

# SY09 - TP03

## Théorie de la décision

Bertrand Bon - Antoine Hars

June 15, 2013

### Introduction

Le but de ce tp est de mettre en pratique deux théorie de la décision qui sont le classifieur euclidien et la règle de Bayes.

### Exercice 1 : Classifieur euclidien

Le but de la première partie de ce tp est d'étudier les performances du classifieur euclidien sur des échantillons issus de deux classes  $\omega_1$  et  $\omega_2$  de  $\mathbb{R}^2$  dont les distributions sont normales et de paramètres  $(\mu_1, \Sigma_1)$  et  $(\mu_2, \Sigma_2)$ .

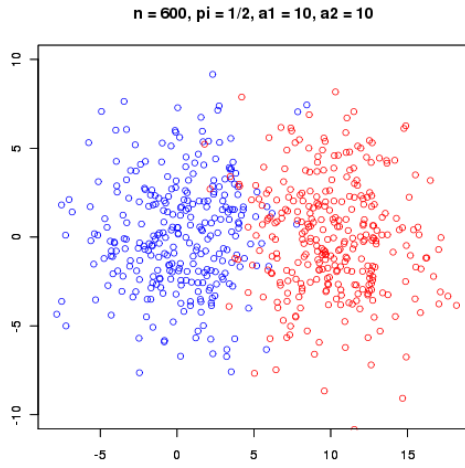
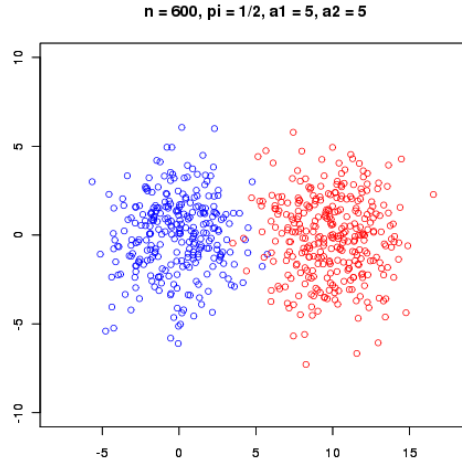
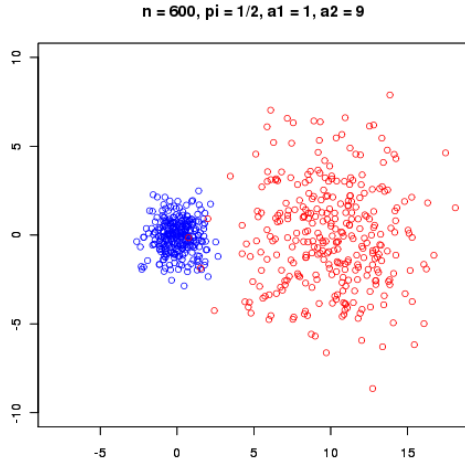
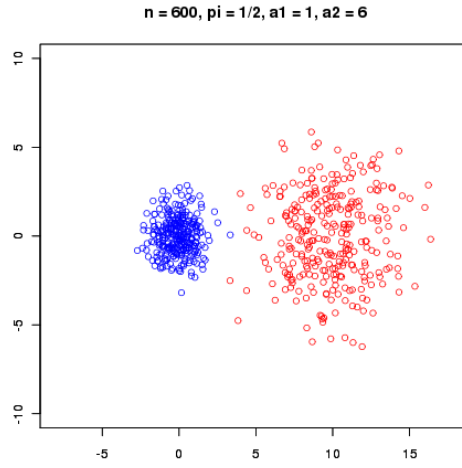
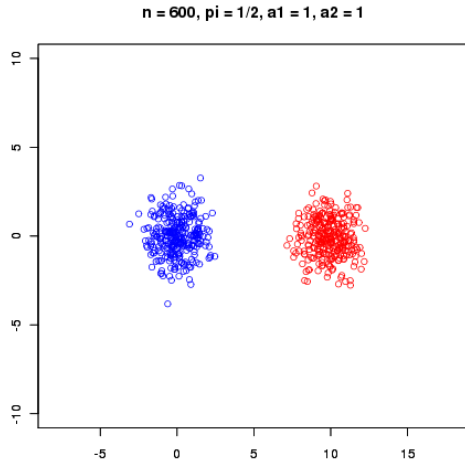
#### 1. Simulation d'un échantillon.

