

ESSEC

Cours « Management bancaire »

Séance 3

Le risque de crédit

Le scoring

Les méthodes d'évaluation du risque de crédit pour les PME et les ménages

- **Caractéristiques**

- ❖ Comme les montants des crédits et des pertes possibles sont relativement peu importantes, les banques ne peuvent passer beaucoup de temps sur l'analyse du crédit.
- ❖ Les données des petites et moyennes entreprises et des ménages sont standard et permettent un traitement de masse informatisé.
- ❖ Les modèles peuvent être utilisés à plusieurs fins :
 - *Evaluer les risques : calculer la probabilité d'être un bon client*
 - *Décider : accorder ou ne pas accorder le crédit*
 - *Proposer : marketing*

Rappel du problème et exemples (1)

- Formalisation du problème
 - ❖ Discriminer au mieux les différents groupes d'une population dont les individus sont décrits par des variables afin de prendre une décision.

Rappel du problème et exemples (2)

- **Exemple: le risque de crédit pour les clients particuliers d'un réseau bancaire**
 - ❖ Le problème: discriminer les clients en fonction de leur degré de risque (les bons et les mauvais clients)
 - ❖ Information disponible: caractéristiques des clients (âge, situation familiale, salaire, ancienneté dans l'entreprise, etc.)
 - ❖ Objectif : accorder un crédit au client et éventuellement de déterminer le montant, la durée et le taux du crédit.
 - ❖ **Exercice: trouver d'autres exemples de tels problèmes dans la banque.**

Les étapes de la mise en place d'un outil d'aide à la décision (1)

- Construire une base de données historique
 - ❖ Définir la base de données
 - *Définir précisément l'événement étudié : qu'est-ce qu'un mauvais payeur ?*
 - *Définir les caractéristiques à retenir : âge, salaire, ancienneté dans l'entreprise, etc.*
 - ❖ Collecter les informations
 - ❖ Retraiter la base de données
 - *Valeurs manquantes, valeurs aberrantes*
 - ❖ Redresser éventuellement la base de données
 - *Comment traiter les dossiers refusés ?*

Les étapes de la mise en place d'un outil d'aide à la décision (2)

- Construire une règle de décision à partir d'une méthode d'analyse discriminante (détaillée plus loin pour le scoring)
 - ❖ Choisir une méthode et un modèle
 - ❖ Estimer le modèle et définir la règle de décision

Les étapes de la mise en place d'un outil d'aide à la décision (3)

- **Mettre en œuvre**
 - ❖ Mettre en place des outils informatiques
 - ❖ Mettre en place des procédures
 - ❖ Former les utilisateurs
- **Evaluer la règle de décision et suivre le modèle sous-jacent**

Les étapes de la mise en place d'un outil d'aide à la décision (4)

- Réactualiser régulièrement le modèle et la règle de décision
 - ❖ Ajouter de nouvelles variables ou supprimer d'anciennes variables
 - ❖ Estimer à nouveau les paramètres du modèle
 - ❖ Changer la valeur des paramètres de la règle de décision (la valeur du seuil pour une fonction de score par exemple)

Les méthodes d'analyse discriminantes (1)

- L'analyse discriminante regroupent différentes méthodes:
 - ❖ Les méthodes de scoring :
 - *Méthodes statistiques qui permettent d'affecter à chaque individu un score (une note) représentatif de son profil. La comparaison de ce score à un seuil fixé aboutit à une décision : accorder un crédit à un client de la banque par exemple.*
 - ❖ Les méthodes d'arbre
 - *Méthodes descriptives fondées sur un arbre décisionnel mettant en valeur les variables discriminantes importantes et les liens entre ces variables. Le cheminement d'un individu dans l'arbre permet de le classer et d'aboutir à une décision.*

Les méthodes d'analyse discriminantes (2)

- ❖ Les réseaux de neurones
 - *Modélisations sophistiquées de la relation entre les inputs (variables décrivant les individus) et les outputs (classement des individus permettant de prendre une décision) adaptées à des données complexes.*
- **Toutes ces méthodes permettent de classer les individus et de prendre une décision de façon automatique (gain de temps dans l'analyse des dossiers et rapidité de décision).**

Objectifs des méthodes d'analyse discriminante (1)

- Les différentes méthodes d'analyse discriminante partagent les mêmes objectifs :
 - ❖ Un objectif descriptif
 - *Mettre en évidence les variables qui permettent de séparer au mieux les groupes d'individus.*
 - *Déterminer si les différences entre les groupes sont significatives.*
 - *Donner une représentation graphique qui rende compte de cette séparation.*
 - *L'analyse discriminante à but descriptif repose sur la notion de distance.*

