caractéristiques $S=\alpha' X$ permettant de rendre les deux lois conditionnelles de S les plus « différentes » possible. Cette seconde approche conduit aux méthodes d'analyse discriminante (linéaire).

Dans cet article, nous nous proposons d'introduire des relations d'ordre sur les scores, un score étant préférable à un autre, s'il conduit à une sélection plus précise des individus.

Dans le paragraphe 2, nous introduisons les courbes de performance et de sélection. Le premier type de courbe a une interprétation simple, puisqu'elle donne le taux de défaillance dans la population sélectionnée en fonction de la part de marché. La courbe de sélection, liée bijectivement à la courbe de performance sera elle plus utile d'un point de vue mathématique. Ces deux courbes sont construites en utilisant l'approche par modèle qualitatif conditionnel, où la variable de non défaillance est expliquée en fonction des caractéristiques.

Dans le paragraphe 3, nous établissons que certaines propriétés (croissance, convexité...) des courbes précédentes caractérisent des idées de liaison croissante entre la variable de non défaillance et le score. Il apparaît alors souhaitable que ces propriétés soient satisfaites par les scores établis en pratique, et nous vérifions notamment qu'elles le sont par les scores optimaux.

Les courbes de performance ou de sélection sont utilisées dans le paragraphe 4 pour comparer entre eux deux scores donnés. On introduit ainsi un ordre partiel sur les scores, qui se révèle utile à la fois pour choisir entre deux scores concurrents et pour assurer le suivi dans le temps d'un score donné.

Le paragraphe 5 est consacré au problème de la dualité. Est-il possible de retrouver un ordre équivalent à partir d'une démarche du type analyse discriminante, fondée sur l'analyse des lois caractéristiques X sachant la non défaillance $Y=1$ ou sachant la défaillance $Y=0$. Ceci nous conduit à définir et étudier les courbes de discrimination.

Divers exemples correspondants aux modèles usuels apparaissant en crédit et en assurance (modèles logit, probit, de durée, d'analyse discriminante) sont discutée dans le paragraphe 6.

2 Courbes de performance et de sélection

Un score donné S fournit plusieurs modes de sélection : ceux-ci dépendent en effet du seuil s auquel est comparé le score. Pour construire un résumé de la qualité du score S, il est naturel d'introduire des mesures scalaires de qualité pour chacun des modes de sélection. Nous sommes ainsi conduits à retenir comme résumés de qualité des courbes paramètrées par le niveau s.

DÉFINITION 1 :

(i) la **courbe de performance** du score S est la courbe d'équation paramétrique :

$$x(s)=P[S\geqslant s],$$
$$y(s)=\frac{P[Y=0/S\geqslant s]}{P[Y=0]}.$$

Nous notons $y=\mathscr{P}(x)$ son équation implicite.

(ii) La **courbe de sélection** du score S est la courbe d'équation paramétrique :

$$\begin{cases} x(s)=P[S\geqslant s], \\ y(s)=P[S\geqslant s/Y=0]. \end{cases}$$

Nous notons $y=\mathscr{S}(x)$ son équation implicite.

Comme : $P[S\geqslant s/Y=0]=\dfrac{P[Y=0/S\geqslant s]}{P[Y=0]}P[S\geqslant s]$,

ces deux courbes sont liées par :

$$\mathscr{S}(x)=x\,\mathscr{P}(x).$$

La courbe de performance donne la qualité de la population retenue, mesurée en terme de proportion relative de défaillants, en fonction de la part de marché $P[S\geqslant s]$, associée au mode de sélection. Cette courbe a son graphe dans le carré $[0, 1]\times[0, 1]$, et on s'attend, si le score a été correctement choisi, qu'elle soit croissante (plus bas est le seuil s, moins bonne est la qualité) et passe par les points (0, 0) (une très grande sévérité devrait permettre de se garder contre les défaillances potentielles) et (1, 1) (obtenu en prenant pour s la plus petite valeur possible).

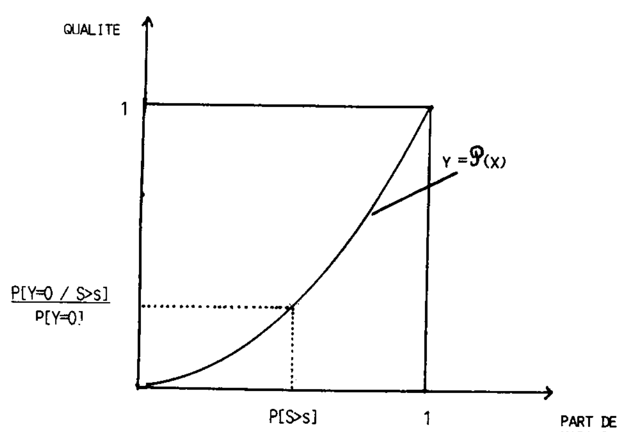

FIGURE 1

Courbe de performance

La courbe de sélection donne la probabilité d'être sélectionné d'un défaillant en fonction de la part de marché. Il s'agit d'une courbe, dont la forme est assez semblable à celle d'une courbe de performance.