En utilisant la fonction **mvrnorm** de la bibliothèque **MASS**, nous avons écrit la fonction **simul** qui nous retourne un échantillon de taille  $n$  tiré suivant une proportion  $\pi$  d'exemples issus de la classe  $\omega_1$ . Pour chaque exemple, nous avons, dans un premier temps, tiré au hasard la classe dont il est issu, avant de le générer en utilisant les paramètres adéquats :

```
simul <- function (n, pi, mu1, mu2, sigma1, sigma2) {  
  
  result = matrix(nrow = n, ncol = 3)  
  for (k in 1:n) {  
    rand = sample(0:1, 1)  
    result[k, 3] = rand  
  
    if (result[k, 3] == 0) {  
      result[k, c(1, 2)] = mvrnorm(1, mu1, sigma1)  
    } else {  
      result[k, c(1, 2)] = mvrnorm(1, mu2, sigma2)  
    }  
  }  
  return (result)  
}
```

Nous avons utilisé pour cette fonction les cinq situations suivantes :

$n = 600$ ,  $\pi = 1/2$ ,  $\mu_1 = (0,0)^T$ ,  $\mu_2 = (10,0)^T$ ,  $\Sigma_1 = a_1 \text{Id}$  et  $\Sigma_2 = a_2 \text{Id}$  avec  $(a_1 = 1, a_2 = 1)$ ,  $(a_1 = 1, a_2 = 6)$ ,  $(a_1 = 1, a_2 = 9)$ ,  $(a_1 = 5, a_2 = 5)$  et  $(a_1 = 10, a_2 = 10)$  :



La première classe est représentée en bleu et la seconde en rouge.

Nous pouvons observer que plus la variance  $\Sigma$  d'une loi normale augmente, plus l'échantillon lui correspondant est dispersé. Dans ce cas, plus le coefficient  $a$  de la variance augmente, plus les individus des deux classes ont tendance à se mélanger, ce qui augmente la probabilité d'erreur entre les deux classes.

## 2. Estimation de la probabilité d'erreur.

Pour chacun des cinq situations, nous avons cherché à estimer la probabilité d'erreur associée au classifieur euclidien.

Pour cela, l'échantillon simulé est coupé en deux : la première moitié forme un échantillon d'apprentissage permettant d'estimer les moyennes  $\mu_1$  et  $\mu_2$  et la seconde moitié forme un ensemble test permettant d'estimer le taux d'erreur.

Pour réaliser cela, nous avons écrit plusieurs fonctions :

- La fonction **calculMu** qui va nous calculer les estimations des moyennes  $\mu_1$  et  $\mu_2$  de l'échantillon d'apprentissage :

```
calculMu <- function (ech) {  
  
  mu = matrix(c(0, 0, 0, 0), 2, 2)  
  tt1 = 0  
  tt2 = 0  
  
  for (i in 1:nrow(ech[1:(nrow(ech)), ])) {  
    if (ech[i,3] == 0) {  
      mu[1] = mu[1] + ech[i,1]  
      mu[2] = mu[2] + ech[i,2]  
      tt1 = tt1 + 1  
  
    } else {  
      mu[3] = mu[3] + ech[i,1]  
      mu[4] = mu[4] + ech[i,2]  
      tt2 = tt2 + 1  
    }  
  }  
  
  mu[1] = mu[1] / tt1  
  mu[2] = mu[2] / tt1  
  mu[3] = mu[3] / tt2  
  mu[4] = mu[4] / tt2  
  
  return (mu)  
}
```