Objectifs des méthodes d'analyse discriminante

(2)

- ❖ Un objectif décisionnel
 - *Construire une règle d'affectation des nouveaux individus à l'un des groupes.*
 - *L'analyse discriminante à but décisionnel repose sur le concept de probabilité.*

Etapes de la construction d'un score

- **Construire un modèle**
 - ❖ Choisir un modèle (exemples : linéaire ou logistique)
 - ❖ Séparer la base de données historiques en deux échantillons : un échantillon d'apprentissage (pour estimer le modèle) et un échantillon test (pour tester le modèle)
 - ❖ Sélectionner les variables explicatives du modèle et estimer les paramètres du modèle à partir de l'échantillon d'apprentissage
 - ❖ Interpréter des paramètres estimés (signe des paramètres)
- **Construire une règle de décision fondée sur le modèle**
 - ❖ Déterminer un seuil pour la règle de décision
- **Mesurer la qualité d'ajustement du modèle sur un échantillon**

Le modèle linéaire (1)

- Un modèle vise à expliquer une variable par d'autres variables :

$$Y_i = \sum_{j=1}^p \beta_j X_{ij} + u_i$$

où Y_i représente la variable à expliquer qualitative ($Y_i = 0$ si le client i est un mauvais client et $Y_i = 1$ si le client i est un bon client).

- *Pour un client de l'échantillon d'apprentissage, la valeur de cette variable est connue. Pour un futur client, cette valeur n'est pas connue au moment de la décision mais le modèle en fournira une estimation qui permettra de prendre une décision.*

Le modèle linéaire (2)

où $(X_{ij})_{j=1, p}$ représente les p variables explicatives qui peuvent être qualitatives (propriétaire ou locataire, marié ou non marié) ou quantitatives (âge, nombre d'enfants, salaire, etc.).

- *Pour un client de l'échantillon d'apprentissage ou un futur client, la valeur de ces variables est connue.*

Le modèle linéaire (3)

- Rappel de l'équation

$$Y_i = \sum_{j=1}^p \beta_j X_{ij} + u_i$$

où $(\beta_j)_{j=1,p}$ représentent les p paramètres du modèle.

- *Ces paramètres sont estimés à partir de l'échantillon d'apprentissage.*

où u_i : le terme d'erreur du modèle.

- *Le modèle n'est pas parfait!*

Le modèle linéaire (4)

- Définition du score

$$Z_i = \sum_{j=1}^p \beta_j X_{ij}$$

- ❖ Le score noté Z_i synthétise les caractéristiques du client i au regard de la variable à expliquer.
- ❖ Plus la valeur du score Z_i est élevée, plus le client i a de chance d'être un bon client.

Le modèle linéaire (5)

- Les limites du modèle linéaire
 - ❖ Remarque 1 : connaissant les variables X , l'anticipation de la variable Y est égale à la probabilité que le client i soit un bon client.

$$E(Y_i / X) = 1 \times p_i + 0 \times (1 - p_i) = p_i \text{ où } p_i = P(Y_i = 1 / X)$$

Le modèle linéaire (6)

- ❖ Remarque 2 : une probabilité est comprise entre 0 et 1. Or, dans le modèle linéaire, cette probabilité estimée peut être en dehors du segment $[0, 1]$, ce qui pose problème...

$$E(Y_i / X) = p_i = \sum_{j=1}^p \beta_j X_{ij}$$

- ❖ Question: quel modèle proposer ?

