

# **CALCUL D'UN SCORE ( SCORING)**

## **Application de techniques de discrimination**

### **LES OBJECTIFS DU SCORING**

- **SÉLECTION DES RISQUES**
- **PRÉVISION DES DÉFAUTS**
- **SUIVI ET CONTRÔLE**

*Pierre-Louis GONZALEZ*

## Différents types de score

- **Score d'appétence**

Score de propension à consommer, score d'affinité. C'est la probabilité d'un client d'être intéressé par un produit ou un service donné.

- **Score de comportement (risque)**

C'est la probabilité chez un client ouvrant un compte bancaire, souscrivant une carte de crédit, demandant un découvert ou contractant un crédit, de rencontrer un incident de paiement ou de remboursement.

## Différents types de score

- **Score d'octroi ( ou score d'acceptation)**

C'est un score de risque calculé pour un client qui est nouveau ou a une faible activité avec la banque. Le risque est calculé en temps réel, au moment où le client sollicite la banque, sur la base de données déclaratives éventuellement croisées avec des données de géomarketing.

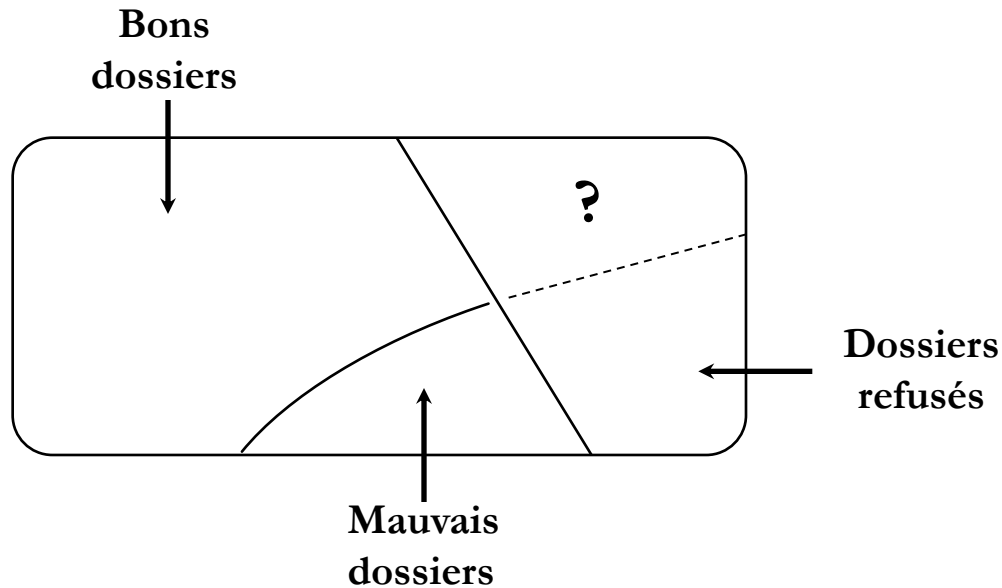
# LES DIFFÉRENTES ÉTAPES DE RÉALISATION

- 1. ÉCHANTILLONNAGE**
- 2. COLLECTE DE L'INFORMATION**
- 3. REDRESSEMENT**
- 4. SÉLECTION DES CRITÈRES**
- 5. CONSTRUCTION DU MODÈLE**
- 6. SIMULATION**
- 7. MISE EN ŒUVRE**

# 1. ÉCHANTILLONNAGE

**Objectif:** Construire un échantillon représentatif de la demande et du comportement payeur.

## 1-1. Prise en compte des dossiers refusés



## LES TROIS STRATES DE LA DEMANDE

### **Problème:**

Un score calculé uniquement sur les dossiers acceptés ne s'applique pas à l'ensemble de la demande.