- La fonction **regleEuclidienne** retournant la classe retenue par le classifieur euclidien pour une observation d'un échantillon :

```
regleEuclidienne <- function (x, mu1, mu2) {  
  
  if ((abs(mu2["x"] - x[1])) > (abs(mu1["x"] - x[1]))) {  
    return (0)  
  } else {  
    return (1)  
  }  
}
```

- La fonction **erreurEstimee** qui retourne la probabilité d'erreur estimée sur un échantillon (l'échantillon de test dans notre cas) :

```
erreurEstimee <- function (ech, regle, mu1, mu2) {

  r = 0
  classement = apply(ech, 1, regle, mu1 = mu1, mu2 = mu2)

  for (i in 1:nrow(ech)) {

    if (classement[i] != ech[i, 3]) {
      r = r + 1
    }
  }

  r = r / nrow(ech) * 100
  return (r)
}
```

Le calcul de la probabilité d'erreur en fonction des estimations de moyennes  $\mu_1$  et  $\mu_2$  nous donne les valeurs suivantes :

$(a_1, a_2)$	$\mu_1$	$\mu_2$	Probabilité d'erreur (%)
(1, 1)	(-0.038, -0.005)	(9.852, -0.146)	0
(1, 6)	(0.096, 0.133)	(10.163, -0.361)	1
(1, 9)	(-0.127, 0.041)	(9.489, -0.062)	1.33
(5, 5)	(-0.145, -0.385)	(10.094, 0.048)	0
(10, 10)	(-0.079, 0.160)	(9.798, 0.267)	3

### 3. Probabilité d'erreur moyenne.

Nous avons répété les 2 premières questions 10 fois afin d'estimer la moyenne et la variance des résultats obtenus.

Pour ce faire, nous avons réalisé une fonction nous permettant d'exécuter toutes les étapes rencontrées précédemment et ce 10 fois de suite afin d'obtenir notre probabilité d'erreur moyenne :

```
erreurMoyenne <- function (n, pi, mu1, mu2, sigma1, sigma2, regle, intervalle) {  
  
  result = matrix(c(0, 0, 0, 0), 1, 4)  
  
  for (i in 1:10) {  
  
    # Création de l'échantillon.  
    ech = simul(n, pi, mu1, mu2, sigma1, sigma2)  
  
    # Calcul des estimateurs des moyennes mu1 et mu2  
    esti = calculMu(ech[1:(n * pi), ])  
  
    # Calcul de la probabilité d'une erreur.  
    e = erreurEstimee(ech[(n * pi + 1):n, ], regle, esti[,1], esti[,2]) / 100  
    result[1] = result[1] + e  
    result[2] = e^2 / 10  
  }  
  
  # Moyenne  
  result[1] = result[1] / 10  
  
  # Variance  
  result[2] = result[2] - result[1]^2  
  
  # Écart-type  
  e = 10 / (10 - 1) * result[2]  
  
  # Intervalle de Confiance de niveau 5%  
  result[3] = result[1] - intervalle * e / sqrt(10)  
  result[4] = result[1] + intervalle * e / sqrt(10)  
  
  return (result)  
}
```

Cette fonction nous permet de récupérer les données suivantes :

$(a_1, a_2)$	Moyenne empirique	Variance empirique	Intervalle de Confiance
(1, 1)	0	0	0
(1, 6)	0.009	$-7.067e^{05}$	[0.0086, 0.0086]
(1, 9)	0.025	$-5.139e^{04}$	[0.0250, 0.0249]
(5, 5)	0.013	$-1.733e^{04}$	[0.0133, 0.0133]
(10, 10)	0.065	$-3.737e^{03}$	[0.0647, 0.0646]

Nous pouvons observer, au niveau de l'intervalle de confiance que lorsque nous avons des classes avec une variance élevé, la moyenne d'erreur augmente du fait que les points sont plus dispersés dans le plan que lorsque la variance des classes est plus faible.