Le modèle logisitique

- Amélioration du modèle linéaire

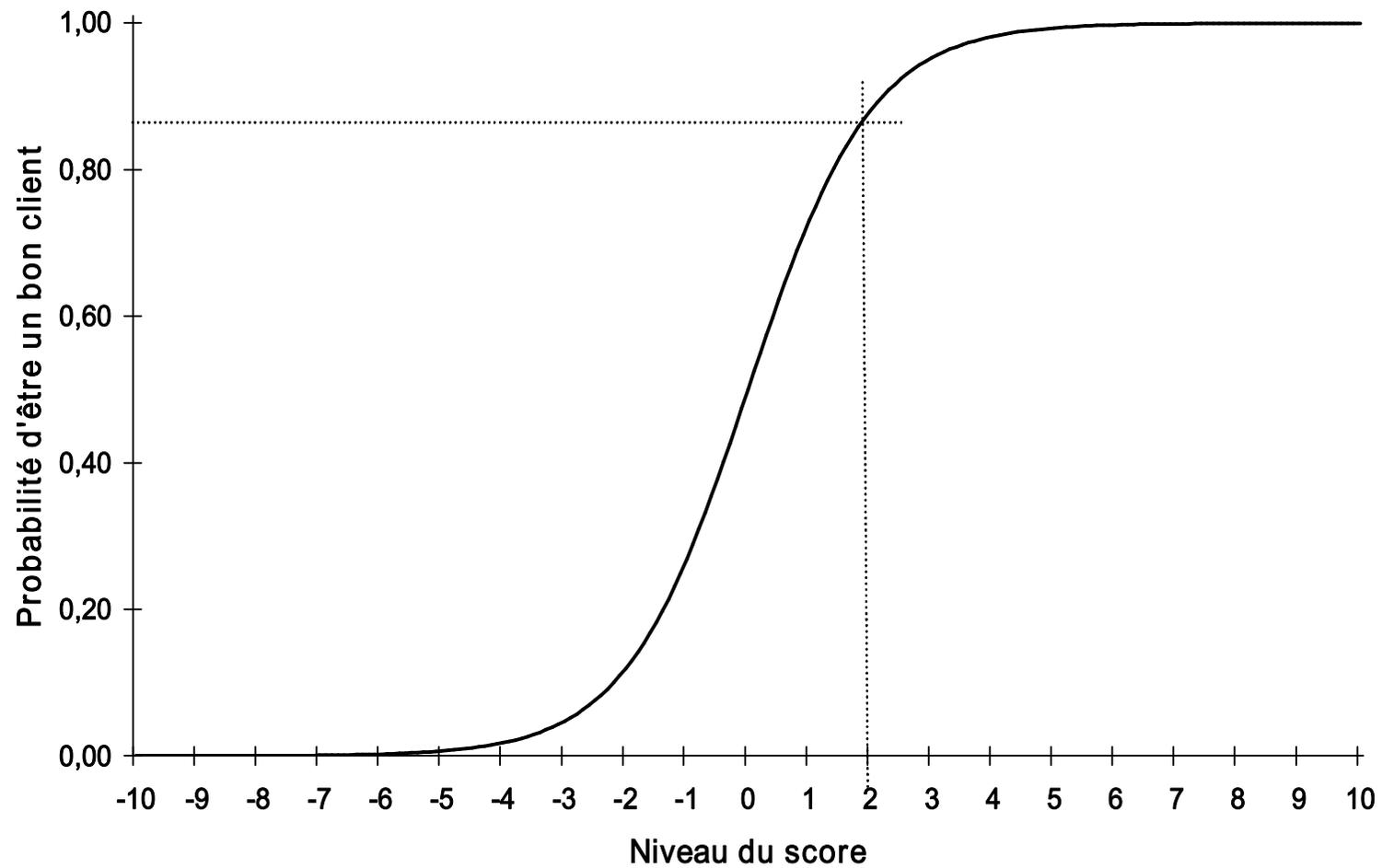
$$E(Y_i / X) = p_i = F\left(\sum_{j=1}^p \beta_j X_{ij}\right)$$

- Le modèle logistique

$$F(x) = \frac{e^x}{1 + e^x}, 0 \leq F(x) \leq 1.$$

- ❖ La relation entre la probabilité d'être un bon client et le niveau du score est représentée ci-dessous :

Le modèle logisitique: probabilité et score



La sélection des variables (1)

- Idée :
 - ❖ Les variables explicatives sont choisies de manière à prédire l'appartenance des individus aux groupes.
 - ❖ Lorsqu'une variable est introduite dans le modèle, elle apporte de l'information pour discriminer les individus appartenant à des groupes différents, mais elle apporte aussi du bruit.
 - ❖ La pertinence de la variable dépend du ratio information/bruit.

La sélection des variables (2)

- **Méthode :**
 - ❖ La contribution marginale de chaque variable au modèle est considérée et testée.
 - ❖ Une variable donnée est retenue d'après un critère objectif : comparaison du maximum de vraisemblance de deux modèles (l'un comprenant la variable l'autre pas).