Ces courbes sont évidemment très utiles du point de vue contrôle de la sélection. La courbe de performance permet ainsi de connaître la dégradation de clientèle consécutive à un accroissement de part de marché (à score S donné).

On notera que les courbes précédentes sont indépendantes de la représentation retenue pour le score. Ainsi si le score initial S est remplacé par le score $S^* = h(S)$, où h est une fonction strictement croissante, la courbe de performance associée à S^* est d'équation paramétrique :

$$\begin{cases} x^*(s) = P[S^* > s] = P[S > h^{-1}(s)] = x[h^{-1}(s)], \\ Y^*(s) = y[h^{-1}(s)]. \end{cases}$$

Ceci équivaut à un changement bijectif de paramètrage et ne change ni la courbe, ni son équation implicite.

3 Quelques propriétés souhaitables d'un score

Il est souhaitable qu'un score soit en liaison croissante avec la variable Y à prévoir. Il existe diverses définitions de l'idée de liaison croissante. Nous donnons ci-dessous plusieurs d'entre elles et leurs caractérisations en terme de courbes de performance et de discrimination (les démonstrations sont fournies dans l'annexe).

Propriété 1 : (i) La courbe de performance (resp. : de sélection) est en dessous de la droite $y = 1$ (resp. : de la première bissectrice) si et seulement si $\mathrm{Cov}[g(Y), h(S)] \geqslant 0$, pour toutes fonctions g et h croissantes.

(ii) La courbe de performance est croissante si et seulement si : $\mathrm{Cov}[g(Y), h(S)/S \geqslant s] \geqslant 0$, pour toutes fonctions g et h croissantes et pour tout niveau s.

(iii) La courbe de sélection est croissante, convexe si et seulement si la prévision : $E[g(Y)/S = s]$ est fonction croissante de s, ceci pour toute fonction croissante g.

Comme annoncé diverses notions de croissance sont caractérisables à partir des courbes introduites précédemment. On peut vérifier que ces notions sont de plus en plus contraignantes au sens où la condition intervenant en (iii), implique celle intervenant en (ii), qui elle même implique celle intervenant en (i) (voir Suppes [1970]). L'analyse des courbes de performance et de sélection permettra en pratique de savoir quelle notion de liaison croissante est satisfaite. Il est intuitif qu'un « bon score » devrait satisfaire la condition la plus contraignante (iii). Cette intuition est confortée par le résultat ci-dessous.

Propriété 2 : Soient X un ensemble de variables explicatives et $S = P[Y=1/X]$ le score optimal associé.
Alors la courbe de sélection de S est croissante, convexe.

Preuve : Comme :

$$E[g(Y)/S] \geqslant g(1) + [g(1) - g(0)]\, P[Y=1/S],$$

il suffit de vérifier que $P[Y=1/S]$ est une fonction croissante de S.

Or on voit immédiatement que :

$$P[Y=1/S] = P[Y=1/X] = S. \quad \square$$

Observer une courbe de sélection non convexe implique donc que le score considéré n'est pas un score optimal.

Finalement, il est intéressant d'expliciter l'expression de la courbe de sélection; ceci nous conduira à une interprétation de celle-ci en terme de courbe de Lorenz (voir Rothschild-Stiglitz *a*), *b*)) et nous permettra d'utiliser les propriétés classiques de telles courbes.

Désignons par $f(s)$ la densité du score S (en supposant pour simplifier que la loi S est continue), par $G(s)$ sa fonction de survie. Nous avons :

$$x(s) = P[S \geqslant s] = G(s),$$

$$y(s) = P[S \geqslant s/Y=0] = \frac{P[Y=0,\ S \geqslant s]}{P[Y=0]} = \frac{\int_s^{\bar{s}} P[Y=0/S=u]\, f(u)\, du}{\int_{\underline{s}}^{\bar{s}} P[Y=0/S=u]\, f(u)\, du},$$

où $\underline{s}$ (resp $\bar{s}$) désigne les valeurs minimale et maximale que peut atteindre le score S. Si la fonction $P[Y=0/S] = 1 - P[Y=1/S]$ est décroissante, la courbe de sélection conduit à calculer la proportion de la probabilité de défaillance attribuable à la proportion x d'individus de probabilités de défaillance les plus faibles. Ce n'est autre que la courbe de Lorenz associée à la variable $P[Y=0/S]$.

