Solution du problème concernant l'A. F. D. sur dix points

N.B. *Les dix points sont les mêmes que pour le problème où l'ACP est appliquée. Les valeurs numériques des réponses 1. et 2. sont donc inchangées.*

1. $g = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{10}\sum_{i=1}^{10} x_i^1 \\ \frac{1}{10}\sum_{i=1}^{10} x_i^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{10}(0+0+1\ldots+1) \\ \frac{1}{10}(0+0+1\ldots+2) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

   $T' = \begin{pmatrix} -1 & -1 & 0 & 0 & 1 & 1 & 0 & -1 & 1 & 0 \\ -1 & -1 & 0 & 0 & 1 & 1 & -1 & 0 & 0 & 1 \end{pmatrix}$

2. $V = T' D_p T = \frac{1}{10} T'T = \begin{pmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{pmatrix} = \frac{1}{5}\begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}$

3. D'où $V^{-1} = 5 \times \frac{1}{9-4}\begin{pmatrix} 3 & -2 \\ -2 & 3 \end{pmatrix} = \begin{pmatrix} 3 & -2 \\ -2 & 3 \end{pmatrix}$. On vérifie que l'on a bien :

   $V^{-1}V = \begin{pmatrix} 3 & -2 \\ -2 & 3 \end{pmatrix} \frac{1}{5}\begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

4. $g_A = \begin{pmatrix} g_{A,1} \\ g_{A,2} \end{pmatrix}$ avec $\begin{cases} g_{A,1} = (0+0+2+2+1+2)/6 = 7/6 \\ g_{A,2} = (0+0+2+2+0+1)/6 = 5/6 \end{cases}$

   $g_B = \begin{pmatrix} g_{B,1} \\ g_{B,2} \end{pmatrix}$ avec $\begin{cases} g_{B,1} = (1+1+0+1)/4 = 3/4 \\ g_{A,2} = (1+1+1+2)/6 = 5/4 \end{cases}$

5. Soit H l'hyperplan séparateur de Fisher, c'est-à-dire l'hyperplan médiateur du segment $[g_A, g_B]$ selon la métrique $M = V^{-1}$:

   $$H = \{x \mid \left(x - \frac{g_A + g_B}{2}\right)' V^{-1}(g_A - g_B) = 0\}.$$

   $g_A - g_B = \begin{pmatrix} 7/6 - 3/4 \\ 5/6 - 5/4 \end{pmatrix} = \begin{pmatrix} (14-9)/12 \\ (10-15)/12 \end{pmatrix} = \frac{5}{12}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$

   $\frac{1}{2}(g_A + g_B) = \frac{1}{2}\begin{pmatrix} 7/6 + 3/4 \\ 5/6 + 5/4 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 23/12 \\ 25/12 \end{pmatrix} = \begin{pmatrix} 23/24 \\ 25/24 \end{pmatrix}$.

