# Méthodes de discrimination

## Partie IV – Comparaison de méthodes

2019

Brigitte Gelein – bgelein@ensai.fr
Marine Depecker

École nationale de la statistique et de l'analyse de l'information

B. Gelein
M. Depecker

# Sommaire


| | |
|---|---|
| Contexte | 4 |
| 1. Matrice de confusion | 5 |
| 2. Courbe ROC | 13 |
| 3. Aire sous la courbe ROC | 37 |
| 4. Courbe LIFT | 41 |
| 5. Aire sous la courbe LIFT | 48 |
| 6. Conclusion | 50 |




# Contexte

- Comparaison des résultats des méthodes de **discrimination (classement)**
  - *Analyse discriminante – (LDA ou QDA)*
  - *Arbre de classement – (Classification trees)*
  - *Bayésien naïf*
  - *Les K plus proches voisins*

- Problématique de **classement**
  - **Variable à expliquer *Y* nominale (ou ordinale)**
    - « Classes », « étiquettes », « labels », etc.
  - Variables explicatives *X* de nature diverse

# 1 - Matrice de confusion

**Contexte de la classification binaire**

**Principe :** confronter la vraie valeur avec la prédiction

Valeurs prédites

Valeurs observées

| | Y=1 | Y=0 | Total |
|---|---|---|---|
| Y=1 | n11 | n10 | n11+n10 |
| Y=0 | n01 | n00 | n01+n00 |
| Total | n11+n01 | n10+n00 | n |



1 - Matrice de confusion

**Principe :** confronter la vraie valeur avec la prédiction

Valeurs prédites

Valeurs observées

| | Y=1 | Y=0 | Total |
|---|---|---|---|
| Y=1 | **n11** | **n10** | n11+n10 |
| Y=0 | **n01** | **n00** | n01+n00 |
| Total | n11+n01 | n10+n00 | n |

# 1 - Matrice de confusion

| | | Valeurs prédites | | |
|---|---|---|---|---|
| Valeurs observées | | Y=1 | Y=0 | Total |
| | Y=1 | **n11** | **n10** | n11+n10 |
| | Y=0 | **n01** | **n00** | n01+n00 |
| | Total | n11+n01 | n10+n00 | n |

## Grandeurs d'intérêt

- Vrais positifs **VP = n11** ; Vrais négatifs **VN = n00**
- Faux positifs **FP = n01** ; Faux négatifs **FN = n10**

## Indicateurs

- **Taux de bien classés** $\frac{n11+n00}{n}$, estime P(Ŷ=Y)
- **Taux d'erreur** $\frac{n01+n10}{n}$, estime P(Ŷ≠Y)



# 1 - Matrice de confusion

| | | Valeurs prédites | | |
|---|---|---|---|---|
| Valeurs observées | | Y=1 | Y=0 | Total |
| | Y=1 | **n11** | **n10** | n11+n10 |
| | Y=0 | **n01** | **n00** | n01+n00 |
| | Total | n11+n01 | n10+n00 | n |

## Indicateurs (suite)

- **Taux de VP** $\frac{n11}{n11+n10}$, estime la **sensibilité** P(Ŷ=1/ Y=1)
- **Taux de FP** $\frac{n01}{n01+n00}$, estime P(Ŷ=1/ Y=0)
- **1–Taux de FP** $\frac{n00}{n01+n00}$, estime la **spécificité** P(Ŷ=0/ Y=0)

# 1 - Matrice de confusion

Pour vérifier que le % d'individus correctement classés est significativement meilleur que par un classement aléatoire, on calcule le **Q-Press** :

$$Q_{press} = \frac{(n-(c \times k))^2}{n \times (k-1)}$$

*Notations*

- n = taille échantillon
- k = nombre de groupes $\Rightarrow$ ici k= 2
- c = nombre d'individus bien classés $\Rightarrow$ c=n11+n00
- Sous H0 (classe comme le hasard) , Q-Press suit un $\chi^2$ à 1 degré de liberté



# 1 - Matrice de confusion

**Sensibilité et spécificité : une autre interprétation**

- Notion de **score**
  - **Discrimination** de deux groupes G1 (les positifs, Y=1) par rapport à G2 (les négatifs, Y=0) à partir d'un ***score***
  - ***Règle de décision :*** *si score ≥ seuil alors G1, sinon G2*
- En fonction du seuil ***z*** de séparation du score :
  - **Sensibilité** : *Sensi* (z) = P(score ≥ z / G1)
    - ➢ probabilité de bien détecter un positif
  - **Spécificité** : *Spéci* (z) = P(score < z / G2)
    - ➢ probabilité de bien détecter un négatif

## 1 - Matrice de confusion

**Lien entre *scoring* et classification binaire**

- Un classifieur binaire peut s'écrire sous la forme :

$$C(x) = 2 \cdot \mathrm{I}\left\{P(Y = +1 \mid X = x) \geq \frac{1}{2}\right\} - 1$$

où $\mathrm{I}\{.\}$ est la fonction indicatrice

- Ou plus généralement :

$$C_z(x) = 2 \cdot \mathrm{I}\{s(x) \geq z\} - 1$$

où z est un seuil et s une fonction de score



## 1 - Matrice de confusion

**Choix d'un classifieur : identification du seuil *z***

- Pour un modèle de score, on peut chercher la valeur de *z* qui maximise la sensibilité **ET** maximise la spécificité (minimise les faux positifs)

**Un bon classifieur permet de capturer le plus possible de vrais positifs avec le moins possible de faux positifs**

# 2 – Courbe ROC

**Contexte de la classification binaire**

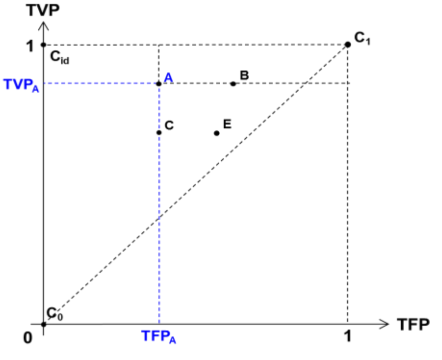
- **ROC** : « Receiver Operating Characteristic »
- Plan ROC
  - en ordonnée : estimation de la **sensibilité** (TVP)
  - en abscisse : estimation de **1 - spécificité** (TFP)
- Un **classifieur** est caractérisé par un **point unique** dans l'espace ROC
- **Le plan ROC permet de comparer des classifieurs (et/ou des fonctions de score)**