La sélection des variables

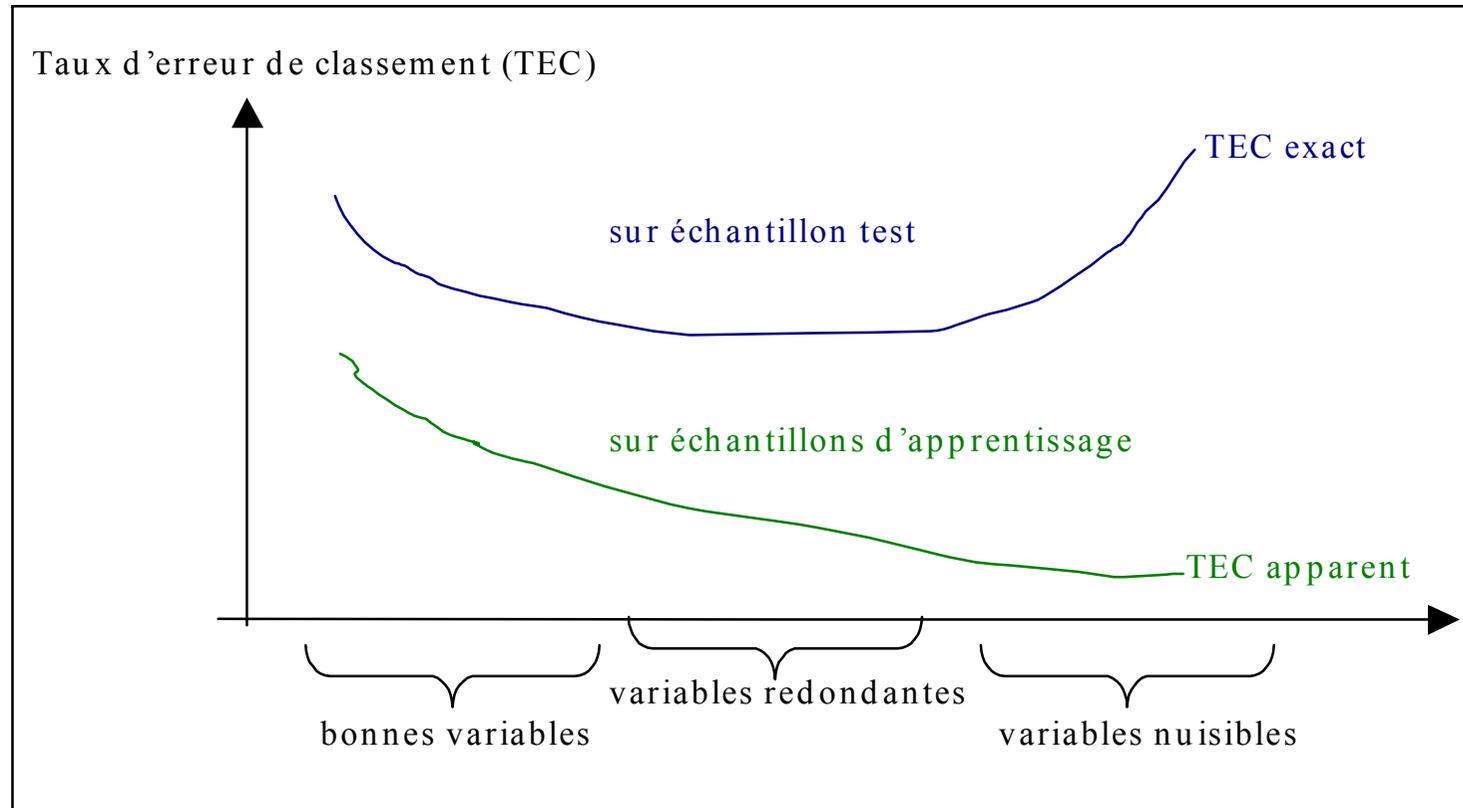


Figure 1 : Evolution du tax d'erreur de classement en fonction du nombre de variables explicatives dans le modèle.

La règle de décision (1)

- La règle de décision repose sur le choix d'un score limite (seuil) noté Z^l .
 - ❖ Si $Z_i < Z^l$, alors la décision est « non ».
 - ❖ Si $Z_i \geq Z^l$, alors la décision est « oui ».
- Choisir un score limite Z^l revient à choisir une probabilité limite notée $p^l : p^l = F(Z^l)$
 - ❖ Si la probabilité estimée du client i d'être « un bon client » est inférieure à la probabilité limite, c'est-à-dire si $p_i < p^l$, alors la décision est « non ».
 - ❖ Si la probabilité estimée du client i d'être « un bon client » est supérieure à la probabilité limite, c'est-à-dire si $p_i \geq p^l$, alors la décision est « oui ».

La règle de décision (2)

- En pratique, pour une valeur du score limite donnée, de bons clients auront un score inférieur au score limite et de mauvais clients auront un score supérieur au score limite.
 - ❖ Encore une fois, le modèle n'est pas parfait !
 - ❖ Si on fixe une valeur basse pour le score limite, on acceptera tous les bons clients mais aussi beaucoup de mauvais.
 - ❖ Si on fixe une valeur trop élevée pour le score limite, on n'acceptera pas de mauvais clients mais très peu de bons clients.
 - ❖ Pour déterminer la valeur optimale du score limite, il faut effectuer au préalable une analyse de rentabilité des bons clients et des mauvais clients. La valeur optimale du score limite est la valeur qui maximise le profit de la banque.