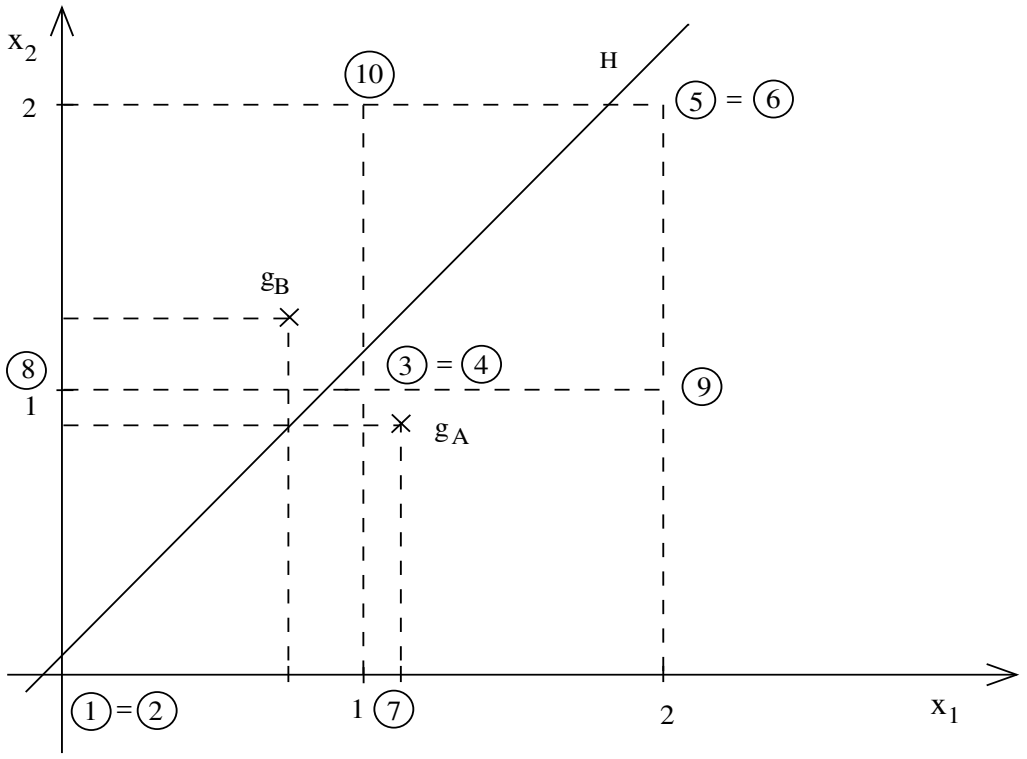
   Or $V^{-1}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 3 & -2 \\ -2 & 3 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = 5\begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

   En posant $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, on en déduit $H = \{x \mid x_1 - 23/24 - x_2 + 25/24 = 0\}$. Finalement :

   $$H = \left\{x \mid x_2 - x_1 = 1/12\right\}$$

6. Affectation des 10 points.

A) *Méthode graphique* : les 10 points sont représentés dans le plan euclidien par leurs coordonnées qui sont les valeurs prises par les variables x^1 et x^2. On représente également les centres de gravité g_A et g_B, ainsi que l'hyperplan H qui sépare le plan en deux régions d'affectation $\mathcal{R}_A$ qui contient g_A et $\mathcal{R}_B$ qui contient g_B : voir la figure ci-dessous.



N.B. *Approximation utilisée :* $g_A \approx \begin{pmatrix} 1.16 & 0.83 \end{pmatrix}'$.

Comme g_B est au dessus de H, et g_A au dessous, tout point au dessus de H est affecté à C_B, et tout point au dessous à C_A. Donc les points 8 et 10 sont affectés à C_B et tous les autres à $C - A$. Les seuls points mal classés sont les points 3 et 4 : ils appartiennent à C_B mais sont affectés à tort à C_A.

B) *Méthode algébrique* : On rappelle qu'un point x est affecté à la classe C_A si et seulement si $\left(x - \frac{g_A + g_B}{2}\right)' V^{-1}(g_A - g_B)$ est positif. D'après les calculs précédents, cette condition est réalisée ssi $x_1 - x_2 + \frac{1}{12}$ est positif. Donc tous les points sont affectés à C_A sauf les points 8 et 10 qui sont affectés à C_B.

7. Tableau de classement

(Affectation)

(Appartenance)		A	B
	A	6	0
	B	2	2

8. Taux de bien classés dans $C_A = 100\%$ (6/6)
   Taux de bien classés dans $C_B = 50\%$ (2/4)
   Taux global de bien classés 80% (6 + 2/10)

9. Ici $q = 2$, donc B = matrice variance interclasses = $m_A m_B (g_A - g_B)(g_A - g_B)'$.

   Or $m_A = 0.6, m_B = 0.4$ et $g_A - g_B = \frac{5}{12}\begin{pmatrix}1\\-1\end{pmatrix}$, d'où :

$$B = \left(\frac{5}{12}\right)^2 \times \frac{6}{10} \times \frac{4}{10} \times \begin{pmatrix}1\\-1\end{pmatrix}\begin{pmatrix}1 & -1\end{pmatrix} = \frac{1}{12} \times \frac{2}{4} \times \begin{pmatrix}1 & -1\\-1 & 1\end{pmatrix} = \frac{1}{24}\begin{pmatrix}1 & -1\\-1 & 1\end{pmatrix}.$$