# 2 – Courbe ROC

*Exemple :*

- A est meilleur que C (même TFP mais TVP plus élevé)
- A est meilleur que B (même TVP mais TFP plus faible)
- C est meilleur que E (même TVP mais TFP plus faible)
- B semble meilleur que E (presque même TFP mais TVP plus élevé)




- **E est globalement plus mauvais**
- **A est globalement meilleur**

## 2 – Courbe ROC

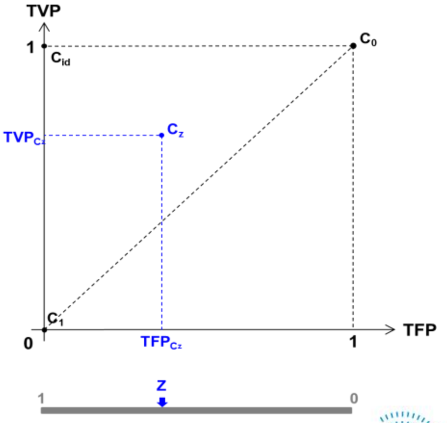
### Point de vue du scoring

- Un classifieur s'écrit

$$C_z(x) = 2 \cdot \mathrm{I}\{s(x) \geq z\} - 1$$

- En faisant varier le seuil z, on obtient des couples ($TFP_z$, $TVP_z$)

- Chaque point caractérise un classifieur qui correspond à la fonction de score s avec un seuil spécifié z




B. Gelein
M. Depecker



## 2 – Courbe ROC

### Point de vue du scoring

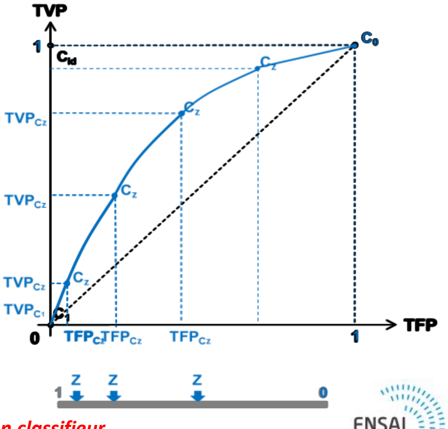
- Un classifieur s'écrit

$$C_z(x) = 2 \cdot \mathrm{I}\{s(x) \geq z\} - 1$$

- En faisant varier le seuil z, on obtient des couples ($TFP_z$, $TVP_z$)

- Chaque point caractérise un classifieur qui correspond à la fonction de score s avec un seuil spécifié z

➡ En faisant varier le seuil z, on définit la courbe ROC



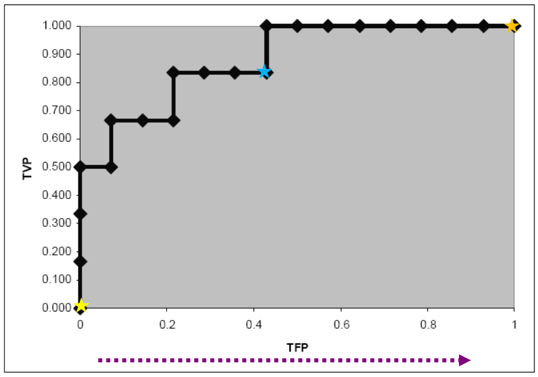

***Chaque point de la courbe ROC caractérise un classifieur qui correspond à la fonction de score s avec un seuil spécifié z***

B. Gelein
M. Depecker

| Individu | Score | Y observé | TFP | TVP |
|---|---|---|---|---|
| | | | 0 | 0.000 |
| 1 | 1 | 1 | 0.000 | 0.167 |
| 2 | 0.95 | 1 | 0.000 | 0.333 |
| 3 | 0.9 | 1 | 0.000 | 0.500 |
| 4 | 0.85 | 0 | 0.071 | 0.500 |
| 5 | 0.8 | 1 | 0.071 | 0.667 |
| 6 | 0.75 | 0 | 0.143 | 0.667 |
| 7 | 0.7 | 0 | 0.214 | 0.667 |
| 8 | 0.65 | 1 | 0.214 | 0.833 |
| 9 | 0.6 | 0 | 0.286 | 0.833 |
| 10 | 0.55 | 0 | 0.357 | 0.833 |
| 11 | 0.5 | 0 | 0.429 | 0.833 |
| 12 | 0.45 | 1 | 0.429 | 1.000 |
| 13 | 0.4 | 0 | 0.500 | 1.000 |
| 14 | 0.35 | 0 | 0.571 | 1.000 |
| 15 | 0.3 | 0 | 0.643 | 1.000 |
| 16 | 0.25 | 0 | 0.714 | 1.000 |
| 17 | 0.2 | 0 | 0.786 | 1.000 |
| 18 | 0.15 | 0 | 0.857 | 1.000 |
| 19 | 0.1 | 0 | 0.929 | 1.000 |
| 20 | 0.05 | 0 | 1.000 | 1.000 |

## Exemple

Courbe ROC




**« meilleurs »** **« moins bons »**

Source : R. Rakotomalala

B. Gelein
M. Depecker

| Individu | Score | Y observé | TFP | TVP |
|---|---|---|---|---|
| | | | 0 | 0.000 |
| 1 | 1 | 1 | 0.000 | 0.167 |
| 2 | 0.95 | 1 | 0.000 | 0.333 |
| 3 | 0.9 | 1 | 0.000 | 0.500 |
| 4 | 0.85 | 0 | 0.071 | 0.500 |
| 5 | 0.8 | 1 | 0.071 | 0.667 |
| 6 | 0.75 | 0 | 0.143 | 0.667 |
| 7 | 0.7 | 0 | 0.214 | 0.667 |
| 8 | 0.65 | 1 | 0.214 | 0.833 |
| 9 | 0.6 | 0 | 0.286 | 0.833 |
| 10 | 0.55 | 0 | 0.357 | 0.833 |
| 11 | 0.5 | 0 | 0.429 | 0.833 |
| 12 | 0.45 | 1 | 0.429 | 1.000 |
| 13 | 0.4 | 0 | 0.500 | 1.000 |
| 14 | 0.35 | 0 | 0.571 | 1.000 |
| 15 | 0.3 | 0 | 0.643 | 1.000 |
| 16 | 0.25 | 0 | 0.714 | 1.000 |
| 17 | 0.2 | 0 | 0.786 | 1.000 |
| 18 | 0.15 | 0 | 0.857 | 1.000 |
| 19 | 0.1 | 0 | 0.929 | 1.000 |
| 20 | 0.05 | 0 | 1.000 | 1.000 |