Propriété 3 : Lorsque la condition 4 (iii) est satisfaite, la courbe de sélection coïncide avec la courbe de Lorenz associée à la variable $P[Y=0/S]$.

4 Comparaison de scores

Les courbes de performance et de sélection se révèlent importantes à la fois pour comparer entre eux des scores différents et pour assurer un bon suivi d'un score donné. Remarquons que les courbes introduites dépendent à la fois du score S considéré et de la population Pop sur laquelle on veut évaluer sa qualité.

DÉFINITION 2 : Le score S_1 est plus performant sur la population Pop_1 que le score S_2 ne l'est sur la population Pop_2 : $(S_1, Pop_1) \gtrsim (S_2, Pop_2)$, si et seulement si la courbe de performance (resp de sélection) de (S_1, Pop_1) est en dessous de celle de (S_2, Pop_2).

En pratique l'ordre de préférence ainsi défini peut être utilisé de deux façons extrêmes :

(i) Pour comparer deux scores S_1, S_2 sur une même population Pop. Ceci permet par exemple de voir dans quelle mesure un score venant d'être construit améliore ou non un score plus ancien.

(ii) Pour comparer les performances d'un même score S sur deux populations différentes. Il peut s'agir de sous-populations des clients potentiels d'une période donnée, l'étude permettant alors de détecter les sous-populations vis-à-vis desquelles le score est le plus (resp le moins) performant. Il peut aussi s'agir de la population globale considérée à deux dates différentes, cette approche permettant l'étude des déformations de la population et des propriétés du score dans le temps.

L'ordre de performance étant fondé sur la comparaison de courbes est un ordre partiel. Deux scores appliqués à une même population peuvent se révéler non comparables, le premier donnant de meilleurs résultats sur certaines sous populations, le second sur d'autres.

L'ordre de performance est compatible avec certaines propriétés naturelles des scores optimaux.

PROPRIÉTÉ 4 : Considérons deux scores optimaux S_1 et S_2 associés respectivement aux informations (X, Z) et X :

$$S_1 = P[Y=1/X, Z], \qquad S_2 = P[Y=1/X].$$

Le score prenant en compte le plus d'information est le plus performant (à population donnée).

Preuve : Nous avons en effet l'égalité :

$$P[Y=O/X] = E[P(Y=O/X, Z)/X].$$

Cette interprétation en terme d'espérance conditionnelle implique que la courbe de Lorenz de $P[Y=O/X]$ est au-dessus de la courbe de Lorenz de $P[Y=O/X, Z]$. □

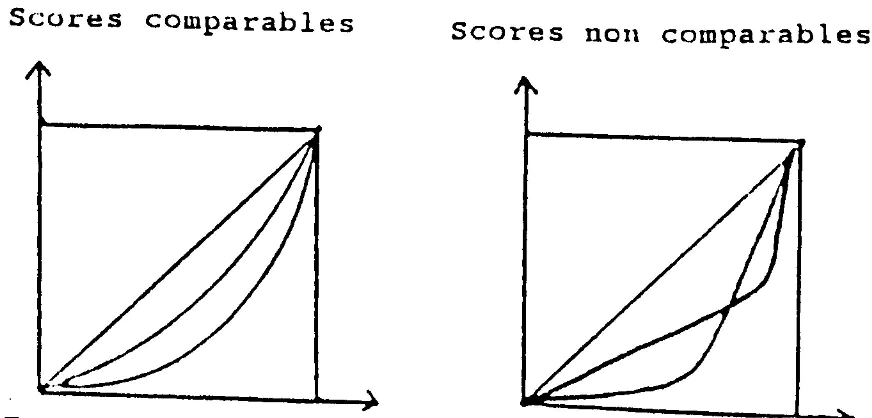

FIGURE 2

Comparaison des courbes de performance

Nous plaçant dans les deux situations extrêmes d'information :

absence d'information : $S_2 = P[Y=1]$,

information complète : $S_1 = P[Y=1/Y] = Y$,

nous en déduisons des encadrements des courbes de performance et de sélection associées aux scores optimaux. Les distributions des scores étant ici discrètes, les courbes limites sont obtenues en prenant le prolongement usuel par randomisation. Ainsi les courbes de sélection associées aux scores canoniques sont comprises entre la première bissectrice et la courbe d'équation.

$$y=0, \quad \text{si} \quad x \leqslant P[Y=1],$$
$$y=1+\frac{x-1}{P[Y=0]}, \quad \text{sinon.}$$

L'encadrement des courbes de performance s'en déduit en divisant les deux courbes limites par x.