## Exercice 2 : Règle de Bayes

On considère un problème de détection de cibles dans lequel la classe  $\mu_1$  correspond aux missiles et la classe  $\mu_2$  correspond aux avions. Chaque cible est décrite par deux variables  $X_1$  et  $X_2$  issues de deux capteurs différents.

Chaque variable suit, dans chaque classe, une loi normale avec les paramètres suivants :

$f_{11}(x_1) \sim \mathcal{N}(-1, 1)$ ,  $f_{21}(x_1) \sim \mathcal{N}(1, 1)$  et  $f_{12}(x_2) = f_{22}(x_2) \sim \mathcal{N}(0, 1)$ . On suppose l'indépendance conditionnelle de  $X_1$  et  $X_2$ . Les densités conditionnelles du vecteur  $\mathbf{X} = (X_1, X_2)^T$  sont donc

$f_1(\mathbf{x}) = f_{11}(x_1)f_{12}(x_2)$  dans la classe  $\omega_1$  et  $f_2(\mathbf{x}) = f_{21}(x_1)f_{22}(x_2)$  dans la classe  $\omega_2$ . On suppose que les distributions sont connues et ne sont donc pas estimées à partir d'un échantillon.

### 1. Montrer que les distributions $f_1$ et $f_2$ sont des distributions normales.

Nous avons donc :  $f_{11}(x_1) \sim \mathcal{N}(-1, 1)$ ,  $f_{21}(x_1) \sim \mathcal{N}(1, 1)$  et  $f_{12}(x_2) = f_{22}(x_2) \sim \mathcal{N}(0, 1)$ .

Vu que  $X_1$  et  $X_2$  sont indépendants, les densités conditionnelles du vecteur  $\mathbf{X} = (X_1, X_2)^T$  sont donc  $f_1(\mathbf{x}) = f_{11}(x_1)f_{12}(x_2)$  et  $f_2(\mathbf{x}) = f_{21}(x_1)f_{22}(x_2)$ .

La fonction de densité d'une variable aléatoire suivant une loi normale est de la forme  $f(x) = \frac{1}{\sigma(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$

Donc  $f_1(\mathbf{x}) = f_{11}(x_1)f_{12}(x_2) = \frac{1}{\sigma_{11}(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}(\frac{x_1-\mu_{11}}{\sigma_{11}})^2} \cdot \frac{1}{\sigma_{12}(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}(\frac{x_2-\mu_{12}}{\sigma_{12}})^2}$

$$f_1(\mathbf{x}) = \frac{1}{2\pi\sigma_{11}\sigma_{12}} e^{-\frac{1}{2\sigma_{11}^2}(x_1-\mu_{11})^2 - \frac{1}{2\sigma_{12}^2}(x_2-\mu_{12})^2}$$

$$f_1(\mathbf{x}) = \frac{1}{2\pi \det(\Sigma)^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)}$$

$$\text{avec } \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \mu_1 = \begin{pmatrix} \mu_{11} \\ \mu_{12} \end{pmatrix}, \Sigma_1 = \begin{pmatrix} \sigma_{11}^2 & 0 \\ 0 & \sigma_{12}^2 \end{pmatrix} \text{ et } \Sigma_1^{-1} = \begin{pmatrix} \frac{1}{\sigma_{11}^2} & 0 \\ 0 & \frac{1}{\sigma_{12}^2} \end{pmatrix}$$

$$\text{Nous avons donc } f_1(\mathbf{x}) \sim \mathcal{N}(\mu_1, \Sigma_1) \text{ avec } \mu_1 = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \text{ et } \Sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Nous appliquons le même raisonnement pour  $f_2$  et nous obtenons un  $f_2(\mathbf{x}) \sim \mathcal{N}(\mu_2, \Sigma_2)$