**Au seuil z = 1**

| | Prédit | | |
|---|---|---|---|
| Observé | Y=1 | Y=0 | Total |
| Y=1 | **1** | **5** | 6 |
| Y=0 | **0** | **14** | 14 |
| Total | 1 | 19 | 20 |

TVP=1/6=0,167 TFP=0/14=0

**Au seuil z = 0,95**

| | Prédit | | |
|---|---|---|---|
| Observé | Y=1 | Y=0 | Total |
| Y=1 | **2** | **4** | 6 |
| Y=0 | **0** | **14** | 14 |
| Total | 2 | 18 | 20 |

TVP=2/6=0,333 TFP=0/14=0

Source : R. Rakotomalala

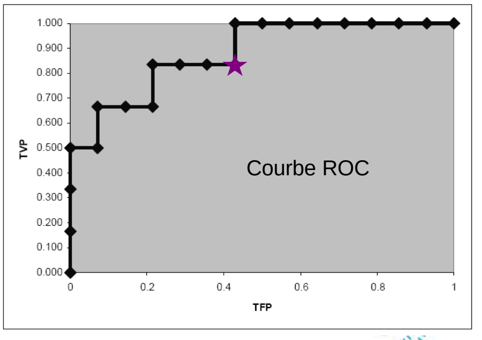
B. Gelein
M. Depecker

| Individu | Score | Y observé | TFP | TVP |
|---|---|---|---|---|
| | | | 0 | 0.000 |
| 1 | 1 | 1 | 0.000 | 0.167 |
| 2 | 0.95 | 1 | 0.000 | 0.333 |
| 3 | 0.9 | 1 | 0.000 | 0.500 |
| 4 | 0.85 | 0 | 0.071 | 0.500 |
| 5 | 0.8 | 1 | 0.071 | 0.667 |
| 6 | 0.75 | 0 | 0.143 | 0.667 |
| 7 | 0.7 | 0 | 0.214 | 0.667 |
| 8 | 0.65 | 1 | 0.214 | 0.833 |
| 9 | 0.6 | 0 | 0.286 | 0.833 |
| 10 | 0.55 | 0 | 0.357 | 0.833 |
| **11** | **0.5** | **0** | **0.429** | **0.833** |
| 12 | 0.45 | 1 | 0.429 | 1.000 |
| 13 | 0.4 | 0 | 0.500 | 1.000 |
| 14 | 0.35 | 0 | 0.571 | 1.000 |
| 15 | 0.3 | 0 | 0.643 | 1.000 |
| 16 | 0.25 | 0 | 0.714 | 1.000 |
| 17 | 0.2 | 0 | 0.786 | 1.000 |
| 18 | 0.15 | 0 | 0.857 | 1.000 |
| 19 | 0.1 | 0 | 0.929 | 1.000 |
| 20 | 0.05 | 0 | 1.000 | 1.000 |

**Au seuil z = 0,5**

| | Prédit | | |
|---|---|---|---|
| Observé | Y=1 | Y=0 | Total |
| Y=1 | 5 | 1 | 6 |
| Y=0 | 6 | 8 | 14 |
| Total | 11 | 9 | 20 |

TVP=5/6=0,833 TFP=6/14=0,429




Source : R. Rakotomalala

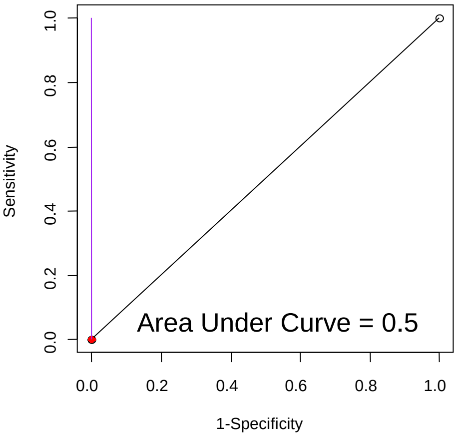


**Package R : TeachingDemos**

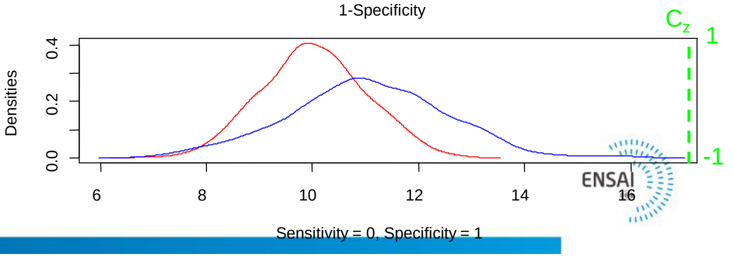
roc.demo(x = rnorm(25, 10, 1),
y = rnorm(25, 11, 1.5))

<u>**Ici paramètre x de roc.demo :**</u>
distribution du score pour
la modalité 0 de Y

<u>**Ici paramètre y de roc.demo :**</u>
distribution du score pour
la modalité 1 de Y




- Densité des positifs : *bleu*
- Densité des négatifs : *rouge*

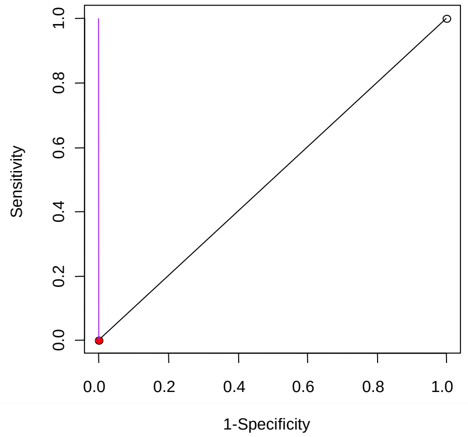


**Package R : TeachingDemos**

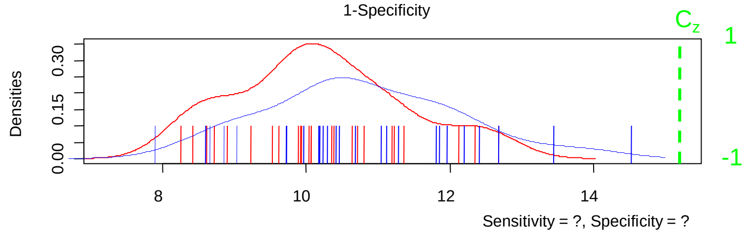
roc.demo(x = rnorm(25, 10, 1),
 y = rnorm(25, 11, 1.5))