Vu l'intérêt d'une courbe de sélection croissante, convexe on peut se demander s'il est facile de corriger un score défectueux, c'est-à-dire dont la courbe de sélection ne satisfait pas ces propriétés. Une réponse partielle est fournie par la propriété ci-dessous.

PROPRIÉTÉ 5 : Soit $\tilde{S}$ un score. Le score optimal fondé sur $\tilde{S}$: $S = P[Y=1/\tilde{S}]$ est plus performant que $\tilde{S}$.

Preuve : Elle résulte immédiatement de la définition de la courbe de sélection. Celle correspondant à $\tilde{S}$ s'obtient en cumulant des valeurs de $P[Y=O/\tilde{S}]$, qui ne sont pas forcément les plus petites; elle est donc toujours au-dessus de celle fondée sur S. □

L'amélioration précédente ne sera effective que si $\tilde{S}$ et S ne sont pas en liaison croissante. Il sera utile pour l'étude d'un score défectueux de tracer la courbe $P[Y=1/\tilde{S}]$ en fonction de $\tilde{S}$ pour voir à quels niveaux de score la croissance n'est pas satisfaite.

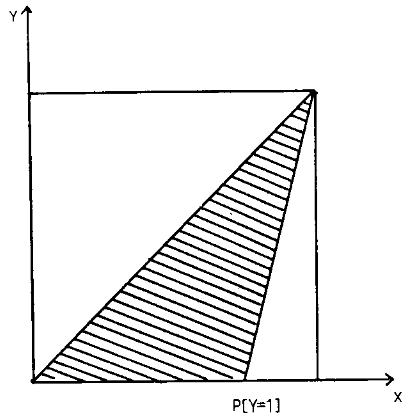

FIGURE 3

Encadrement des courbes de sélection

5 Courbes de discrimination

Les courbes de performance et de sélection, et l'ordre qui s'en déduit, sont directement définis à partir de la loi conditionnelle de la variable Y sachant les scores. Nous allons maintenant considérer le problème sous sa forme duale définie à partir des lois conditionnelles des scores sachant la valeur de Y, et nous allons chercher un ordre reflétant l'idée qu'un score est plus discriminant qu'un autre score. Dans la suite nous notons f_0 et G_0 (resp. f_1 et G_1) la densité et la fonction de survie du score S sachant que $Y=0$ (resp. $Y=1$).

5.1. Definitions

DÉFINITION 3 : La **courbe de discrimination** du score S est l'application $\mathscr{D}=G_1 \circ G_0^{-1}$ de [0, 1] dans [0, 1].

Cette application est croissante, passe par les points [0, 0] et [1, 1]. Elle est clairement invariante par transformation croissante inversible du score S. En effet, si h est une telle transformation, si $S^*=h(S)$, nous avons :

$$G_1^*(s)=P(S^* \geqslant s/Y=1)=P[S>h^{-1}(s)/Y=1]=G_1 \circ h^{-1}(s)$$

et

$$G_0^*(s)=G_0 \circ h^{-1}(s).$$

Nous en déduisons que :

$$G_1^* \circ G_0^{*-1} = G_1 \circ G_0^{-1}.$$

Pour de bons scores. la courbe de discrimination devrait de plus être concave.

> PROPRIÉTÉ 6 : La courbe de discrimination est concave si et seulement si : $E[g(Y)/S=s]$ est fonction croissante de s, pour toute fonction croissante g.

Preuve : Nous avons :

$$\frac{d\mathscr{D}(x)}{dx} = \frac{f_1[G_0^{-1}(x)]}{f_0[G_0^{-1}(x)]}.$$

La fonction $\mathscr{D}$ est donc concave si et seulement si le rapport des densités $\frac{f_1(s)}{f_0(s)}$ est fonction croissante de s. Or ce rapport s'écrit :

$$\frac{f_1(s)}{f_0(s)} = \frac{P[Y=0]}{P[Y=1]} \frac{P[Y=1/S=s]}{1-P[Y=1/S=s]};$$

La croissance de $\frac{f_1(s)}{f_0(s)}$ équivaut donc à la croissance de $P[Y=1/S=s]$. □

Finalement il nous reste à discuter l'intérêt intuitif de la courbe de discrimination pour évaluer le « pouvoir discriminant » d'un score. Pour cela examinons deux cas extrêmes. Le premier correspond à un score absolument non discriminant, c'est-à-dire à $G_1 = G_0$; dans ce cas la courbe de discrimination est confondue avec la première bissectrice.

