

## Indices de réfraction d'éclats de verre

Anne B. Dufour &amp; Jean R. Lobry

21 août 2008

**Table des matières**

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Caractérisation des fragments de verre</b>	<b>2</b>
2.1	Un jeu de données classique . . . . .	2
2.2	L'indice de réfraction . . . . .	2
2.3	Composition chimique . . . . .	3
2.4	L'indice de réfraction et la composition chimique . . . . .	6
2.5	Les types de verre . . . . .	8
<b>3</b>	<b>Hétérogénéité de l'indice de réfraction</b>	<b>12</b>
3.1	Le jeu de données . . . . .	12
3.2	Comparaison avec fg1 . . . . .	12
3.3	Comparaison avec d'autres écart-types . . . . .	13
3.4	Rapprochement des échantillons Q et B . . . . .	14
3.5	Rapprochement des échantillons Q et A . . . . .	15

**1 Introduction**

Qui dit statistiques forensiques pense immédiatement à l'analyse des traces ADN parce que cette activité est maintenant dominante. Mais il ne faut pas oublier que des traces non biologique sont également présentes sur des scènes de crime. James M. Curran rapporte [4] que pour la période 2001-2002 les preuves basées sur l'analyse d'éclats de verre ont représenté près du quart des cas de preuves non-biologiques en Nouvelle-Zélande<sup>1</sup>.

Toujours d'après James M. Curran [4], les preuves liées au verre interviennent principalement dans des cas où une fenêtre a été brisée, mais pas seulement : par exemple dans un délit de fuite on peut retrouver des éclats d'ampoules des phares, dans une agression des éclats de bouteille en verre ou bien de lunettes, dans une dispute domestique des éclats de récipients en verre, dans un accident de laboratoire des fragments de verrie.

<sup>1</sup>En France, cette pratique est moins répandue, il n'y a pas raison particulière à cette différence autre que culturelle (Laurent Pene, INPS, communication personnelle)

Si une fenêtre est brisée (vol à la roulotte, cambriolage) des éclats de verre, typiquement plus petits qu'un grain de sable, peuvent être transférés sur un suspect. La question est donc de faire le rapprochement entre des éclats de verre trouvés sur une scène de crime et ceux trouvés sur un suspect.

## 2 Caractérisation des fragments de verre

### 2.1 Un jeu de données classique

On dans  $\mathbb{R}$  en standard le jeu données `fgl` sur des éclats de verre dans le paquet `MASS` [11].

```
library(MASS)
data(fgl)
summary(fgl)
```

RI	Na	Mg	Al	Si
Min. : -6.8500	Min. : 10.73	Min. : 0.0000	Min. : 0.290	Min. : 69.81
1st Qu.: -1.4775	1st Qu.: 12.91	1st Qu.: 2.115	1st Qu.: 1.190	1st Qu.: 72.28
Median : -0.3200	Median : 13.30	Median : 3.480	Median : 1.360	Median : 72.79
Mean : 0.3654	Mean : 13.41	Mean : 2.685	Mean : 1.445	Mean : 72.65
3rd Qu.: 1.1575	3rd Qu.: 13.82	3rd Qu.: 3.600	3rd Qu.: 1.630	3rd Qu.: 73.09
Max. : 15.9300	Max. : 17.38	Max. : 4.490	Max. : 3.500	Max. : 75.41

K	Ca	Ba	Fe	type
Min. : 0.0000	Min. : 5.430	Min. : 0.0000	Min. : 0.00000	WinF : 70
1st Qu.: 0.1225	1st Qu.: 8.240	1st Qu.: 0.0000	1st Qu.: 0.00000	WinNF: 76
Median : 0.5550	Median : 8.600	Median : 0.0000	Median : 0.00000	Veh : 17
Mean : 0.4971	Mean : 8.957	Mean : 0.1750	Mean : 0.05701	Con : 13
3rd Qu.: 0.6100	3rd Qu.: 9.172	3rd Qu.: 0.0000	3rd Qu.: 0.10000	Tabl : 9
Max. : 6.2100	Max. : 16.190	Max. : 3.1500	Max. : 0.51000	Head : 29