**<u>Ici paramètre x de roc.demo :</u>**
distribution du score pour
la modalité 0 de Y

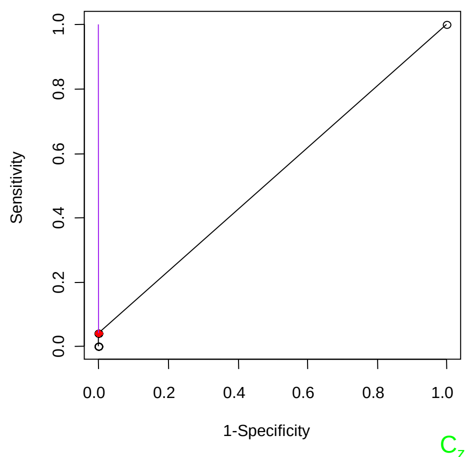
**<u>Ici paramètre y de roc.demo :</u>**
distribution du score pour
la modalité 1 de Y

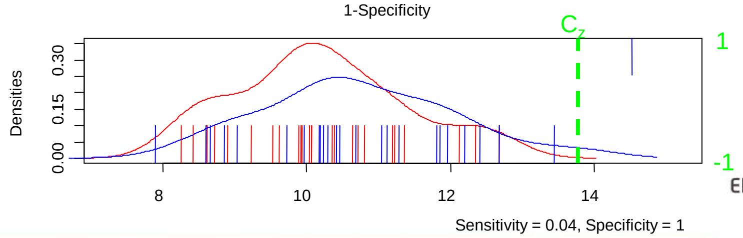



- Echantillonnage
  - *Positifs*
  - *Négatifs*



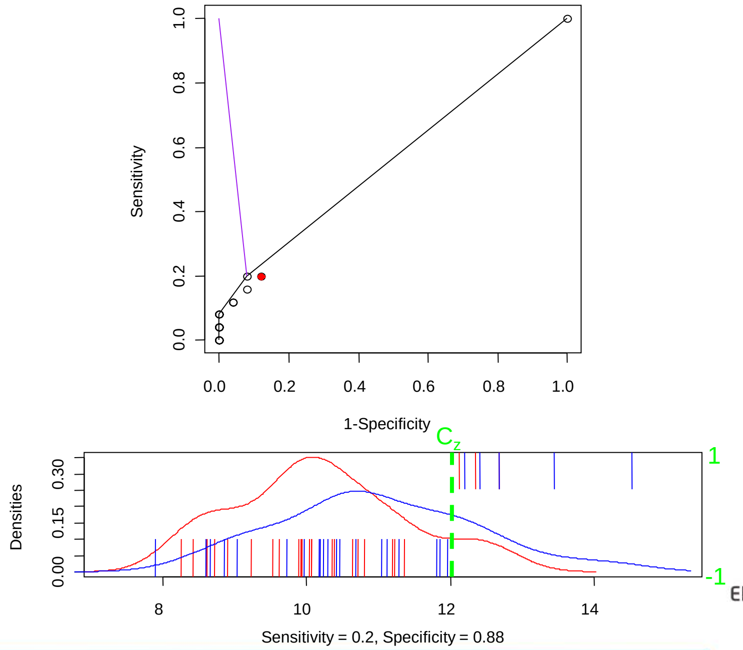

B. Gelein
M. Depecker



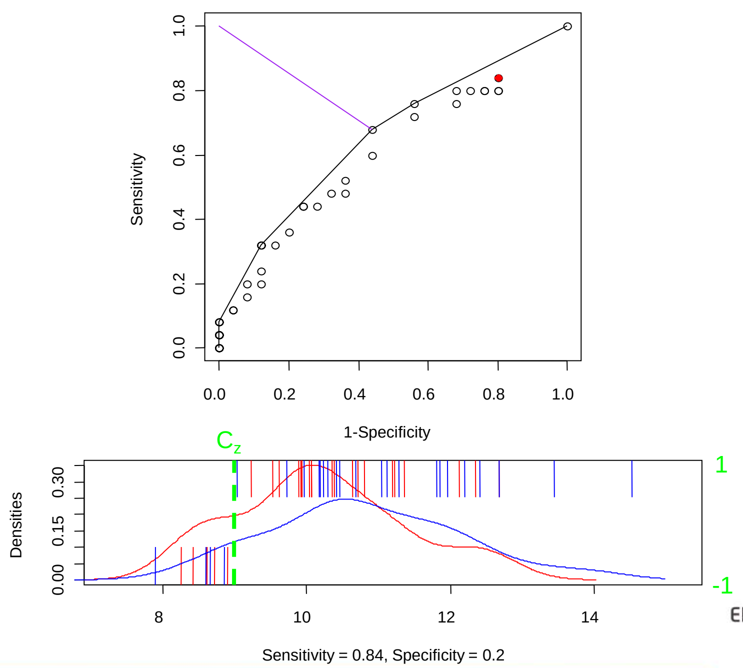




ENSAI

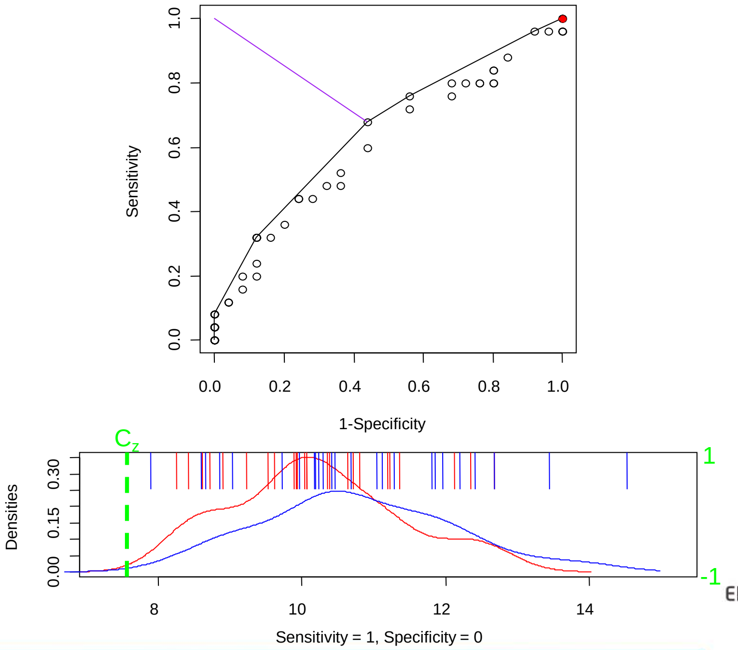
B. Gelein
M. Depecker




B. Gelein
M. Depecker




B. Gelein
M. Depecker

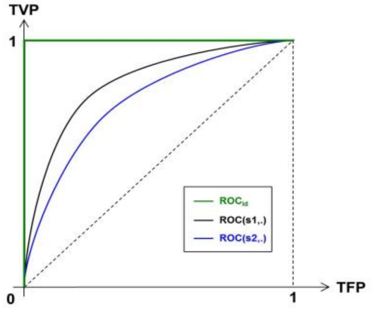





## 2 – Courbe ROC

# Propriétés de la ROC : une courbe paramétrique

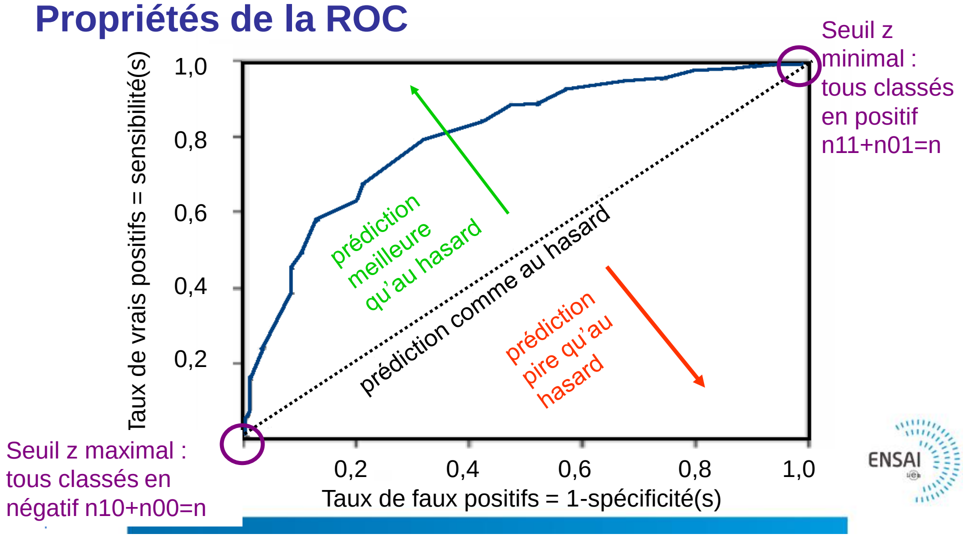
- Collection de classifieurs binaires : $\{C_z(x) = 2 \cdot \mathbb{I}\{s(x) \geq z\} - 1, z \in \mathbb{R}\}$
- Taux de faux positifs (TFP) : $\text{TFP}_s(z) = \mathbb{P}\{s(X) \geq z \mid Y = -1\}$
- Taux de vrais positifs (TVP) : $\text{TVP}_s(z) = \mathbb{P}\{s(X) \geq z \mid Y = +1\}$