# 1. ÉCHANTILLONNAGE

## 1.2 Prise en compte de la dimension temporelle

Deux plans possibles :

- **Observer une coupe instantanée**

**Inconvénient:** Certains dossiers sont considérés comme « bons » alors qu'ils deviendront « mauvais » par la suite.

- **Observer une population de dossiers terminés**

**Inconvénient:** La structure de la population observée ne correspond pas à la structure actuelle.

## 2. LA COLLECTE DE L'INFORMATION

### **Objectif:**

Bâtir un fichier contenant toutes les informations connues sur les refusés ainsi que les bons et mauvais payeurs.

### **Problèmes:**

- Pas de stockage informatique des observations individuelles
- Pas de conservation des dossiers refusés
- Pas de statistiques permettant d'élaborer le plan de sondage
- Historique trop court ou absent

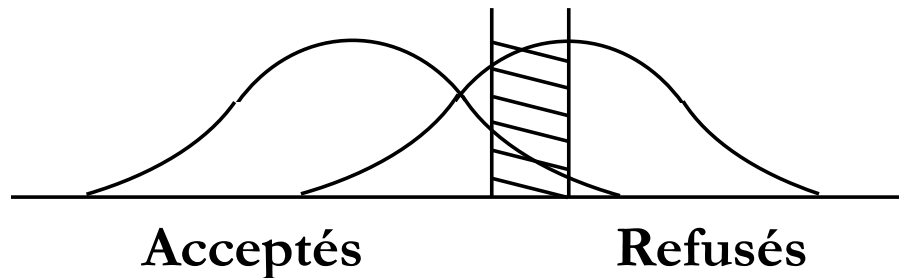
### 3. REDRESSEMENT

#### Objectif :

Redonner à l'échantillon la structure de la **demande actuelle**.

#### Deux familles de méthodes :

- **Score accepté/refusé**



**Hypothèse:** Les refusés d'une tranche ont le même comportement que les acceptés.

- **Simulation du comportement**

**Principe:** Chaque dossier refusé serait devenu bon (ou mauvais) avec une probabilité à estimer.



## 4. SÉLECTION DES CRITÈRES

### **Objectif:**

Lors de cette étape on choisit les variables et les interactions à introduire dans le modèle.

### **Différents types de problèmes se présentent:**

- Découpage/regroupement en catégories
- Choix des interactions
- Choix des variables les plus explicatives
- Choix des variables les moins corrélées entre elles

## 5. CONSTRUCTION DU MODÈLE

**Objectif:** Bâtir le système de pondération des variables (catégories).

### Revenu disponible par personne du ménage

Inférieur à 800 €	Entre 800 € et 1500 €	Plus de 1500 €
0	+ 12	+ 20

### Ratio d'endettement

Inférieur à 10 %	Entre 10 et 20 %	Entre 20 et 30 %	Plus de 30 %
+ 20	+ 16	+ 8	0

## 5. CONSTRUCTION DU MODÈLE

### Situation dans le logement

Propriétaire	Locataire
+ 10	0

### État matrimonial et nombre d'enfants à charge

Enfants	État matrimonial	
	Marié	Autres
0	+ 10	+ 8
1 ou 2	+ 20	+ 5
3 et plus	+ 16	0

← **interactions**

## 5. CONSTRUCTION DU MODÈLE

### Profession et stabilité dans l'emploi

Profession	Travaille dans le même emploi depuis		
	Moins de 4 ans	4 à 10 ans	Plus de 10 ans
Fonctionnaires, retraités	+ 18	+ 30	+ 30
Industriels, gros commerçants, Professions libérales, Cadres supérieurs Employés de bureau	+ 15	+ 22	+ 25
Artisans, petits commerçants Exploitants agricoles Cadres moyens	+ 5	+ 12	+ 17
Employés de commerce Ouvriers, Autres	0	+ 5	+ 10