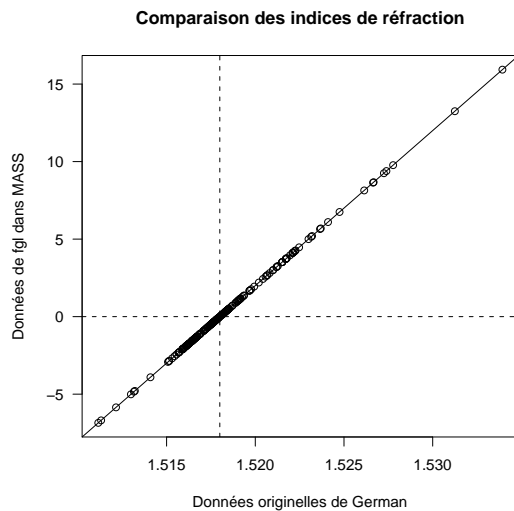
Ce jeu de données est un classique, on le retrouve dans les paquets  $\mathbb{R}$  `ml-bench` [6], `FKBL` [1], `kknn` [9], et `mda` [7]. Le jeu de données original a été créé en 1987 par B. German dans le dépôt de l'UCI [2] (<http://archive.ics.uci.edu/ml/datasets/Glass+Identification>), le donateur est Vina Spiehler. C'est un jeu de données qui a été utilisé dans de nombreuses publications. On peut essayer de l'importer directement dans  $\mathbb{R}$  pour comparer avec `fgl`.

```
fglb <- read.table("http://archive.ics.uci.edu/ml/machine-learning-databases/glass/glass.data",
  sep = ",", row.names = 1)
colnames(fglb) <- colnames(fgl)
fglb$type <- factor(fglb$type, labels = levels(fgl$type))
```

### 2.2 L'indice de réfraction

Comparons les indices de réfraction :

```
plot(fglb$RI, fgl$RI, main = "Comparaison des indices de réfraction",
  xlab = "Données originelles de German", ylab = "Données de fgl dans MASS",
  las = 1)
abline(h = 0, lty = 2)
abline(v = 1.518, lty = 2)
abline(-1518, 1000)
```



Par rapport aux données originelle, les valeurs de indices de réfraction ont donc été centrées à 1.518 et multipliées par 1000.

```
all.equal(fgl$RI/1000 + 1.518, fglb$RI)
[1] TRUE
```

L'intérêt est de faciliter la lecture. Dans les données originelles les indices de réfraction sont donnés avec 6 chiffres significatifs mais les deux premiers ne sont pas informatifs :

```
range(fglb$RI)
[1] 1.51115 1.53393
```

On utilisera donc par la suite les données transformées.

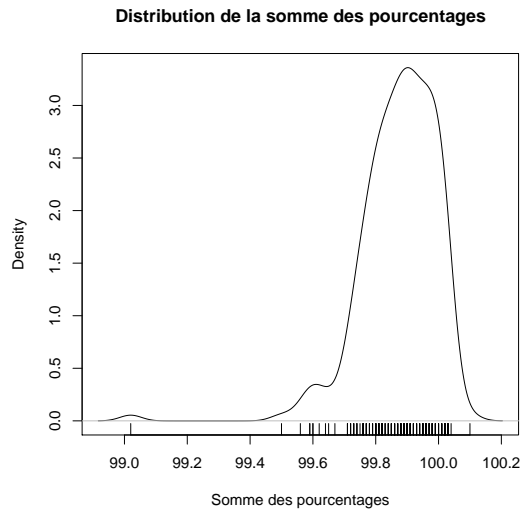
### 2.3 Composition chimique

Les 7 colonnes suivantes donnent la composition chimique des éclats de verre, elles sont identiques entre les deux jeux de données :

```
all.equal(fgl[, 2:8], fglb[, 2:8])
[1] TRUE
```

Les compositions sont exprimées en pourcentage, la somme des colonnes devrait donc faire 100, vérifions :

```
total <- rowSums(fgl[, 2:9])
plot(density(total), main = "Distribution de la somme des pourcentages",
     xlab = "Somme des pourcentages")
rug(total)
```