ROC : $z \mapsto (\text{TFP}_s(z), \text{TVP}_s(z))$

## 2 – Courbe ROC

### Propriétés de la ROC






## 2 – Courbe ROC

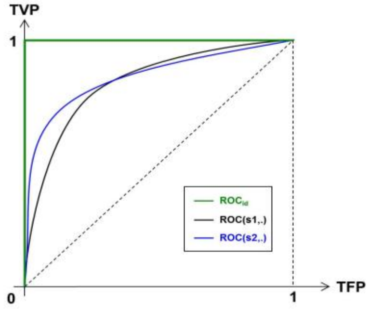
### Propriétés de la ROC

- Elles est **invariante** pour toute transformation monotone croissante du score
- **C'est un outil de comparaison de modèles (scores et classifieurs)**

B. Gelein
M. Depecker

# 2 – Courbe ROC

## Propriétés de la ROC




- La courbe ROC induit un **ordre partiel** sur l'ensemble des fonctions de score $\mathcal{S}$
- $s_1$ est plus performante que $s_2$ ***ssi***

$$\forall \alpha \in \left]0,1\right[,\ \mathrm{ROC}(s_1,\alpha) \geq \mathrm{ROC}(s_2,\alpha)$$

- La courbe ROC optimale correspond à la probabilité $\eta$

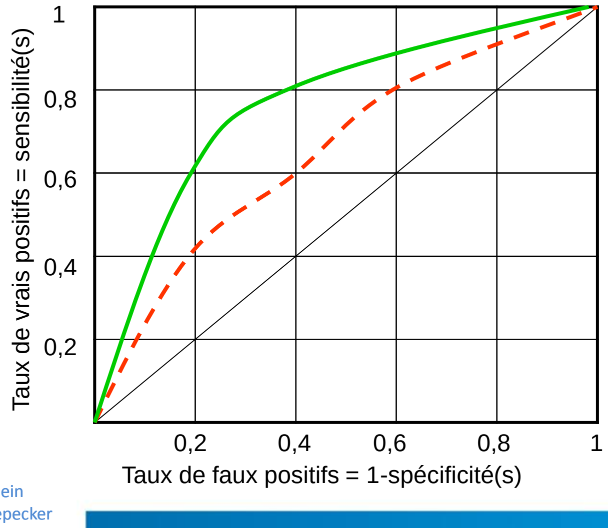
$$\forall s \in \mathcal{S}, \forall \alpha \in \left]0,1\right[,$$
$$\mathrm{ROC}^*(\alpha) = \mathrm{ROC}(\eta,\alpha) \geq \mathrm{ROC}(s,\alpha)$$