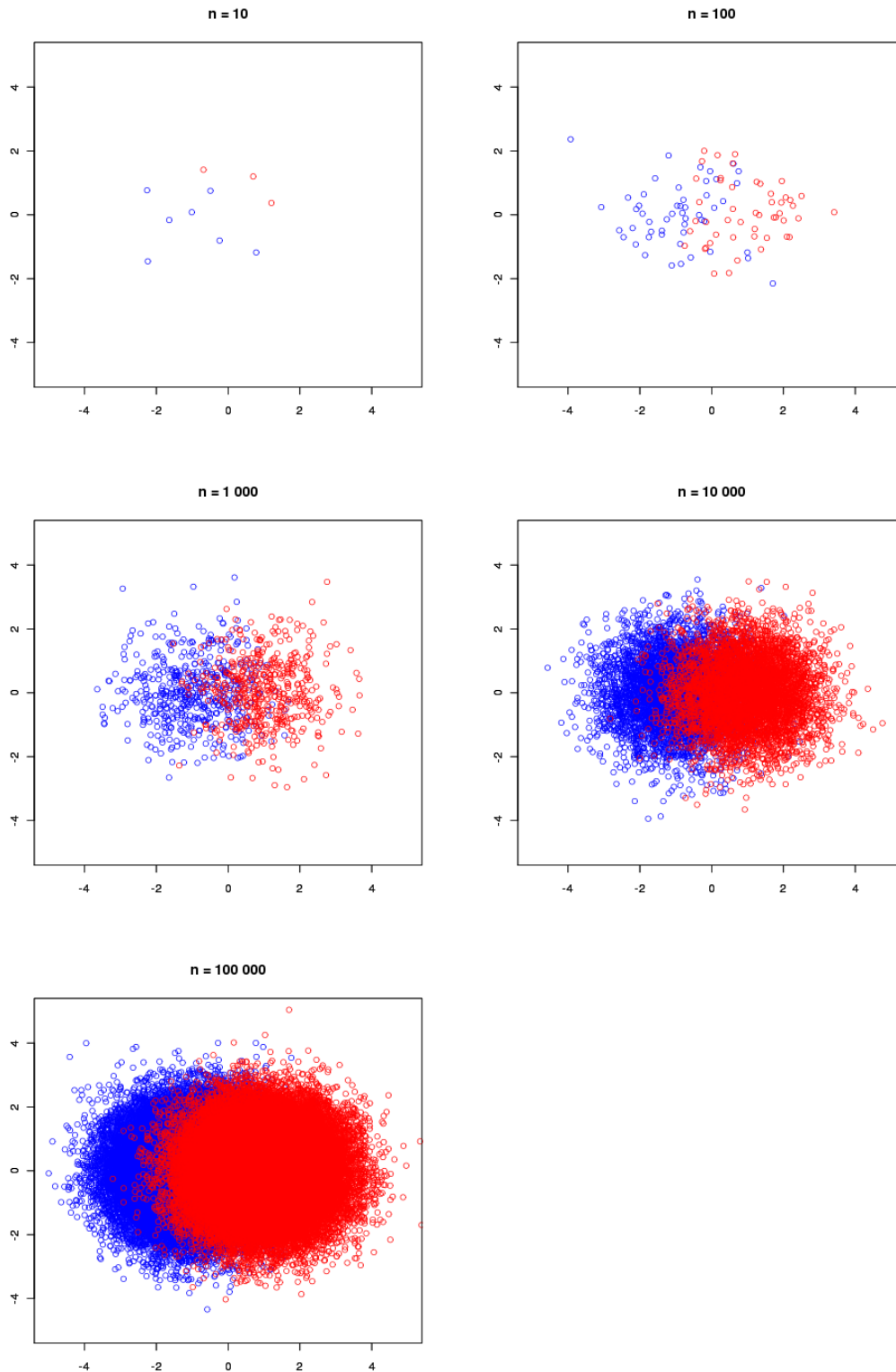
$$\text{avec } \mu_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ et } \Sigma_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

## 2. Générer un échantillon de $n$ réalisations issues des deux classes $\omega_1$ et $\omega_2$ .

Pour générer un échantillon de  $n$  réalisations issues des deux classes  $\omega_1$  et  $\omega_2$  en proportions égales ( $\pi_1 = \pi_2 = 0.5$ ), nous avons utilisé la fonction **simul** écrite précédemment.