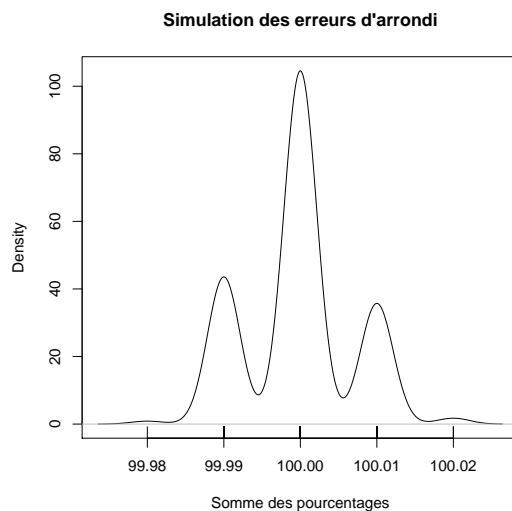


Ce n'est pas le cas. Les valeurs sont discrètes avec un pas de 0.01, ce qui est normal puisque les pourcentages sont donnés avec deux chiffres après la virgule. Mais on trouve des individus à plus de 100 % et un près de 99 % :

```
fgl[which.min(total), ]
      RI   Na  Mg  Al   Si   K   Ca Ba Fe type
150 -1.57 12.16 3.52 1.35 72.89 0.57 8.53 0 0 Veh
```

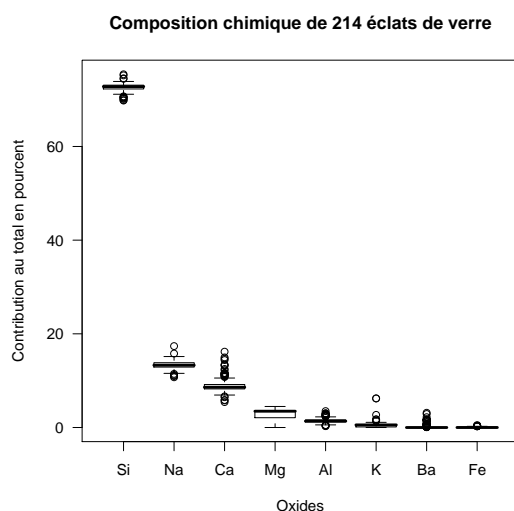
Faisons une petite expérience pour voir ce à quoi on s'attendrait avec les erreurs dues aux arrondis.

```
experience <- function() {
  x <- apply(fgl[, 2:9], 2, function(x) x[sample(1:nrow(fgl),
    1)])
  x <- 100 * x/sum(x)
  x <- round(x, 2)
  return(sum(x))
}
res <- replicate(nrow(fgl), experience())
plot(density(res), main = "Simulation des erreurs d'arrondi", xlab = "Somme des pourcentages")
rug(res)
```



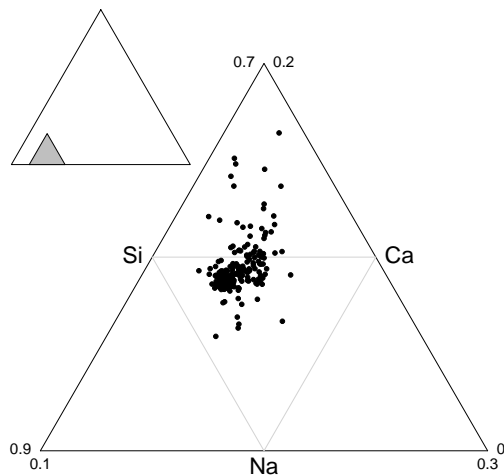
On a une distribution centrée autour de 100 % et beaucoup plus resserrée que celle qui est observée. Il y a donc un sens à ce que la somme des pourcentage ne fasse pas exactement 100 %, peut-être des éléments sous forme de traces qui n'ont pas été dosés. Les éléments dosés sont les 8 oxides suivants :  $\text{Na}_2\text{O}$ ,  $\text{MgO}$ ,  $\text{Al}_2\text{O}_3$ ,  $\text{SiO}_2$ ,  $\text{K}_2\text{O}$ ,  $\text{CaO}$ ,  $\text{BaO}$  et  $\text{Fe}_2\text{O}_3$ . Une représentation graphique simple des données :

```
boxplot(fgl[, rev(1 + order(colMeans(fgl[, 2:9])))], las = 1, main = paste("Composition chimique de",
  nrow(fgl), "éclats de verre"), xlab = "Oxides", ylab = "Contribution au total en pourcent")
```



montre que ces éclats de verre sont chimiquement bien définis avec une prédominance d'oxide de silicium ( $\text{SiO}_2$ ) qui contribue à près de 73 % du total. Il est suivi par l'oxide de sodium ( $\text{Na}_2\text{O}$ ) et l'oxide de calcium ( $\text{CaO}$ ) qui contribuent à hauteur de 13 % et 9 %, respectivement. Ces trois oxides contribuent donc à eux seuls à 95 % de la composition totale des éclats de verre. On peut donc faire une représentation triangulaire pour ces trois principaux composants.

```
library(ade4)
triangle.plot(fgl[, c("Si", "Na", "Ca")])
```