   Pour calculer B, on aurait pu aussi utiliser sa définition dans le cas général, i.e. B est la matrice variance du nuage $\mathcal{M}_G = \{(g_A; m_A = 0.6), (g_B; m_B = 0.4)\}$. Le calcul est alors un peu moins simple. Plus précisément, le tableau des données associé à ce nuage, noté G, est défini par $G' = \begin{pmatrix}g_A - g & g_B - g\end{pmatrix}$ avec :

$$g_A - g = \begin{pmatrix}7/6\\5/6\end{pmatrix} - \begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}1/6\\-1/6\end{pmatrix} \quad \text{et} \quad g_B - g = \begin{pmatrix}3/4\\5/4\end{pmatrix} - \begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}-1/4\\1/4\end{pmatrix}.$$

   D'où $B = G' D_m G = \begin{pmatrix}1/6 & -1/4\\-1/6 & 1/4\end{pmatrix}\begin{pmatrix}3/5 & 0\\0 & 2/5\end{pmatrix}\begin{pmatrix}1/6 & -1/6\\-1/4 & 1/4\end{pmatrix} = \frac{1}{24}\begin{pmatrix}1 & -1\\-1 & 1\end{pmatrix}$.

10. On a $W = m_A V_A + m_B V_B$ ou, ce qui est plus simple ici, $W = V - B$. D'où :

$$W = \frac{1}{5}\begin{pmatrix}3 & 2\\2 & 3\end{pmatrix} - \frac{1}{24}\begin{pmatrix}1 & -1\\-1 & 1\end{pmatrix} = \frac{1}{120}\begin{pmatrix}67 & 53\\53 & 67\end{pmatrix}.$$

11. Soit u le vecteur axial factoriel discriminant qui, au signe près, est unique ici car $q = 2$. On a :

$$u = \frac{g_A - g_B}{\|g_A - g_B\|_{V^{-1}}},$$

    avec $g_A - g_B = \frac{5}{12}\begin{pmatrix}1\\-1\end{pmatrix}$. D'où :

$$\|g_A - g_B\|^2_{V^{-1}} = \left(\frac{5}{12}\right)^2 \begin{pmatrix}1 & -1\end{pmatrix} V^{-1} \begin{pmatrix}1\\-1\end{pmatrix} = \left(\frac{5}{12}\right)^2 \begin{pmatrix}1 & -1\end{pmatrix}\begin{pmatrix}5\\-5\end{pmatrix} = \left(\frac{5}{12}\right)^2 \times 10,$$

d'où $u = \frac{1}{\sqrt{10}}\begin{pmatrix}1\\-1\end{pmatrix}$.

NB : on peut changer l'orientation de u.

Le facteur discriminant b est donné par :

$$b = V^{-1}u = \frac{1}{\sqrt{10}}V^{-1}\begin{pmatrix}1\\-1\end{pmatrix} = \sqrt{\frac{5}{2}}\begin{pmatrix}1\\-1\end{pmatrix}.$$

12. Comme $q = 2$, la valeur propre discriminante est égale à $\lambda = m_A m_B \|g_A - g_B\|^2_{V^{-1}}$. D'après ce qui précède :

$$\lambda = \frac{6}{10} \times \frac{4}{10} \times \left(\frac{5}{12}\right)^2 \times 10 = \frac{3 \times 4 \times 5}{12^2} = \frac{5}{12}$$

On aurait pu aussi calculer BV^{-1}, puis sa trace qui est égale à λ, puisque dans le cas $q = 2$, il n'y a qu'une seule valeur propre non triviale.

$$BV^{-1} = \frac{1}{24}\begin{pmatrix}1 & -1\\-1 & 1\end{pmatrix}\begin{pmatrix}3 & 2\\2 & 3\end{pmatrix} = \frac{5}{24}\begin{pmatrix}1 & -1\\-1 & 1\end{pmatrix}.$$

D'où $\lambda = \mathrm{tr}(BV^{-1}) = \frac{5}{24} + \frac{5}{24} = \frac{5}{12}$.