Considérons maintenant l'autre cas extrême, où les distributions G_0 et G_1 seraient très concentrées autour de points a_0 et a_1, $a_1 > a_0$. Pour toute valeur x non proche de zéro ou de un, on a : $G_0^{-1}(x) \simeq a_0$ et donc $G_1 \circ G_0^{-1}(x) \simeq G_1(a_0) = 1$. La courbe de discrimination serait alors proche de la fonction $H(x)=0$, si $x=0$, $H(x)=1$, si $x>0$. Ceci nous conduit à introduire l'ordre suivant.

> DÉFINITION 4 : Le score S_1 est sur une population Pop_1 plus discriminant que le score S_2 ne l'est sur la population Pop_2, si la courbe de discrimination associée à (S_1, Pop_1) est au-dessus de celle associée à (S_2, Pop_2).

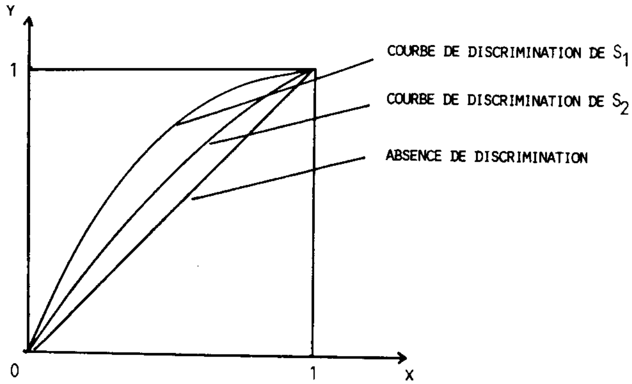

FIGURE 4

Comparaison des courbes de discrimation

5.2. Lien avec les courbes de sélection

Courbes de discrimination et de sélection sont en liaison biunivoque.

PROPRIÉTÉ 7: Si la fonction $\mathcal{S}$ est inversible, on a :

$$\mathcal{S}(x)=(p_1 \mathcal{D}+p_0 \operatorname{Id})^{-1}(x),$$

où $p_1=\mathrm{P}[\mathrm{Y}=1]$, $p_0=\mathrm{P}[\mathrm{Y}=0]$,
et où $(\quad)^{-1}$ désigne l'application inverse.

Preuve : La courbe de sélection admet l'équation paramétrique :

$$\begin{cases} x(s)=\mathrm{P}[\mathrm{S}\geqslant s]=p_1 \mathrm{G}_1(s)+p_0 G_0(s), \\ y(s)=\mathrm{P}[\mathrm{S}\geqslant s/\mathrm{Y}=0]=\mathrm{G}_0(s). \end{cases}$$

Effectuant le changement de paramètre $u=\mathrm{G}_0(s)$, nous avons :

$$\begin{cases} x(u)=(p_1 \mathcal{D}+p_0 \operatorname{Id})(u), \\ y(u)=u, \end{cases}$$

d'où nous déduisons :

$$\mathcal{S}(x)=(p_1 \mathcal{D}+p_0 Id)^{-1}(x). \quad \square$$

La condition d'inversibilité de $\mathcal{S}$ est en fait assez faible, la fonction $\mathcal{S}$ étant croissante et continue. Il suffit donc qu'elle soit strictement croissante.

On déduit immédiatement de la propriété précédente le corollaire suivant :

COROLLAIRE 8 : (S_1, Pop_1) est plus performant que (S_2, Pop_2), si et seulement si il est plus discriminant.

On retrouve bien ainsi l'équivalence des approches par modèle qualitatif dichotomique et par analyse discriminante.

6 Exemples

Dans ce paragraphe, nous donnons quelques exemples de détermination analytique de courbes de performance et de courbes de discrimination. Ces exemples sont déduits d'une formulation de la loi jointe de la variable Y et du score S, ceci à titre de simplification. En pratique ils devraient évidemment être déduits d'une formulation de la loi jointe de Y et des variables X servant à construire le score.

Selon que cette loi jointe du couple Y, S est définie à partir de la loi conditionnelle de S sachant Y ou de la loi conditionnelle de Y sachant S, il sera plus facile de déterminer la courbe de discrimination ou la courbe de performance. Pour les divers modèles, nous mettrons en évidence les paramètres servant à mesurer la plus ou moins grande performance du score.