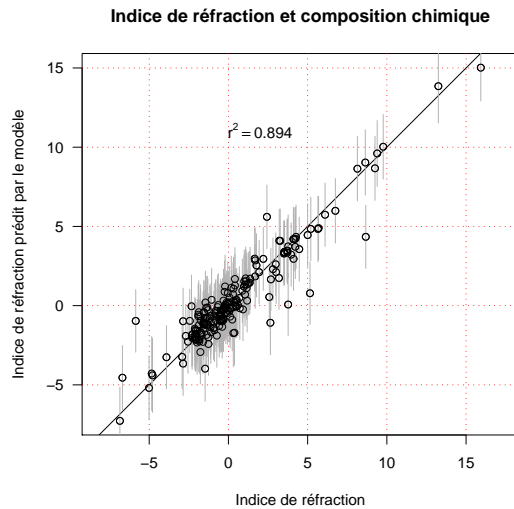
On constate que si la contribution de  $\text{Na}_2\text{O}$  est relativement constante, on a plus de variabilité pour les deux autres oxides (forme allongée du nuage de point vers le haut).

## 2.4 L'indice de réfraction et la composition chimique

L'indice de réfraction est une propriété physique globale des éclats de verre, on devrait logiquement pouvoir le prédire à partir de leur composition chimique. Essayons un modèle linéaire simple sans interactions avec ordonnée à l'origine sélectionné sur critère BIC [10].

```
data <- fgl[, 1:9]
(lm1BIC <- step(lm(RI ~ ., data), trace = 0, k = log(nrow(data))))
Call:
lm(formula = RI ~ Na + Mg + K + Ca + Ba, data = data)
Coefficients:
(Intercept)      Na      Mg      K      Ca      Ba
  -48.802      1.247      1.717      1.208      2.986      2.813

x <- fgl$RI
y <- predict(lm1BIC)
plot(x, y, las = 1, asp = 1, xlab = "Indice de réfraction", ylab = "Indice de réfraction prédit par le modèle"
      main = "Indice de réfraction et composition chimique")
abline(c(0, 1))
grid(col = "red")
r2 <- round(cor(x, y)^2, 3)
text(2, 10, bquote(r^2 == .(r2)), pos = 3)
suppressWarnings(cint <- predict(lm1BIC, interval = "predict"))
segments(x, cint[, "lwr"], x, cint[, "upr"], col = grey(0.7))
points(x, y)
```

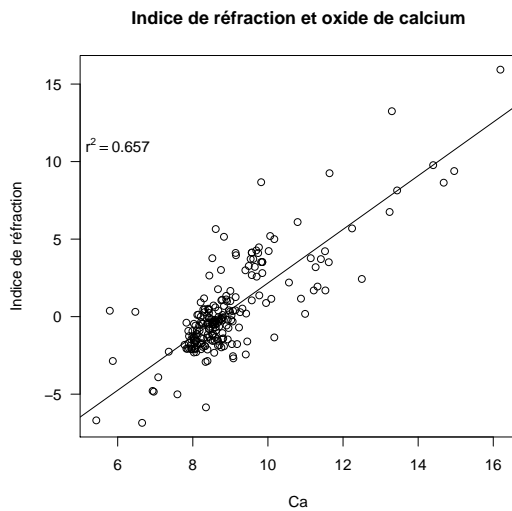


On arrive donc à bien prédire l'indice de réfraction à partir de la composition chimique. Une analyse de la variance montre :

```
options(show.signif.stars = FALSE)
anova(lm1BIC)
Analysis of Variance Table
Response: RI
      Df Sum Sq Mean Sq F value    Pr(>F)
Na     1   72.33   72.33   72.551 3.279e-15
Mg     1   64.88   64.88   65.082 5.643e-14
K      1  265.79  265.79  266.602 < 2.2e-16
Ca     1 1111.29 1111.29 1114.696 < 2.2e-16
Ba     1  242.74  242.74  243.487 < 2.2e-16
Residuals 208  207.36    1.00
```

que c'est l'oxide de Calcium (CaO) qui a le plus d'effet, ce que l'on peut vérifier par une représentation directe :

```
x <- fgl$Ca
y <- fgl$RI
plot(x, y, las = 1, xlab = "Ca", ylab = "Indice de réfraction",
      main = "Indice de réfraction et oxide de calcium")
abline(lm(y ~ x))
r2 <- round(cor(x, y)^2, 3)
text(6, 10, bquote(r^2 == .(r2)), pos = 3)
```