***(argument de Neymann-Pearson)***

où $\quad \eta(x) = P(Y = +1 \mid X = x)$



# 2 – Courbe ROC



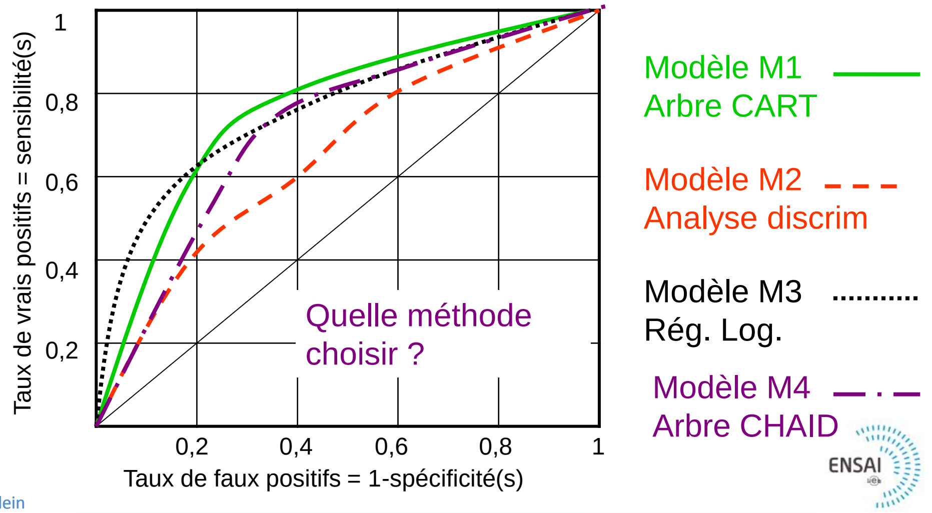

Modèle(s) M1

Modèle(s) M2

La courbe de M1 est toujours « au-dessus » de celle de M2 :

les classifieurs de M1 sont meilleurs en prédiction quel que soit le seuil z

## 2 – Courbe ROC



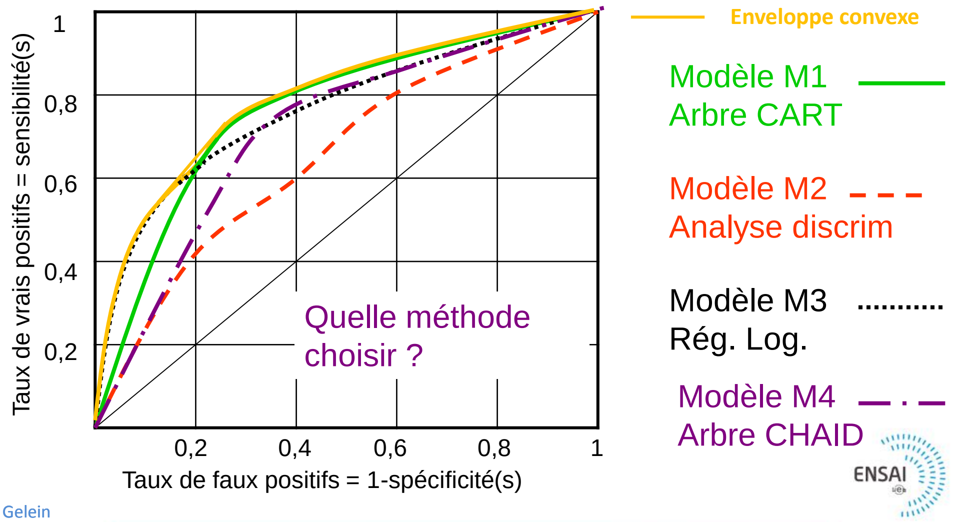



## 2 – Courbe ROC

**Enveloppe convexe – ou comment éliminer d'office les modèles les moins intéressants ?**

- L'**enveloppe convexe** permet d'effectuer une première sélection dans un **ensemble de modèles** donné
- Elle est formée par **les courbes** ou **parties de courbes**, telles qu'il n'existe pas d'autre courbe « au-dessus » d'elles
- Les courbes situées **sur cette enveloppe** correspondent aux **modèles les plus performants** pour une matrice de coût donnée

## 2 – Courbe ROC






## 2 – Courbe ROC

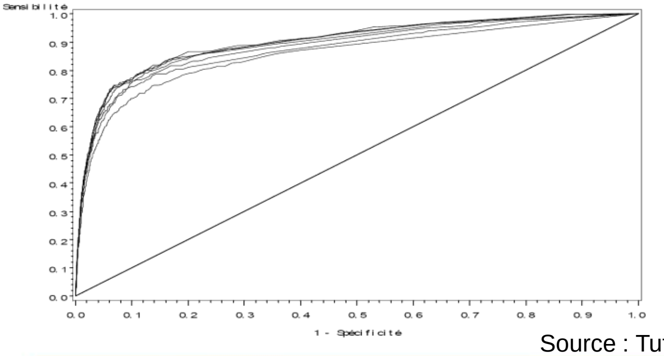
**Enveloppe convexe – ou comment éliminer d'office les modèles les moins intéressants ?**

- Sont **éliminés d'office** les modèles ne participant jamais à cette enveloppe
- Dans l'exemple, l'enveloppe convexe est formée par les courbes de M1 (arbre de classement CART) et M3 (régression logistique).
  - M2 est dominé par tous les modèles, il est donc éliminé.
  - M4 peut être meilleur que M3 dans certains cas, mais pour ces cas là, il s'avère moins bon que M2. M4 est donc éliminé.

## 2 – Courbe ROC

### Autre usage de la courbe ROC

- On peut aussi tracer les courbes ROC correspondant à une entrée progressive de variables dans un modèle



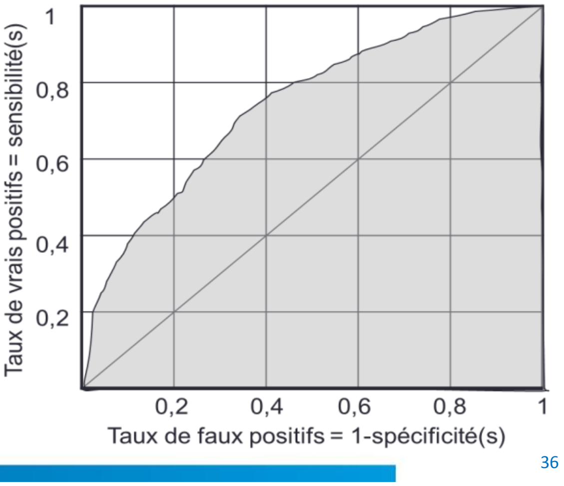

Source : Tufféry



## 3 – Aire sous la courbe ROC

### Aire AUC sous la courbe ROC

$$\mathrm{AUC}(s) = \int_{\alpha=0}^{1} \mathrm{ROC}(s,\alpha)d\alpha$$



## 3 – Aire sous la courbe ROC

**Aire AUC sous la courbe ROC**

- La **capacité prédictive** d'un modèle est d'autant meilleure que l'aire **AUC est proche de 1**
- Si l'**AUC = 0,5** alors le modèle n'est pas meilleur qu'une **prédiction aléatoire** (ROC = diagonale)
- Estimation de la probabilité que pour tout couple (A,B)
  score(individu A)> score(individu B),
  avec A tiré au hasard dans le groupe G1 (à prédire, par ex « positif ») et B dans le groupe G2
  Interprétation : *taux de paires concordantes*