Pour chacun des échantillons, nous avons ensuite déterminé les estimations des différents paramètres de  $f_1$  et  $f_2$ .

Pour les valeurs de  $n$  suivantes : 10, 100, 1000, 10 000, 100 000, nous obtenons donc les graphiques suivants :



Nous pouvons observer les estimations des différents paramètres de  $f_1$  et de  $f_2$  dans le tableau suivant :

Échantillon	$\bar{\mu}_1$	$\bar{\Sigma}_1$	$\bar{\mu}_2$	$\bar{\Sigma}_2$
10	$\begin{pmatrix} -0.570 \\ -0.584 \end{pmatrix}$	$\begin{pmatrix} 0.762 & 0.302 \\ 0.302 & 0.354 \end{pmatrix}$	$\begin{pmatrix} 1.883 \\ 0.327 \end{pmatrix}$	$\begin{pmatrix} 1.954 & -0.092 \\ -0.091 & 0.700 \end{pmatrix}$
100	$\begin{pmatrix} -1.164 \\ 0.114 \end{pmatrix}$	$\begin{pmatrix} 0.802 & 0.156 \\ 0.156 & 1.121 \end{pmatrix}$	$\begin{pmatrix} 0.575 \\ 0.094 \end{pmatrix}$	$\begin{pmatrix} 1.359 & 0.200 \\ 0.200 & 0.937 \end{pmatrix}$
1 000	$\begin{pmatrix} -1.0854 \\ 0.006 \end{pmatrix}$	$\begin{pmatrix} 1.011 & -0.053 \\ -0.053 & 0.830 \end{pmatrix}$	$\begin{pmatrix} 0.961 \\ 0.051 \end{pmatrix}$	$\begin{pmatrix} 1.099 & 0.113 \\ 0.113 & 0.946 \end{pmatrix}$
10 000	$\begin{pmatrix} -1.003 \\ -0.009 \end{pmatrix}$	$\begin{pmatrix} 0.993 & 0.027 \\ 0.027 & 0.982 \end{pmatrix}$	$\begin{pmatrix} 0.987 \\ 0.012 \end{pmatrix}$	$\begin{pmatrix} 0.974 & 0.010 \\ 0.010 & 0.967 \end{pmatrix}$
100 000	$\begin{pmatrix} -1.002 \\ 0.001 \end{pmatrix}$	$\begin{pmatrix} 0.995 & 0.000 \\ 0.000 & 0.997 \end{pmatrix}$	$\begin{pmatrix} 1.001 \\ 0.003 \end{pmatrix}$	$\begin{pmatrix} 1.011 & 0.000 \\ 0.000 & 1.00 \end{pmatrix}$

Nous avons pu observer que plus  $n$  était grand, plus on se rapprochait de la valeur théorique.

### 3. Montrer que les courbes d'iso-densité sont des cercles.

Calculer les courbes d'iso-densité revient à résoudre l'équation suivante :

$f_{w_1}(x) = K$  où  $K$  est une constante.

Dans le cas de  $f_1(x) = K_1$

$$\frac{1}{2}(x - \mu)^T(x - \mu) = -\ln(2\pi K)$$

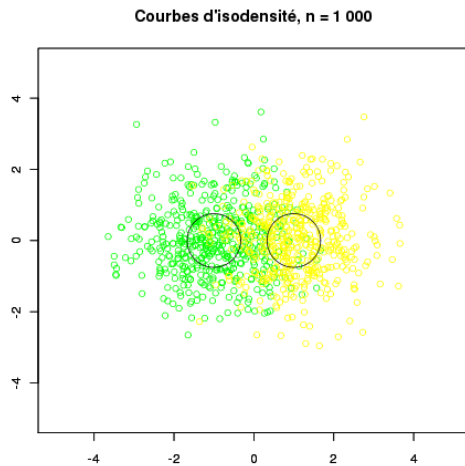
$$\Leftrightarrow (x_1 - \mu_1)^2 + (x_2 - \mu_2)^2 = -2\ln(2\pi K_1)$$

$$\Leftrightarrow (x_1 + 1)^2 + x_2^2 = -2\ln(2\pi K_1)$$

Nous avons donc finalement un cercle de centre  $\mu_1 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$  et de rayon  $\sqrt{-2\ln(2\pi K_1)}$ .