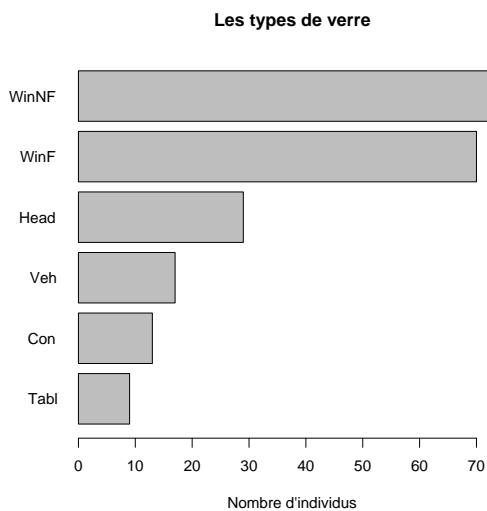


La prise en compte de la concentration en oxide de calcium explique à elle seule 66 % de la variabilité de l'indice de réfraction dans ce jeu de données. La prédiction des propriétés physiques des verres à partir de leur composition chimique est un problème ancien [12] qui a donné lieu à de nombreuses publications de part leur importance économique pour l'optimisation des procédés de fabrication [8].

## 2.5 Les types de verre

L'intérêt de ce jeu de données et que les verres étudiés appartiennent à des groupes.

```
(grps <- table(fgl$type))
WinF WinNF Veh Con Tabl Head
70 76 17 13 9 29
barplot(grps[order(grps)], horiz = TRUE, las = 1, main = "Les types de verre",
        xlab = "Nombre d'individus")
```





La signification des groupes est donnée ci-dessous. Le terme flotté pour un verre fait référence à un procédé de fabrication dans lequel le verre en fusion est versé sur une couche de métal liquide.

**WinNF** *window non-float glass* : fenêtre en verre non-flotté

**WinF** *window float glass* : fenêtre en verre flotté

**Head** *vehicle headlamps* : ampoules d'éclairage de véhicules

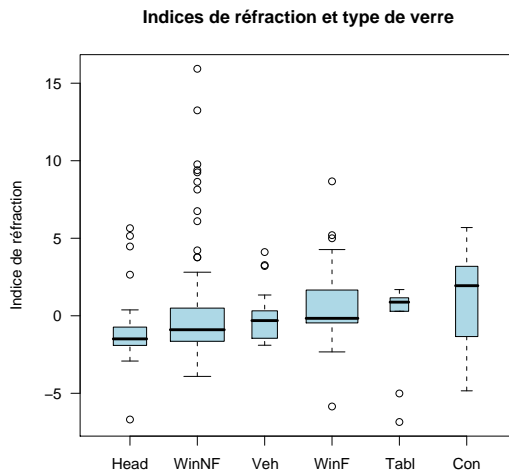
**Veh** *vehicle window glass* : fenêtres de véhicules (tous de type flotté)

**Con** *containers* : récipients

**Tabl** *tableware* : couverts en verre

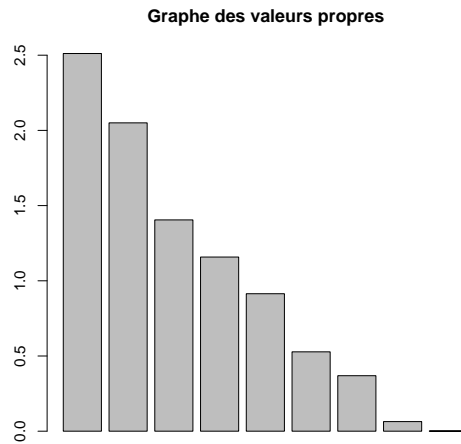
Si on représente la distribution des indices de réfraction par groupe :

```
boxplot(fgl$RI ~ fgl$type, main = "Indices de réfraction et type de verre",
        varwidth = TRUE, col = "lightblue", at = rank(as.vector(by(fgl$RI,
        fgl$type, median))), las = 1, ylab = "Indice de réfraction")
```