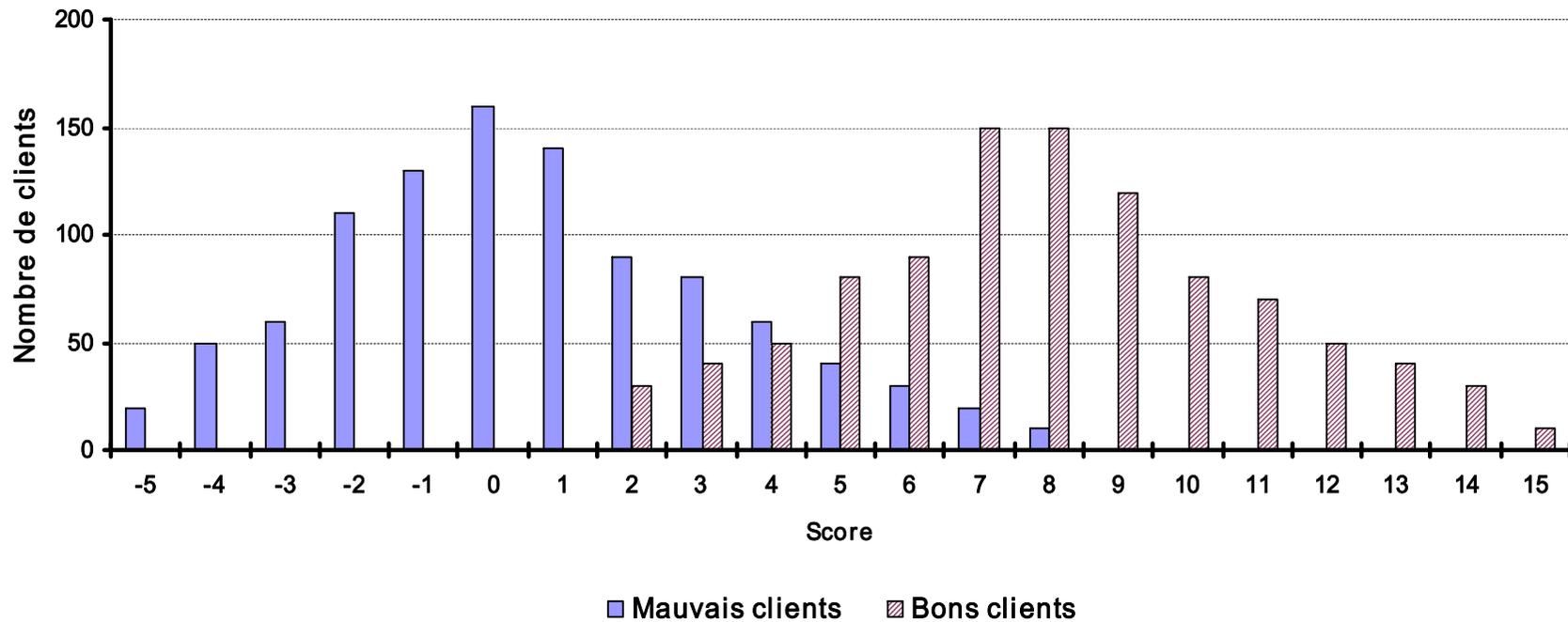
Distribution du score

Distribution du score de tous les clients (bons et mauvais)