## EXEMPLE: UTILISATION DU MODÈLE

<b>Ratio d'endettement : 15 %</b>	<b>16</b>
<b>Revenu disponible par personne : 1300 €</b>	<b>+ 12</b>
<b>Locataire :</b>	<b>+ 0</b>
<b>Marié, sans enfants :</b>	<b>+ 10</b>
<b>Employé de bureau : depuis 6 ans</b>	<b>+ 22</b>
<b>SCORE</b>	<b>+ 60</b>

## 6. SIMULATION

**L'objectif est de choisir une politique de mise en oeuvre**

Tranche de score	Nombre de demandes	Nombre de contentieux	Taux de contentieux		Nombre de demandes cumulées	Nombre de contentieux cumulés	Taux de contentieux cumulés
90 - 100	1 000	10	1 %		1 000	10	1 %
80 - 90	1 500	35	2,3 %		2 500	45	1,8 %
70 - 80	1 500	55	3,6 %		4 000	100	2,5 %
60 - 70	2 000	80	4 %		6 000	180	3 %
50 - 60	2 000	100	5 %		8 000	280	3,5 %
40 - 50	2 000	140	7 %		10 000	420	4,2 %
30 - 40	2 000	180	9 %		12 000	600	5 %
20 - 30	1 000	110	11 %		13 000	710	5,4 %
10 - 20	1 000	130	13 %		14 000	840	6 %
0 - 10	1 000	160	16 %		15 000	1 000	6,5 %

## 6. SIMULATION

	Avant le score		Politique n° 1		Politique n° 2
Seuil de sélection	-		Note supérieure à 50		Note supérieure à 40
Nombre de dossiers produits	10 000		8 000		10 000
Taux de refus	33 %		46 %		33 %
Contentieux sur la production	500		280		420
Taux de contentieux	5 %		3,5 %		4,2 %

## 7. LA MISE EN ŒUVRE

### **Objectif :**

Introduire le score comme outil de sélection, de prévision et de suivi.

### **Les problèmes:**

- Formation des utilisateurs.
- Mise en place des outils informatiques.
- Réactualisation: il est nécessaire de prévoir des outils d'alerte indiquant la détérioration du score: cartes de contrôle.



# COMPARAISONS DE SCORES

## Tables de classement

**Objectif:** Évaluer la capacité prédictive du modèle.

Dans le cas de deux groupes, en fonction d'un seuil  $s$  du score:

$$\begin{cases} \hat{Y}_i = 1 & \text{si score} \geq s \\ \hat{Y}_i = 0 & \text{si score} < s \end{cases}$$

		PRÉDICTION	
		$\hat{Y}_i = 1$	$\hat{Y}_i = 0$
OBSERVATION	$Y_i = 1$	a	b
	$Y_i = 0$	c	d