6.1. Score exponentiel

Ce premier exemple correspond au cas, où le score admet une interprétation en terme de durée. Il serait par exemple assez adapté pour l'assurance automobile, avec un score lié à la durée séparant deux accidents consécutifs. Le modèle est défini à partir des deux lois conditionnelles :

$$\begin{cases} G_1(s) = \exp(-\lambda_1 s), \\ G_0(s) = \exp(-\lambda_0 s), \end{cases} \qquad \text{avec} \quad \lambda_1 < \lambda_0.$$

On a alors pour courbe de discrimination :

$$\begin{aligned} \mathscr{D}(x) &= G_1 \circ G_0^{-1}(x) \\ &= \exp\left[-\lambda_1\left(-\frac{1}{\lambda_0}\operatorname{Log} x\right)\right] \\ &= x^{\lambda_1/\lambda_0}. \end{aligned}$$

La courbe est bien concave et Y dépend donc de façon croissante du score. De plus on note que la comparaison des performances de scores exponentiels peut être effectuée au travers du paramètre λ_1/λ_0, qui dans le cas de l'assurance automobile s'interprête comme le rapport des nombres d'accidents des conducteurs jugés bons et jugés mauvais.

Si $\lambda_1/\lambda_0 = 1$, la courbe correspond à la première bissectrice;

Si λ_1/λ_0 tend vers zéro, la courbe tend vers la fonction limite :

$$H(x) = 0, \quad \text{si } x = 0, \qquad H(x) = 1, \quad \text{si } x > 0.$$

La courbe de performance est définie par :

$$x\mathscr{P}(x)=[p_1 \operatorname{Id}(.)^{\lambda_1/\lambda_0}+p_0 \operatorname{Id}(.)]^{-1}(x).$$

6.2. Cas de scores gaussiens

Les comparaisons de scores par l'intermédiaire de courbes peuvent se ramener au calcul d'indices scalaire, si on dispose d'information complémentaire sur la distribution des scores. Nous avons vu que l'analyse discriminante linéaire reposait sur l'hypothèse que la loi des variables explicatives était normale, de même variance, sur chacune des sous-populations $Y=0$ et $Y=1$. Il est donc intéressant de considérer ce cas appliqué au score S.

Nous supposons donc que le score S admet :

- la loi $N[m_0, \sigma^2]$ sur la sous-population $Y=0$,
- la loi $N[m_1, \sigma^2]$ sur la sous-population $Y=1$, avec $m_1>m_0$.

Nous en déduisons que :

$$G_1(s)=P[S\geqslant s/Y=1]=\Phi\left[\frac{m_1-s}{\sigma}\right], \qquad G_0(s)=\Phi\left(\frac{m_0-s}{\sigma}\right).$$

La courbe de discrimination admet l'équation implicite :

$$y=\mathscr{D}(x)=G_1\circ G_0^{-1}(x)=\Phi\left(\frac{m_1-m_0}{\sigma}+\Phi^{-1}(x)\right).$$

Sous l'hypothèse de normalité, ces courbes appartiennent à une famille paramétrée par $\mu=\dfrac{m_1-m_0}{\sigma}$. Elles sont de plus fonctions croissantes de ce paramètre. Ainsi plus le scalaire μ est grand, plus le score est discriminant, et la comparaison des courbes se ramène à la comparaison des scalaires μ.

Cet indice est proportionnel à la racine carré du rapport de la variance intergroupe $[(m_1-m_0)^2 p_0 p_1]$ par la variance intragroupe (σ^2), rapport qui est souvent utilisé pour évaluer le pouvoir discriminant d'un score (voir BARDOS [1989] p. 79 par exemple).

Il faut noter que l'utilisation de ce rapport pour comparer les scores n'apparaît justifiée que sous les hypothèses distributionnelles restrictives faites sur les lois du score dans chaque population. Afin d'illustrer ce point nous pouvons étudier le cas normal avec variances différentes dans chaque sous-population. Nous avons alors :

$$G_1(s)=\Phi\left[\frac{m_1-s}{\sigma_1}\right], \qquad G_0(s)=\Phi\left[\frac{m_0-s}{\sigma_0}\right],$$

$$y=\mathscr{D}(x)=\Phi\left[\frac{m_1-m_0}{\sigma_1}+\frac{\sigma_0}{\sigma_1}\Phi^{-1}(x)\right].$$

Les courbes de discrimination appartiennent maintenant à une famille à deux paramètres :

$$\mu=\frac{m_1-m_0}{\sigma_1}, \qquad \eta=\frac{\sigma_0}{\sigma_1}.$$

La comparaison des courbes nécessite alors le calcul de deux résumés scalaires. Le rapport de la variance intergroupe par la variance intragroupe, égal à :

$$\frac{p_0 p_1 (m_0-m_1)^2}{p_0 \sigma_0^2+p_1 \sigma_1^2}=\frac{p_0 p_1 \mu^2}{p_0+p_1 \eta^2},$$

est fonction des deux indices précédents, mais son utilisation fait perdre une part de l'information.