## 3 – Aire sous la courbe ROC

**AUC - trois méthodes d'estimation**

1. Méthode des trapèzes
2. Interprétation probabiliste: taux de paires concordantes

$$\forall (X, X') \in \mathcal{X}^2,\ s \in \mathcal{S} \text{ et pour } p = \mathbb{P}\{Y = 1\},$$

$$\begin{aligned}
\mathrm{AUC}(s) &= \mathbb{P}\left\{s(X) > s(X') \mid (Y, Y') = (+1, -1)\right\} \\
&+ \frac{1}{2} \cdot \mathbb{P}\left\{s(X) = s(X') \mid (Y, Y') = (+1, -1)\right\}. \\
&= 1 - \frac{1}{2p(1-p)} \cdot \mathbb{P}\left\{(s(X) - s(X'))(Y - Y') < 0\right\}
\end{aligned}$$

## 3 – Aire sous la courbe ROC

**Estimation du taux de paire concordantes :**

- Soit n1 (resp. n2) le nombre d'observations dans G1 (resp. G2)
- Soient les n1*n2 paires formées d'un individu x1 du groupe G1 et d'un individu x2 du groupe G2
- Parmi ces n1*n2 paires on a :
  - concordance si score(x1) >score(x2)
  - discordance si score(x1) < score(x2)
  - nc = nombre de paires concordantes
  - nd = nombre de paires discordantes
  - ne =n1*n2 - nc - nd = nombre d'ex æquo

$$AUC \approx \frac{nc + \frac{1}{2}\left(n1 \times n2 - nc - nd\right)}{n1 \times n2}$$



## 3 – Aire sous la courbe ROC

**3. Méthode de Mann-Whitney**

- L'AUC peut s'exprimer en fonction de la statistique de test de Mann-Whitney *U* : $AUC = \frac{U}{n1 \times n2}$
- *U* est une statistique de test non-paramétrique permettant d'évaluer l'homogénéité entre deux populations où $R_1$ (resp. $R_2$) est la somme des rangs des individus de $G_1$ (resp. $G_2$), et

$$U = R_1 - \frac{n1(n1+1)}{2} = R_2 - \frac{n2(n2+1)}{2}$$

$$R_1 + R_2 = \frac{N(N+1)}{2}, \qquad avec\ N = n1 + n2$$

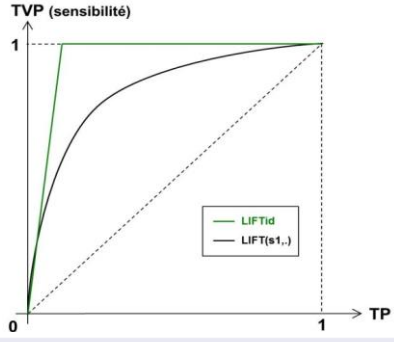
# 4 – Courbe LIFT

**Courbe LIFT ou courbe de gain**

- Objectif : La démarche du ciblage - marketing

*Exemple :* publipostage pour la promotion d'un produit

- Taux de répondants *vs* taux d'individus ciblés
  - Ordonnée :
    estimation de la sensibilité
    Proba(score ≥ s / G1) (TVP)
  - Abscisse :
    estimation du taux de positifs
    Proba(score ≥ s) (TP)




*LIFT* : $z \mapsto (\mathrm{TP}_s(z), \mathrm{TVP}_s(z))$



# 4 – Courbe LIFT

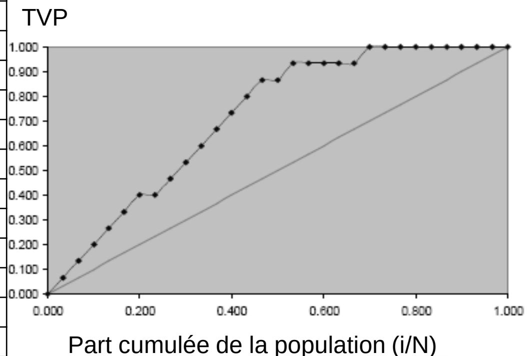
**Courbe LIFT ou courbe de gain**

- Cette courbe représente la proportion de vrais positifs en fonction des individus sélectionnés, lorsque l'on fait varier le seuil z du score
- **Sa forme dépend du taux de positifs *a priori***
- Même ordonnée que la courbe ROC, mais une abscisse généralement plus grande
- La courbe de lift est généralement sous la courbe ROC
- **ROC et LIFT représentent une information similaire**
  - La diagonale du plan représente une prédiction aléatoire
  - La courbe idéale LIFT est la plus proche du coin supérieur gauche

| Individu | Y observé | Score | Part cum. de Population | TVP |
|---|---|---|---|---|
| | | | 0 | 0.000 |
| 1 | 1 | 1.000 | 0.033 | 0.067 |
| 2 | 1 | 1.000 | 0.067 | 0.133 |
| 3 | 1 | 0.999 | 0.100 | 0.200 |
| 4 | 1 | 0.999 | 0.133 | 0.267 |
| 5 | 1 | 0.998 | 0.167 | 0.333 |
| 6 | 1 | 0.992 | 0.200 | 0.400 |
| … | … | … | … | … |
| 19 | 0 | 0.294 | 0.633 | 0.933 |
| 20 | 0 | 0.109 | 0.667 | 0.933 |
| 21 | 1 | 0.073 | 0.700 | 1.000 |
| 22 | 0 | 0.035 | 0.733 | 1.000 |
| 23 | 0 | 0.024 | 0.767 | 1.000 |
| 24 | 0 | 0.016 | 0.800 | 1.000 |
| 25 | 0 | 0.015 | 0.833 | 1.000 |
| 26 | 0 | 0.009 | 0.867 | 1.000 |
| 27 | 0 | 0.004 | 0.900 | 1.000 |
| 28 | 0 | 0.003 | 0.933 | 1.000 |
| 29 | 0 | 0.002 | 0.967 | 1.000 |
| 30 | 0 | 0.000 | 1.000 | 1.000 |