# COMPARAISONS DE SCORES

**SENSIBILITE**

$$\frac{a}{a + b}$$

**SPÉCIFICITÉ**

$$\frac{d}{c + d}$$

**TAUX D'ERREUR PAR EXCÈS**

$$\frac{c}{a + c}$$

**TAUX D'ERREUR PAR DÉFAUT**

$$\frac{b}{b + d}$$

**BIEN CLASSÉS**

$$\frac{a + d}{n}$$

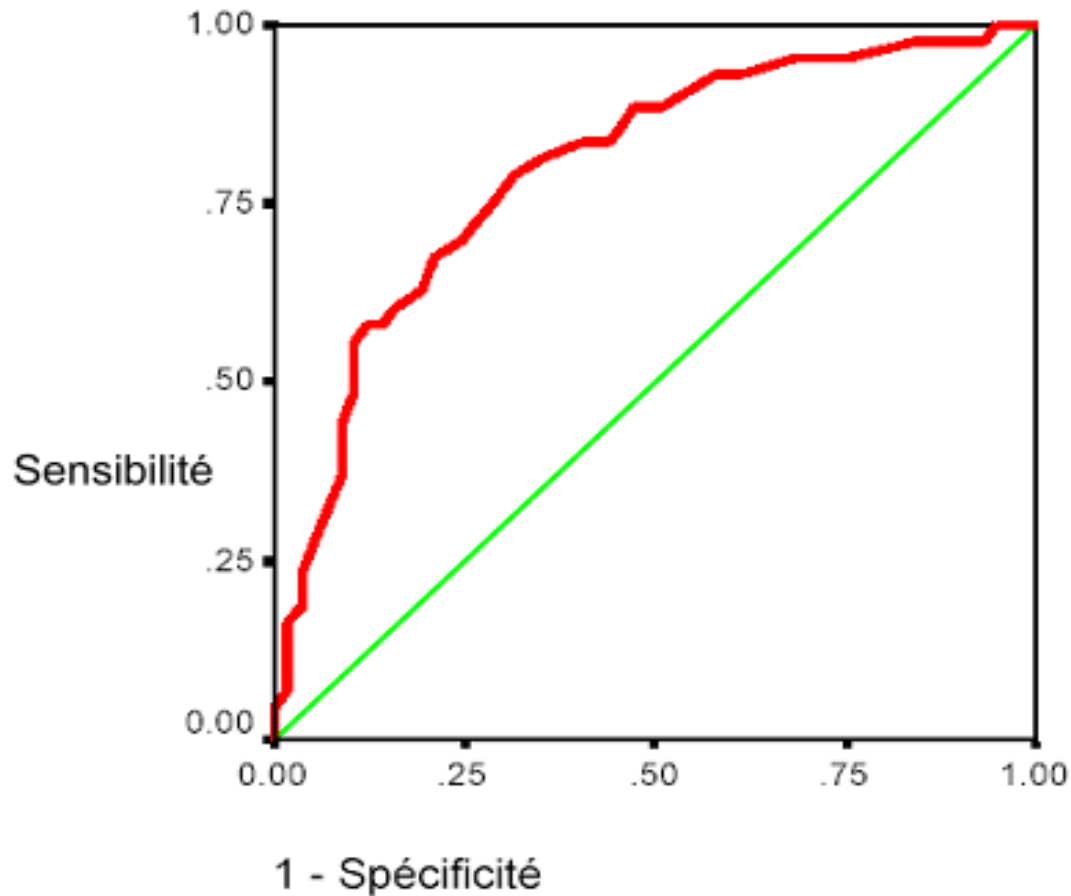
La capacité prédictive  
du modèle est d'autant  
meilleure que :

{ **SENSIBILITE**  
**SPÉCIFICITÉ**  
**BIEN CLASSÉS**

sont élevés.