Dans le cas de  $f_2(x) = K_2$ , nous obtenons un cercle de centre  $\mu_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$  et de rayon  $\sqrt{-2\ln(2\pi K_2)}$ .

Nous avons donc obtenu un graphique de la forme suivante :



Pour ce graphique, nous avons choisi de prendre  $K_1 = K_2 = 0.2$ . Nous pouvons aussi confirmer visuellement ce que la partie calcul à montrer, à savoir que ces courbes d'iso-densité sont des cercles de rayon = 0.671 dans notre cas.



#### 4. Règle de Bayes.

Soient  $\pi_1$  et  $\pi_2$  les probabilités à priori des deux classes, et  $c_{lk}$  le coût associé au choix de l'action  $a_l$  lorsque la vraie classe est  $\omega_k$ . On suppose  $c_{11} = c_{22} = 0$ .

(a) Donner l'expression de la règle de Bayes  $\delta^*$  pour ce problème.

La règle de Bayes s'écrit de la manière suivante :

$$\delta^*(x) = \begin{cases} a_1 & \text{si } \frac{f_1(x)}{f_2(x)} > \frac{c_{12}\pi_2}{c_{21}\pi_1} \\ a_2 & \text{sinon.} \end{cases}$$

Les coûts sont définis ainsi :

	$\omega_1$	$\omega_2$
$a_1$	$c_{11}$	$c_{12}$
$a_2$	$c_{21}$	$c_{22}$

(b) Tracer les frontières de décision correspondantes dans le plan  $(X_1, X_2)$ .

- Dans le cas où  $c_{12} = c_{21} = 1$ ,  $\pi_1 = \pi_2$  :

Nous pouvons écrire la règle de Bayes pour ce cas donc nous avons  $\delta^* = a_1 \Leftrightarrow f_1(x) > f_2(x)$ .

Ici, nous trouvons que  $f_1(x) = f_2(x) \Leftrightarrow (x+1)^2 = (x-1)^2 \Leftrightarrow x = 0$

La frontière de décision est la droite d'équation  $x = 0$  dans ce cas de figure.

- Dans le cas où  $c_{12} = 10$ ,  $c_{21} = 1$ ,  $\pi_1 = \pi_2$  :

Nous pouvons écrire la règle de Bayes pour ce cas donc nous avons  $\delta^* = a_1 \Leftrightarrow f_1(x) > 10.f_2(x)$ .

$$\begin{aligned} \frac{f_1(x)}{f_2(x)} &< 10 \\ \Leftrightarrow e^{-2(x)} &< 10 \\ \Leftrightarrow x &= -\frac{\ln(10)}{2} \end{aligned}$$