## Courbe LIFT

Les données sont triées selon les scores décroissants

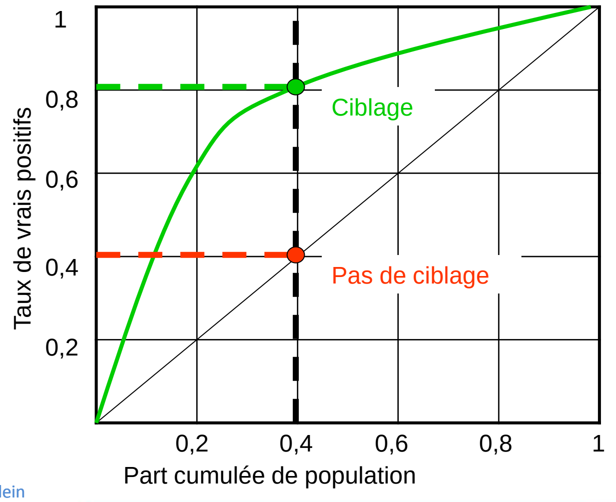



N=30 et N(Y=1)=15

Source : R. Rakotomalala

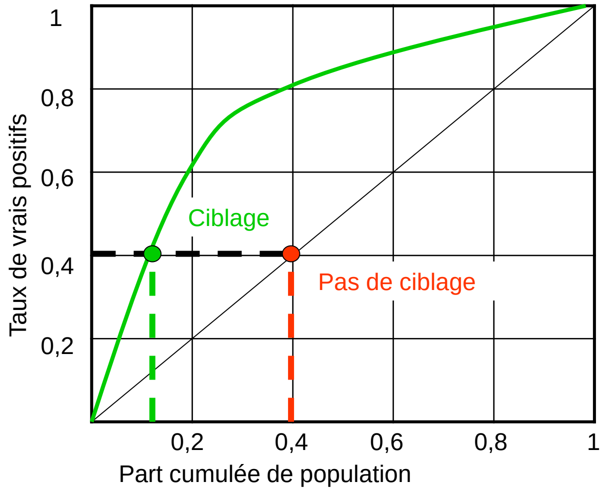


## 4 – Courbe LIFT




Optique
« budget fixé »

# 4 – Courbe LIFT

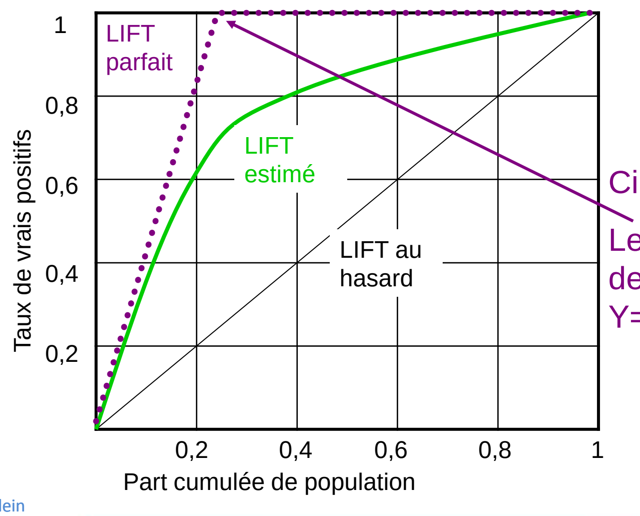



**Objectif =**
part de Y=1

Exemple : « part de marché »



# 4 – Courbe LIFT




Ciblage parfait :

Les premiers individus de la base ont tous Y=1

# 4 – Courbe LIFT

### Lien entre courbe de LIFT et courbe ROC

- Si la courbe ROC(s) domine dans le plan (TVP,TFP) alors la courbe LIFT(s) domine dans le plan (TVP,TP) et la fonction de score s est la plus performante



# 5 – Aire sous la courbe LIFT

### Lien entre les aires sous les courbes LIFT et ROC

- On note souvent AUC, parfois AUL l'aire sous la courbe LIFT

- L'AUL s'exprime simplement à partir de l'AUC

$$AUL = \frac{p}{2} + (1 - p)\,AUC$$

où p = Proba(G1) = probabilité a priori de l'événement Y=1 dans la population

## 5 – Aire sous la courbe LIFT

**Lien entre courbe de LIFT et courbe ROC**

- Cas particuliers :
  - AUC = 1 ⇔ AUL = p/2 + (1 – p) = 1 – p/2
  - AUC = 0,5 ⇔ AUL = p/2 + 1/2 – p/2 = 0,5
  - Si p est petite ⇔ AUC et AUL sont proches
  - AUC(M1) > AUC (M2) ⇔ AUL (M1) > AUL (M2)



## 6 - Conclusion
## 6.1 – Avantages et Limites

**Courbes ROC et LIFT**

- Des critères fonctionnels
  - Faciles à visualiser *a posteriori*
  - Difficiles à optimiser directement

  **Alternative : les aires sous les courbes et critères dérivés**
- La courbe ROC ne permet pas de visualiser le rapport entre les deux classes

  **Difficulté : elle peut induire en erreur dans le cas d'échantillons fortement déséquilibrés**
- La forme de la courbe LIFT dépend de l'équilibre des classes dans l'échantillon

  **Difficulté : elle doit être lue *relativement* à la LIFT idéale**

# Conclusion
# 6.1 – Avantages et Limites

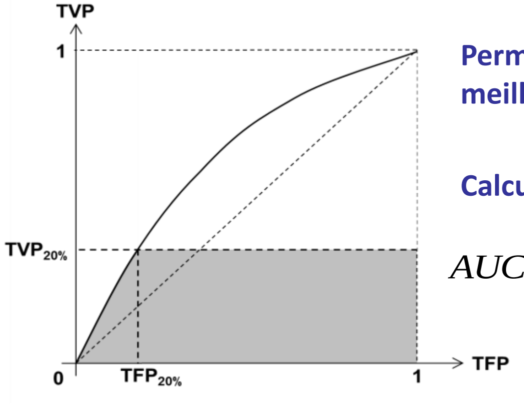
**AUC et AUL**

- Des critères réels
  - Simples à calculer et à optimiser
  - L'AUC a une interprétation probabiliste intuitive
- Des critères globaux
  - deux fonctions scores optimales auront **même AUC** mais des **courbes ROC** possiblement **différentes**…
  - …l'une pouvant être **meilleure que l'autre sur une portion** des individus!
  - Ne permettent pas de se focaliser sur les **meilleures observations**



# Conclusion
# 6.2 – Alternatives

**Exemple: l'AUC partielle**




**Permet de se focaliser sur les x% meilleures observations**

**Calcul simple :**

$$AUC_{tronquée\ x\%} + \left(1 - TFP_{x\%}\right) \times TVP_{x\%}$$

S.Clémençon & N. Vayatis, *Ranking the best instances*, Journal of Machine Learning Research, vol 8, pp 2671-2699, 2007