COMPLÉMENT

Solution à l'aide du logiciel R

Données : lecture et affichage des données.

Première méthode : lire les données sur un fichier en utilisant la commande read.table qui crée un data.frame, noté "don".

```
> don <- read.table("don-AFD.txt", header=T)
> plot(X2~X1,data=don,col=Y)
#
#  Type des données
#
> mode(don)
> is.data.frame(don)
```

Seconde méthode : on construit directement une matrice contenant les données, noté "don1". On doit alors introduire les noms des variables X1, X2 et Y.

```
> don1 <- matrix(c(0,0,1,1,2,2,1,0,2,1,0,0,1,1,2,2,0,1,1,2,"A",
"A","B","B","A","A","A","B","A","B"),ncol=3,byrow=F)
> don1 <- as.data.frame(don1)
> nomligne <- c("i1","i2","i3","i4","i5","i6","i7","i8","i9","i10")
> nomcol <- c("X1","X2","Y")
> dimnames(don1) <- list(nomligne,nomcol)
> plot(X2 ~ X1,data=don1,col=Y)
```

1. Centre de gravité g

   ```
   > g <-mean(don[,1:2])
   ```

   Tableau centré T :

   ```
   > T <- don[,1:2] - g
   ```

   ou bien :

   ```
   > matg <- matrix(rep.int(g,10), ncol=2)
   > T <- don[,1:2] - matg
   ```

2. Matrice variance : si l'on utilise la commande "var", il faut tenir compte du fait que chaque matrice variance calculée par R est corrigée, i.e. R divise par $n-1$ au lieu de n, pour obtenir une estimation sans biais de cette variance.

   ```
   > is.matrix(T)
   > T <- as.matrix(T)
   > is.matrix(T)
   > V <- t(T) %*% T / 10
   ```

   ou

   ```
   > V <- var(don[,1:2])*9/10
   ```

3. Matrice inverse, notée $V1$

   ```
   > V1 <- solve(V)
   ```

4. Centres de gravité g_A et g_B

   On commence par extraire les sous-tableaux de T, notés TA et TB, associés respectivement aux classes C_A et C_B.

   ```
   > rA <- c(1,2,5,6,7,9)
   > rB <- c(3,4,8,10)
   > TA <- don[rA,]
   > TB <- don[rB,]
   ```

   ```
   > gA <-mean(TA[,1:2])
   > gB <-mean(TB[,1:2])
   ```

5. Equation de l'hyperplan séparateur de Fisher
   On rappelle que cette équation s'écrit :

$$\left(x - \frac{g_A + g_B}{2}\right)' V^{-1}(g_A - g_B) = 0$$

   Soit m_{AB} le milieu de $[g_A, g_B]$, noté m_AB en R, et C le vecteur $V^{-1}(g_A - g_B)$:

```
> m_AB <- (gA + gB)/2
> C <- V1 %*% (gA - gB)
```

   Comme le vecteur C est proportionnel à $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$, on en déduit que l'équation s'écrit : $x_1 - m_{AB}[1] - x_2 + m_{AB}[2] = 0$, soit encore $x_2 - x_1 = m_{AB}[2] - m_{AB}[1]$. Or :

```
> m_AB[2]  - m_AB[1]
        X2
0.08333333
```

   On retrouve ainsi le résultat précédant obtenu sans R, puisque $0.08333333 \approx 1/12$.