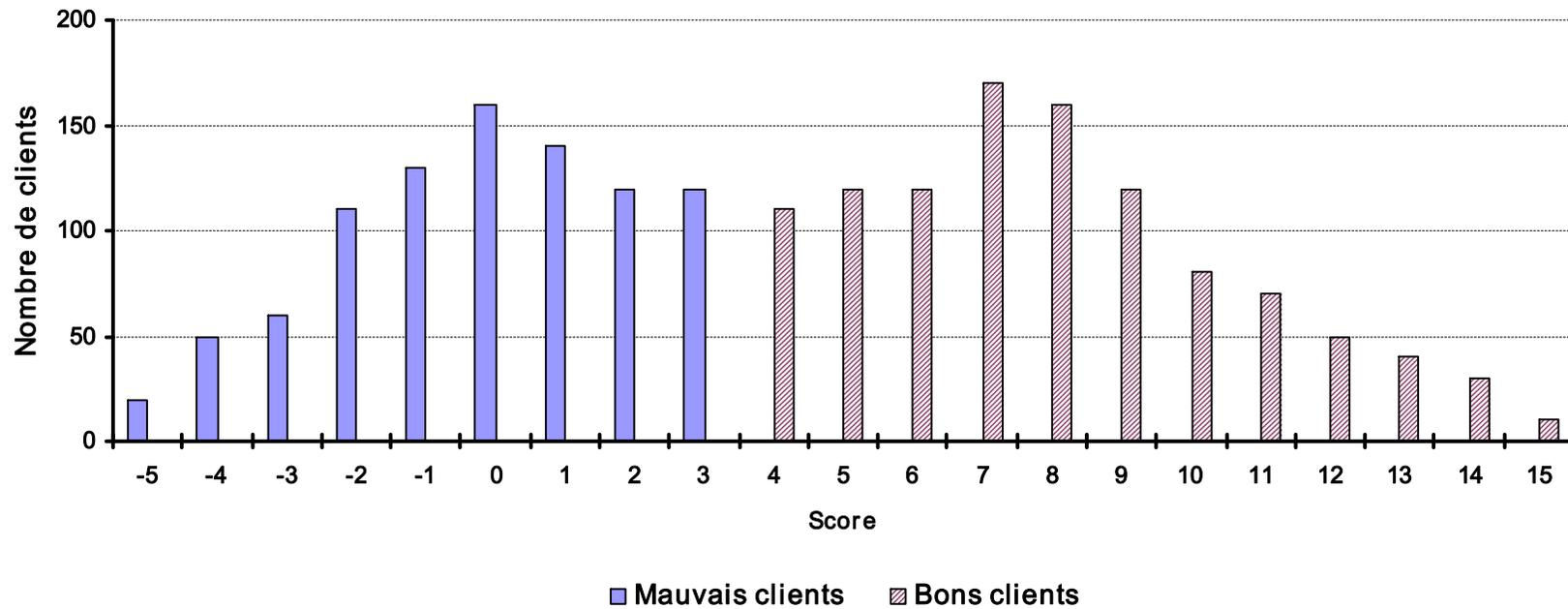
Distribution du score par groupe (1)

Distribution du score des bons clients et des mauvais clients
CAS 1: le score est discriminant



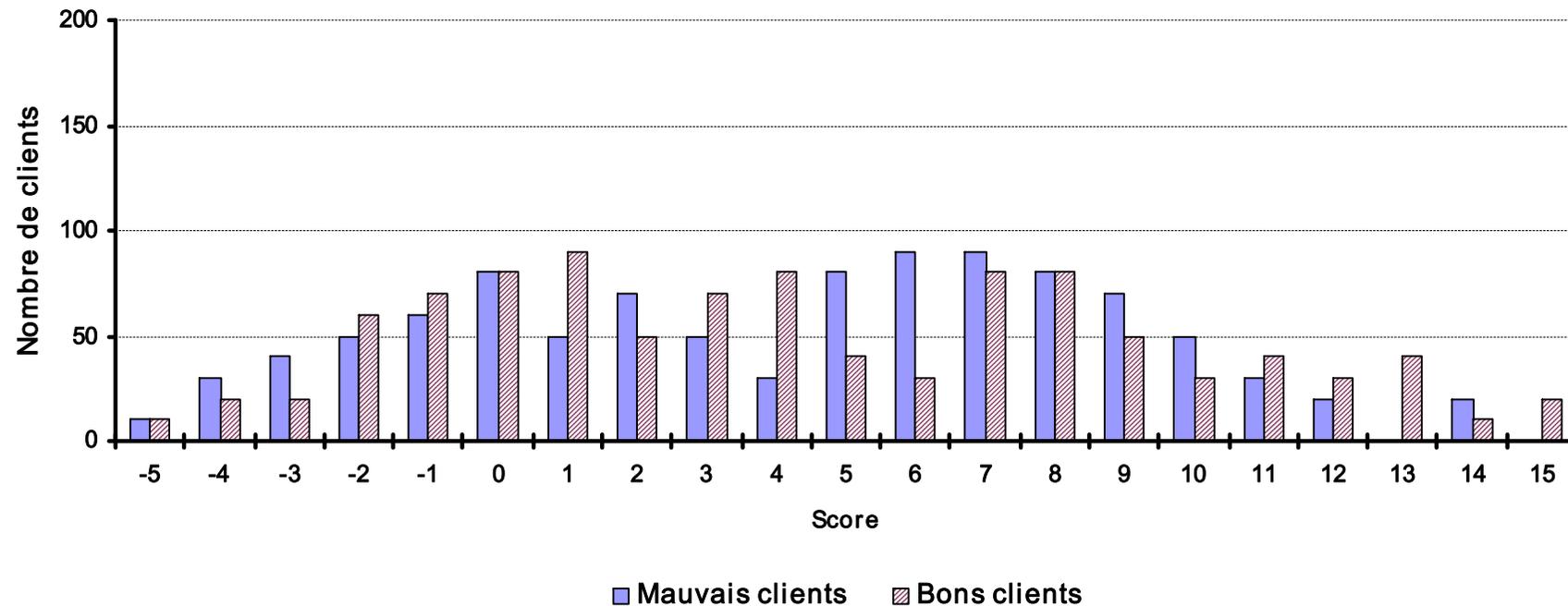
Distribution du score par groupe (2)