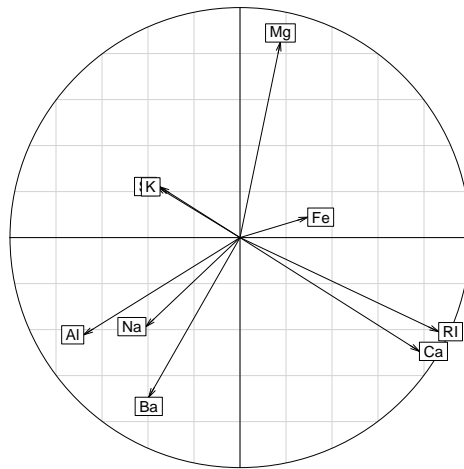
on voit qu'il va être difficile de prédire les groupes à partir du seul indice de réfraction. Faisons une analyse en composantes principale des données.

```
acp <- dudi.pca(fgl[, 1:9], scann = FALSE, nf = 2)
cumsum(100 * acp$eig/sum(acp$eig))
[1] 27.90182 50.68040 66.28978 79.15492 89.31050 95.17311 99.27265 99.98212
[9] 100.00000
barplot(acp$eig, main = "Graphe des valeurs propres")
```



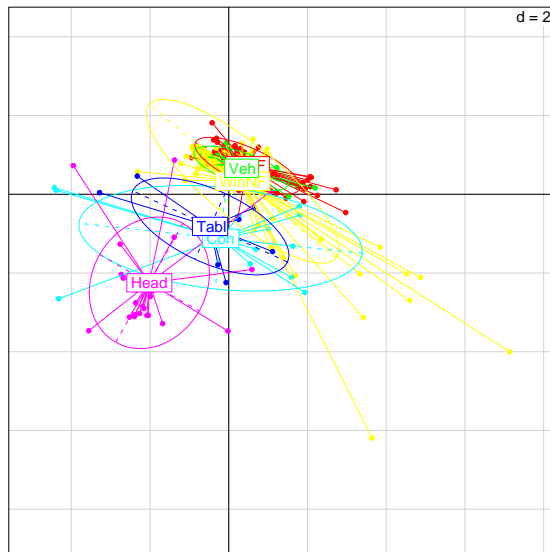
On a conservé ici deux facteurs qui extraient 50 % de la variabilité initiale.

```
s.corcircle(acp$co)
```



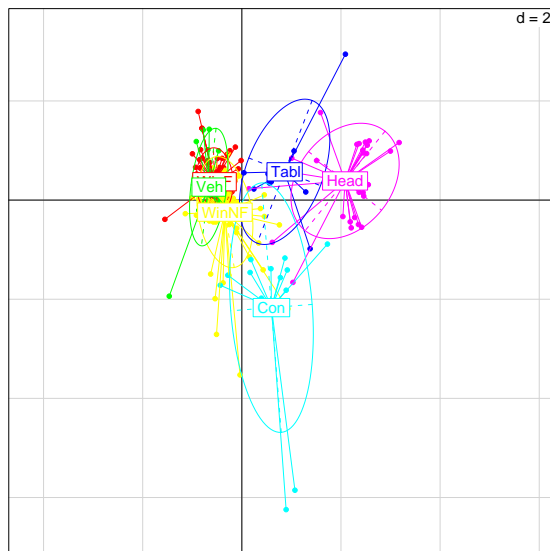
On retrouve ici la forte corrélation positive entre l'indice de réfraction et la concentration en oxide de calcium qui sort en premier facteur de la variabilité du jeu de données. Le deuxième facteur oppose l'oxide de magnésium aux oxides d'aluminium, de sodium et de barium.

```
s.class(acp$li, fgl$type, col = rainbow(6))
```



Sur le premier plan factoriel des individus on note la grande dispersion pour le groupe des verre de fenêtre non-flottés. Les groupes apparaissent sur le deuxième facteur mais sont mal résolus.

```
ad <- discrimin(acp, fgl$type, scann = FALSE, nf = 2)
s.class(ad$li, fgl$type, col = rainbow(6))
```



Une analyse discriminante n'améliore que peu les choses, on isole les ampoules d'éclairage des véhicules sur le premier axe et les récipients sur le second, mais il reste encore beaucoup de chevauchements. On comprend pourquoi ce jeu de donné est très utilisé pour comparer les méthodes de classification.