Dans le cas où les lois ne seraient pas normales, ce rapport peut même se révéler inadéquat, c'est-à-dire non fonction des indices scalaires permettant la comparaison des courbes (comme le montre l'exemple 6.1 des lois exponentielles).

6.3. Modèle Probit

Il s'agit du modèle où $Y=\mathbf{1}_{Z>0}$ avec un couple $\begin{pmatrix} S \\ Z \end{pmatrix}$ suivant une loi normale, de moyenne $\begin{pmatrix} m_S \\ m_Z \end{pmatrix}$, de matrice de variance-covariance :

$$\begin{pmatrix} \sigma_S^2 & \rho\sigma_S\sigma_Z \\ \rho\sigma_S\sigma_Z & \sigma_Z^2 \end{pmatrix}.$$

Comme la variable Y est qualitative, on peut sans perte de généralité fixer $\sigma_Z=1$. Comme les courbes de performance et de discrimination sont invariantes par transformation croissante de S, on peut fixer $m_S=0$ et $\sigma_S=1$.

Sous ces hypothèses, nous avons :

$$x(x)=P[S\geqslant s]=1-\Phi(s),$$

$$\begin{aligned} y(s)&=\frac{P[Y=0/S\geqslant s]}{P[Y=0]}=\frac{P[Z<0,\ S\geqslant s]}{P[Z<0]\,P[S\geqslant s]} \\ &=\frac{P[Z-m_Z<-m_Z,\ S\geqslant s]}{P[Z-m_Z<-m_Z]\,P[S\geqslant s]} \\ &=\frac{P[Z-m_Z>m_Z,\ S\geqslant s]}{P[Z-m_Z>m_Z]\,P[S\geqslant s]} \\ &=\frac{h[m_Z,\ s,\ \rho]}{[1-\Phi(m_Z)][1-\Phi(s)]}, \end{aligned}$$

en notant $h(a, b, c) = \mathrm{P}[\mathrm{U} > a, \mathrm{V} > b]$, avec $\begin{pmatrix} \mathrm{U} \\ \mathrm{V} \end{pmatrix} \sim \left[\begin{pmatrix} \mathrm{O} \\ \mathrm{O} \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right]$.

La courbe de performance a finalement pour équation :

$$\mathscr{P}(x) = \frac{h[m_{\mathrm{Z}}, (1-\Phi)^{-1}(x), \rho]}{[1-\Phi(m_{\mathrm{Z}})]\, x}.$$

Preuve de la propriété 1

• ***Preuve de 1* (i)**

(*a*) Comme la variable Y ne prend que deux valeurs 0 et 1, on peut se restreindre à écrire la condition $\mathrm{Cov}[g(\mathrm{Y}), h(\mathrm{S})] \geqslant 0$, pour les seules fonctions g linéaires croissantes et donc simplement écrire :

$$\mathrm{Cov}[\mathrm{Y}, h(\mathrm{S})] \geqslant 0, \qquad \forall h \text{ croissante.}$$

(*b*) Comme l'ensemble des fonctions croissantes constitue un cône, fermé pour la limite croissante, engendré par les fonctions du type $h_{s_0}(S) = \mathbf{1}_{\mathrm{S} \geqslant s_0}$, la condition est aussi équivalente à :

$$\mathrm{Cov}[\mathrm{Y}, \mathbf{1}_{\mathrm{S} \geqslant s_0}] \geqslant 0, \qquad \forall s_0.$$

(*c*) Développons alors cette condition; nous obtenons :

$$\begin{aligned}&\mathrm{Cov}(\mathrm{Y}, \mathbf{1}_{\mathrm{S} \geqslant s_0})\\&\quad = \mathrm{Cov}(\mathrm{E}(\mathrm{Y}/\mathrm{S}), \mathbf{1}_{\mathrm{S} \geqslant s_0})\\&\quad = \mathrm{Cov}(\mathrm{P}[\mathrm{Y}=1/\mathrm{S}], \mathbf{1}_{\mathrm{S} \geqslant s_0}),\end{aligned}$$

et la condition s'écrit :

$$\mathrm{Cov}(\mathrm{P}[\mathrm{Y}=0/\mathrm{S}], \mathbf{1}_{\mathrm{S} \geqslant s_0}) \leqslant 0, \qquad \forall s_0.$$