6. Affectation des 10 points
   On utilise la commande "lda" du package MASS qu'il faut d'abord appeler. La commande "lda" construit le modèle, et "predict" donne les affectations (composant $class), les probabilités a posteriori ($posterior) et les scores ($x).

```
> library(MASS)
> modele <- lda(Y~., don)
> predict(modele)$class
 [1] A A A A A A A B A B
Levels: A B
```

7. Tableau de classement

```
> table(don$Y,predict(modele)$class)
    A B
  A 6 0
  B 2 2
```

   Cette estimation du tableau de classement est biaisée par excès d'optimisme, puisqu'elle est établie sur les données qui ont servi à construire la règle de décision. Si l'on utilise la validation croisée, on obtient une estimation du tableau avec un pourcentage de bien classés qui chute de 80% à 40% : voir ci-dessous.

```
> prev <- lda(Y~., data=don, CV=TRUE)$class
> prev
 [1] B B A A B B A B A B
Levels: A B
```

```
> table(don$Y,prev)
   prev
    A B
  A 2 4
  B 2 2
```

8. Taux de bien classés

   TBA : taux de bien classés dans la classe C_A
   TBB : taux de bien classés dans la classe C_B
   TBG : taux global de bien classés

```
> tab <- table(don$Y,predict(modele)$class)
> TBA <- tab[1,1]/sum(tab[1,])
> tab[1,]
A B
6 0
> TBA
[1] 1
> TBB <- tab[2,2]/sum(tab[2,])
> TBB
[1] 0.5
> TBG <- sum(diag(tab))/sum(tab)
> TBG
[1] 0.8
```

9. Matrice variance interclasses B

   On utilise la formule $B = m_A\, m_B\, (g_A - g_B)\, (g_A - g_B)'$.

```
> mA <- length(rA)/nrow(don)
> mB <- length(rB)/nrow(don)
> B <- mA * mB * (gA - gB) %*% t(gA - gB)
```

10. Matrice variance intraclasses W

    On utilise la formule $W = V - B$.

```
> W <- V - B
```

11. Vecteur axial factoriel

    Ce vecteur, noté u, est colinéaire à $g_A - g_B$, et est normé à 1 pour la métrique V^{-1}.

```
> norme <- sqrt(t(gA - gB) %*% V1 %*% (gA - gB))
> u <- (gA - gB)/norme
```

    Facteur discriminant $b = V^{-1}u$.

```
> b <- V1 %*% u
```

12. Pouvoir discriminant ou valeur propre discriminante
Comme $q = 2$, la valeur propre λ peut se calculer de plusieurs façons :
a) $\lambda = m_A m_B \|g_A - g_B\|^2_{V^{-1}}$

```
> lambda <-  mA * mB * norme^2
```

b) λ est aussi la trace de BV^{-1}

```
> lambda <-  sum(diag(B %*% V1))
```

c) Par ailleurs, λ est aussi l'unique valeur propre non nulle de BV^{-1}

```
> res <- eigen(B %*% V1)
> res$values
```

Remarque concernant la question 11.

La commande "lda" fournit directement le facteur discriminant, résultat qui est donné par le quatrième composant de la liste générée par "lda", ici la liste "modele" (cf. question 6.) :

```
> modele[4]
$scaling
        LD1
X1 -1.85164
X2  1.85164
```

Néanmoins, par rapport à la valeur de b précédemment trouvée, on observe que ce vecteur diffère de b d'une constante multiplicative : cela est dû au fait que R utilise la métrique W^{-1} au lieu de la métrique V^{-1}, et plus précisément la métrique $(W^\star)^{-1}$ au lieu de la métrique V^{-1}, où $W^\star$ est l'estimation empirique sans biais de W (cf. cours) :

$$W^\star = \frac{n}{n-2} W.$$

En reprenant les calculs du polycopié de cours pour tenir compte de ce facteur correctif $\frac{n}{n-2}$, on obtient facilement que :

$$c = \frac{b}{\sqrt{\frac{n}{n-2}(1-\lambda)}},$$

où c désigne le facteur discriminant calculé avec la métrique $(W^\star)^{-1}$.
A l'aide de R, on observe que cette relation est bien vérifiée :

```
> n <- 10
> c <- b / sqrt(n*(1-lambda)/(n-2))
> c
       [,1]
X1  1.85164
X2 -1.85164
```