# COMPARAISONS DE SCORES

Courbe ROC



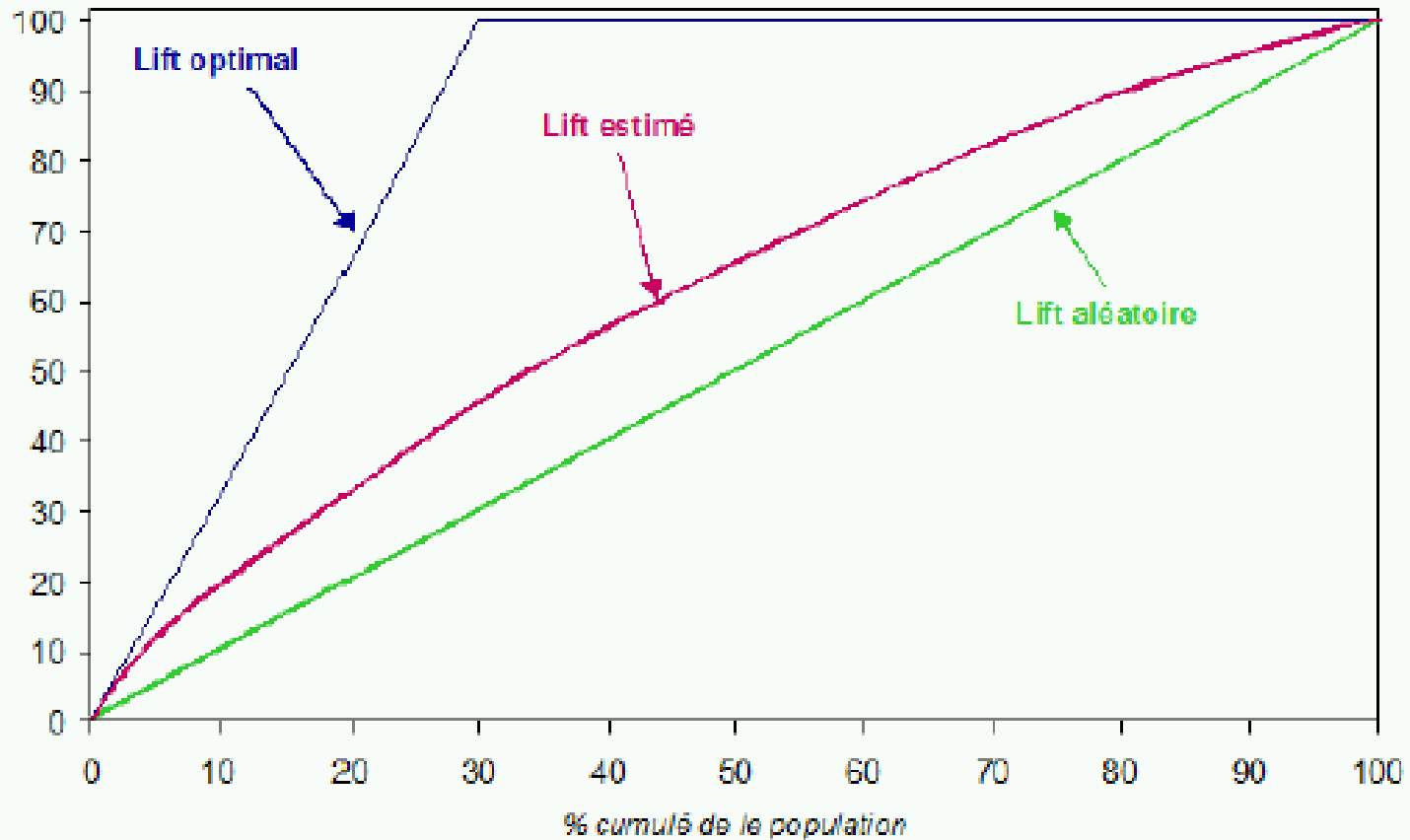
On peut calculer l'aire sous la courbe ROC pour comparer plusieurs modèles.

Il est également possible de déterminer des intervalles de confiance.