Comme :

$$\begin{aligned}&\mathrm{Cov}[\mathrm{P}[\mathrm{Y}=0/\mathrm{S}], \mathbf{1}_{\mathrm{S} \geqslant s_0}]\\&\quad = \mathrm{P}[\mathrm{Y}=0, \mathrm{S} \geqslant s_0] - \mathrm{P}[\mathrm{Y}=0]\,\mathrm{P}[\mathrm{S} \geqslant s_0],\end{aligned}$$

la condition équivaut à :

$$y(s_0) = \frac{\mathrm{P}[\mathrm{Y}=0/\mathrm{S} \geqslant s_0]}{\mathrm{P}[\mathrm{Y}=0]} = \frac{\mathrm{P}[\mathrm{Y}=0, \mathrm{S} \geqslant s_0]}{\mathrm{P}[\mathrm{Y}=0]\,\mathrm{P}[\mathrm{S} \geqslant s_0]} \leqslant 1.$$

• ***Preuve de 1* (ii)**

Le même raisonnement appliqué à la loi conditionnelle à $\mathrm{S} \geqslant s_1$ avec $s_1 < s_0$ conduit à :

$$\frac{\mathrm{P}[\mathrm{Y}=0/\mathrm{S} \geqslant s_0 \textit{ et } S \geqslant s_1]}{\mathrm{P}[\mathrm{Y}=0/\mathrm{S} \geqslant s_1]} \leqslant 1 \quad \Leftrightarrow \quad \frac{y(s_0)}{y(s_1)} = \frac{\mathrm{P}[\mathrm{Y}=0/\mathrm{S} \geqslant s_0]}{\mathrm{P}[\mathrm{Y}=0/\mathrm{S} \geqslant s_1]} \leqslant 1.$$

Elle est donc équivalente à la décroissance de la fonction $y(s)$. Le résultat s'en déduit en remarquant que la fonction $x(s)$ est également décroissante.

• ***Preuve de 1* (iii)**

Notons $f(s)$ la densité du score S et $G(s)$ sa fonction de survie. La courbe de sélection est telle que :

$$x(s) = G(s),$$

$$y(s) = \frac{\int_s^{\bar{s}} P[Y=0/S=u]\, f(u)\, du}{P(Y=0)}.$$

Nous en déduisons que :

$$\mathscr{S}(x) = \frac{1}{P[Y=0]} \int_{G^{-1}(x)}^{\bar{s}} P[Y=0/S=u]\, f(u)\, du.$$

La dérivée première vaut :

$$\frac{d\mathscr{S}(x)}{dx} = \frac{1}{P[Y=0]} P[Y=0/S=G^{-1}(x)],$$

et est toujours positive.

La courbe de sélection est convexe si et seulement si : $P[Y=0/S=G^{-1}(x)]$ est fonction croissante de x, c'est-à-dire si et seulement si $P[Y=0/S=s] = 1 - P[Y=1/S=s]$ est fonction décroissante de s.

• Références bibliographiques

Altman, E. I. et Eisenbeis, R. A. (1978). — « Financial Applications of Discriminant Analysis, a Clarification », *Journal of Financial and Quantitative Analysis*, Mars, pp. 185-205.

Amemiya, T. et Powell, J. L. (1983). — « A Comparison of the Logit Model and Normal Discriminant Analysis when the Independent Variables are Binary », dans Karlin, Amemiya, et Goodman éd. *Studies in Econometrics, Time Series and Multivariate Statistics*, Academic Press, New York.

Bardos, M. (1989). — « P.M.E. du Bâtiment et du Génie Civil et Agricole : Ratios Significatifs et Détection du Risque », *Cahiers Économiques et Monétaires*, 33.

Gouriéroux, C. (1989). — « Économétrie des Variables Qualitatives », *Economica*, Paris, 2e édition.

Lo, A. W. (1986). — « Logit Versus Discriminant Analysis : A Specification Test and Application to Corporate Bankrupcies », *Journal of Econometrics*, *31*, pp. 151-178.

Manski, C. F. et Lerman, S. R. (1977). — « The Estımation of Choice Probabilities from Choice Based Samples », *Econometrica*, 45, pp. 1977-1988.

Romeder, J. M. (1973). — « *Méthodes et Programmes d'Analyse Discriminante* », Dunod.

Rothschild, M. et Stiglitz, J. (1970). — « Increasing Risk : I A Definition », *Journal of Economic Theory.*, 2, pp. 225-243.

Rothschild, M. et Stiglitz, J. (1970). — « Increasing Risk : II Its Economic Consequences », *Journal of Economic Theory*, 3, pp. 66-84.

Suppes, P. (1970). — « *A Probabilistic Theory of Causality* », North-Holland.