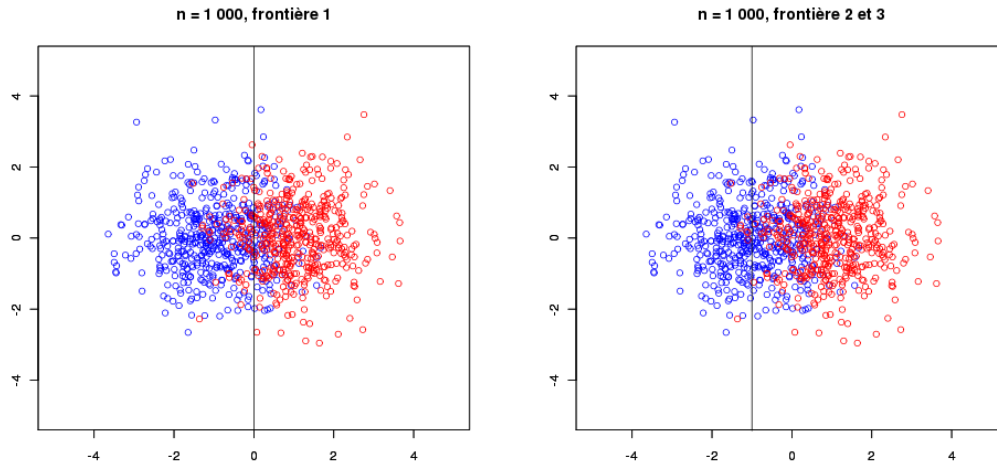
La frontière de décision, pour ce cas, la droite d'équation  $x = -\frac{\ln(10)}{2}$ , soit approximativement la droite d'équation  $x = -1.15$

- Dans le cas où  $c_{12} = c_{21} = 1$ ,  $\pi_2 = 10\pi_1$  :

Nous pouvons écrire la règle de Bayes pour ce cas donc nous avons  $\delta^* = a_1 \Leftrightarrow f_1(x) > 10.f_2(x)$ .

$$\begin{aligned} \frac{f_1(x)}{f_2(x)} &< 10 \\ \Leftrightarrow e^{-2(x)} &< 10 \\ \Leftrightarrow x &= -\frac{\ln(10)}{2} \end{aligned}$$

La frontière de décision, pour ce cas, la droite d'équation  $x = -\frac{\ln(10)}{2}$ , soit approximativement la droite d'équation  $x = -1.15$  comme pour le cas précédent.



À travers ces deux graphiques, nous pouvons dire que dans le premier cas, les erreurs de la première et de la deuxième classe sont du même ordre. Aucune des deux n'est plus pénalisée par rapport à cet frontière.

Pour le second et troisième cas, le coût d'erreur de la première classe, est plus grand que celui de la seconde classe. Par conséquent, la règle a tendance à minimiser les erreurs de la seconde classe.

### (c) Donner une estimation des risques $\alpha$ et $\beta$ .

Pour évaluer une estimation des risques  $\alpha = \mathbb{P}(\delta^*(\mathbf{X}) = a_2 | \omega_1)$  et  $\beta = \mathbb{P}(\delta^*(\mathbf{X}) = a_1 | \omega_2)$ , nous utilisons la fonction **simul** écrite en début de tp.

Ces risques sont définis par  $\alpha = \mathcal{P}(\delta^*(x) = a_2 | \omega_1)$  et  $\beta = \mathcal{P}(\delta^*(x) = a_1 | \omega_2)$ .

Nous obtenons ainsi les données suivantes :

	$\alpha$	$\beta$
$c_{12} = c_{21} = 1, \pi_1 = \pi_2$	0.16	0.17
$c_{12} = 10, c_{21} = 1, \pi_1 = \pi_2$	0.53	0.01
$c_{12} = c_{21} = 1, \pi_2 = 10\pi_1$	0.70	0.03

Dans le premier cas, nous pouvons observer que les estimateurs des risques  $\alpha$  et  $\beta$  sont très proches. Dans le deuxième et le troisième cas,  $\beta$  est plus petit que  $\alpha$ .

## Conclusion

Durant ce TP, nous avons eu la possibilité d'utiliser deux méthodes de classification supervisée : le classifieur euclidien, l'un des plus simples pour se baser un ensemble d'apprentissage pour déterminer le centre de classes. Cette simplicité vient de paire avec une limitation du coté de la gestion des distributions très distribuées dans le plan.

puis la règle de Bayes qui permet de contrôler les risques ce qui ajoute une gestion des erreurs plus poussée. Avec cette méthode, nous avons pu voir que nous pouvions réduire le risque d'erreur au lieu du nombre d'erreurs, nous permettant ainsi de privilégier ou non un certain type d'erreur par rapport à d'autres.