Distribution du score des bons clients et des mauvais clients
CAS 2: le score est parfaitement discriminant



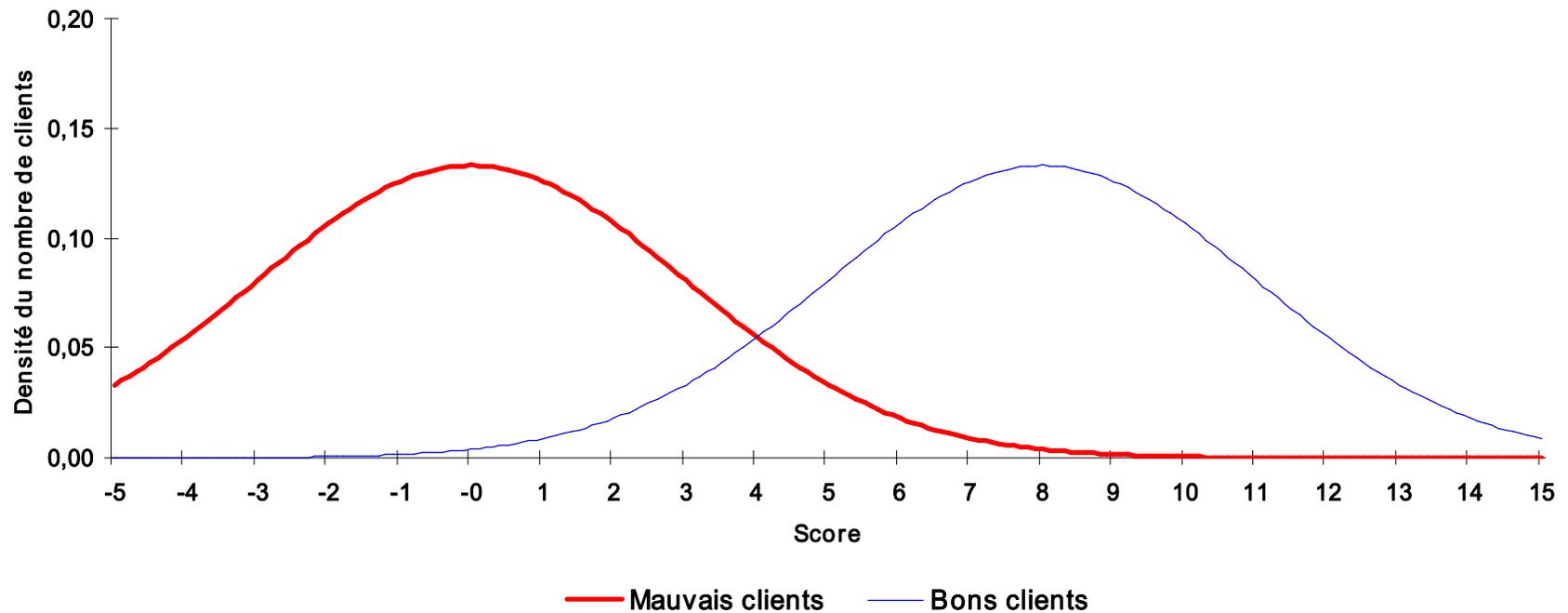
Distribution du score par groupe (3)

Distribution du score des bons clients et des mauvais clients
CAS 3: le score n'est pas discriminant