# COMPARAISONS DE SCORES

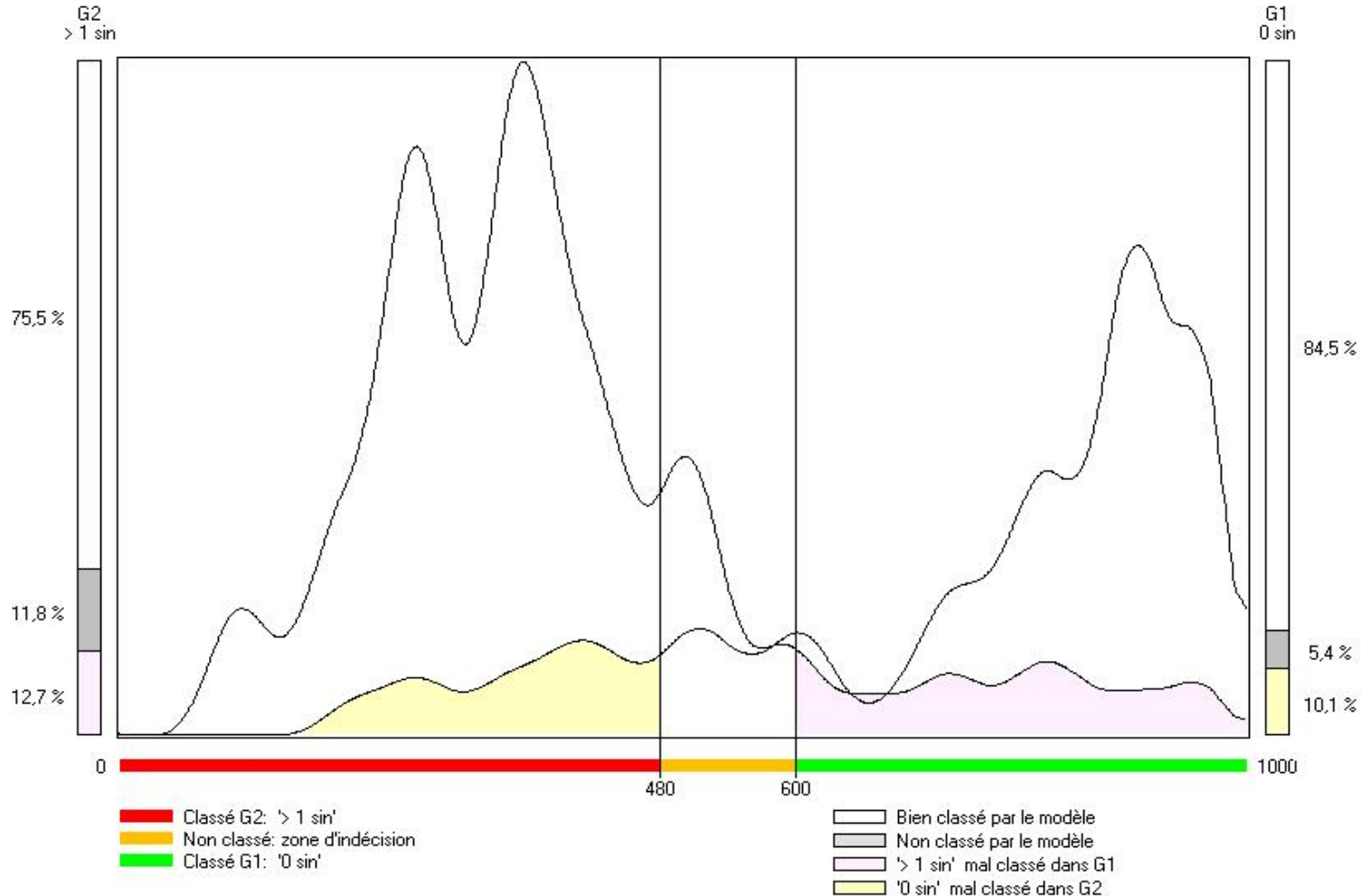
## Courbe de lift

*% de la cible*



# COMPARAISONS DE SCORES: FONCTIONS DE DENSITE

Variable cible: Sinistralité RC - SNB11 2-3 - Modalité cible: 0 sin



**Courbes**

Groupes  
 Total

Score inférieur: 480  
 TEC 1: 10,1 %

Score supérieur: 600  
 TEC 2: 12,7 %

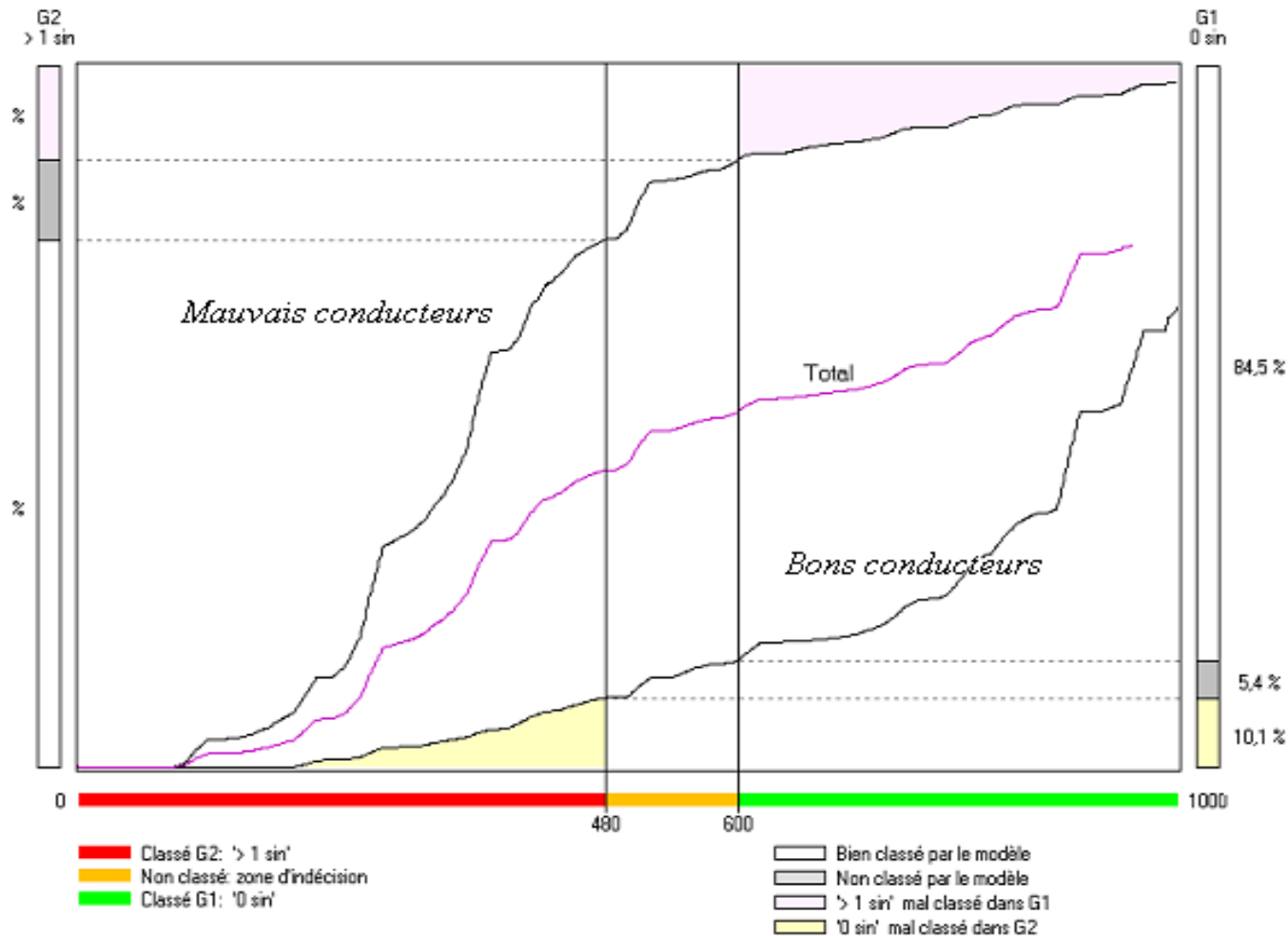
Score unique

Effectif réel:   
 % redressement G2:

Modèle  
 Données  
 Pourcentages  
 Volumes  
 Répartition  
 **Densité**  
 Lift  
 ROC

# COMPARAISONS DE SCORES: FONCTIONS DE REPARTITION

Variable cible: Sinistralité RC - SNB11 2-3 - Modalité cible: 0 sin



Courbes

Groupes

Total

Lissage

Score inférieur

480

TEC 1

10,1 %

Score supérieur

600

TEC 2

12,7 %

Score unique

Effectif réel

% redressement G2

Modèle

Données

Pourcentages

Volumes

▶ Répartition

Densité

Lift

ROC