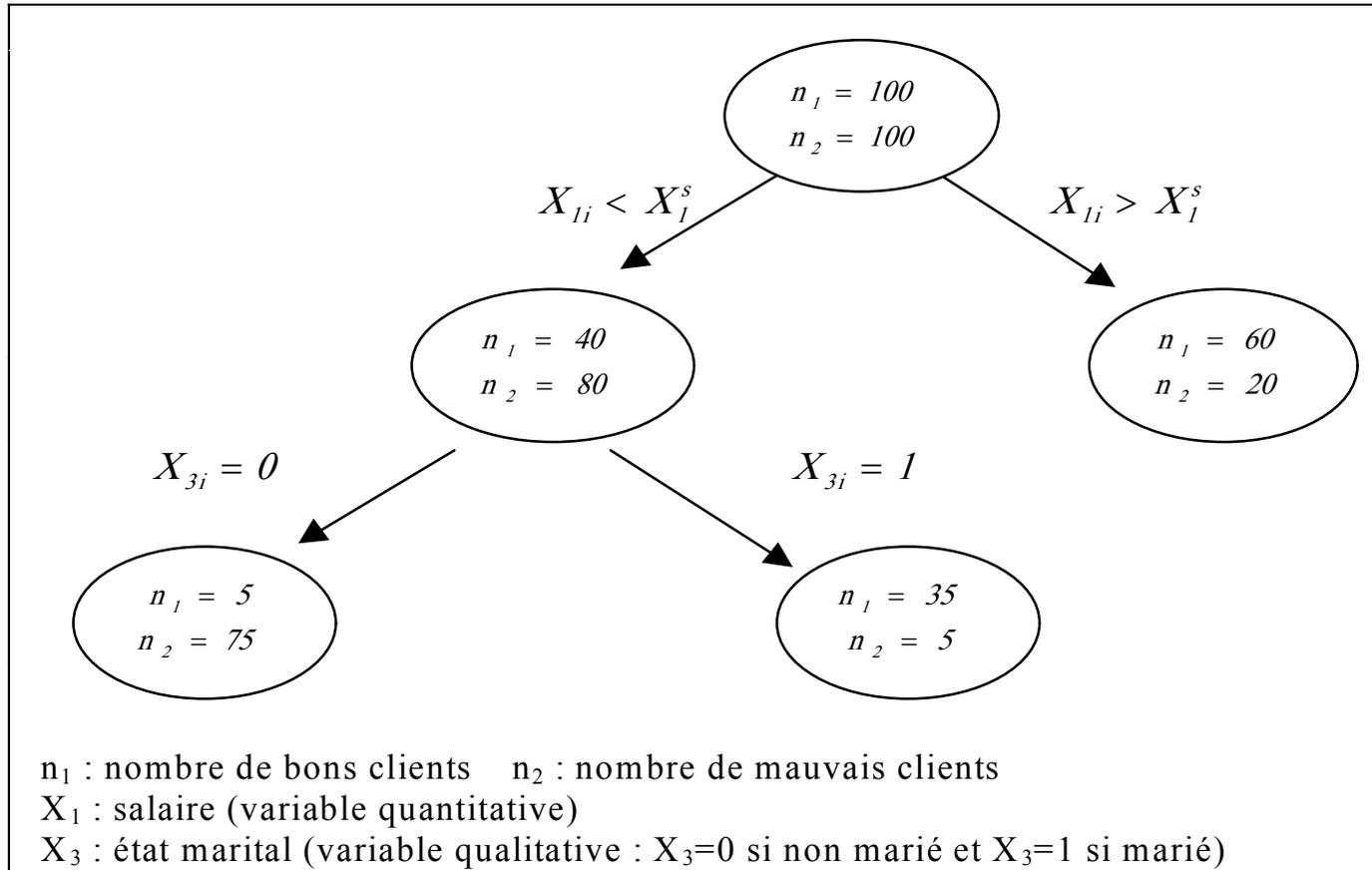


Distribution du score par groupe modélisée

Distribution du score de bons clients et des mauvais clients
Modélisation de l'échantillon



La méthode de l'arbre



Exemple d'arbre de décision binaire à 2 niveaux

Validation de l'analyse discriminante (1)

- Une fois la règle de décision construite (fonction de score, arbre de décision ou réseau de neurones), il est important d'évaluer sa performance. La performance est mesurée par les taux d'erreur de classement.
- Outil d'évaluation ex ante : la matrice de classement ou la matrice de confusion

		Groupes d'affectation	
		Bons clients	Mauvais clients
Groupes d'origine	Bons clients	52	12
	Mauvais clients	9	47

Validation de l'analyse discriminante (2)

- ❖ Dans cet exemple, après application de la méthode, 52 bons clients et 47 mauvais clients ont bien été classés dans leur groupe d'origine.
- ❖ Par contre, 12 bons clients ont été considérés comme mauvais clients et 9 mauvais clients comme bons clients.
- **Outil d'évaluation ex post : comparaison des taux d'erreur de classement anticipé et réalisé**