## 3 Hétérogénéité de l'indice de réfraction

### 3.1 Le jeu de données

L'intérêt de l'indice de réfraction est que celui-ci est obtenu par une analyse non destructive sur des éclats très petits. Pour une analyse donnée on va pouvoir faire plusieurs mesures indépendantes d'une même source. La variabilité des valeurs s'explique d'une part par les erreurs de mesure et d'autre part par l'hétérogénéité spatiale des verres.

L'objectif est d'analyser des morceaux de verre retrouvés sur une scène d'infraction et de les comparer, en règle générale, avec une vitrine, une vitre ou tout autre élément vitré. Prenons un exemple concret<sup>2</sup>. On dispose de deux échantillons de référence notés A et B caractérisés par 27 mesures de l'indice de réfraction :

```
refverre <- read.table("http://pbil.univ-lyon1.fr/R/donnees/refverreAB.txt",
  header = TRUE)
head(refverre)
  A      B
1 1.52106 1.52007
2 1.52115 1.52006
3 1.52114 1.52004
4 1.52124 1.52001
5 1.52122 1.52000
6 1.52124 1.52002
nrow(refverre)
[1] 27
```

On dispose également d'un échantillon dit de question caractérisé par 15 mesures :

```
echaver <- read.table("http://pbil.univ-lyon1.fr/R/donnees/echaver.txt",
  header = TRUE)
head(echaver)
  Q
1 1.52103
2 1.52102
3 1.52100
4 1.52103
5 1.52097
6 1.52103
nrow(echaver)
[1] 15
```

### 3.2 Comparaison avec fgl

Les valeurs sont données avec 6 chiffres significatifs comme dans les données originelles de `fgl`. On utilise la même transformation des données que dans `fgl` pour faciliter la lecture et les comparaisons.

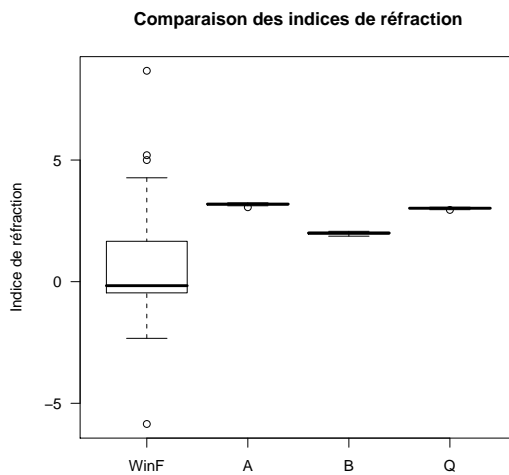
```
refverre <- 1000 * (refverre - 1.518)
summary(refverre)
  A      B
Min. :3.060  Min. :1.870
1st Qu.:3.150  1st Qu.:1.955
Median :3.190  Median :2.000
Mean   :3.178  Mean   :1.990
3rd Qu.:3.205  3rd Qu.:2.030
Max.   :3.240  Max.   :2.070
```

<sup>2</sup>Ce sont de vraies données fournies par Laurence Dujourdy de l'INPS.

```
echaver <- 1000 * (echaver - 1.518)
summary(echaver)
  Q
Min.  :2.950
1st Qu.:3.000
Median :3.020
Mean   :3.013
3rd Qu.:3.030
Max.   :3.060
```

Comparons avec les valeurs de l'indice de réfraction pour le verre flotté de `rgl` pour vérifier qu'il n'y a pas d'erreur grossière.

```
comp <- list(WinF = fgl[fgl$type == "WinF", "RI"], A = refverre$A,
             B = refverre$B, Q = echaver$Q)
boxplot(comp, main = "Comparaison des indices de réfraction", las = 1,
         ylab = "Indice de réfraction")
```



Les valeurs sont parfaitement réalistes puisque dans la gamme des valeurs observées dans `rgl`. On note que la variabilité intra-échantillon est très faible par rapport à la variabilité inter-échantillons, ce qui est une bonne chose. On voit de plus que l'échantillon de question Q est bien plus proche de A que de B.

### 3.3 Comparaison avec d'autres écart-types

D'après la figure 6 de [3] l'écart type pour l'indice de réfraction d'une vitre échantillonnée en 49 points avec 10 mesures à chaque fois ( $n = 490$ ) est de l'ordre de 0.042 (cf. Fig 1). C'est remarquablement proche de ce que l'on a ici :

```
sd(refverre)
  A      B
0.04151339 0.05639705
sd(echaver)
  Q
0.03221949
```

Une étude plus ancienne [5] avait donné une valeur de l'ordre de 0.02 pour l'écart-type de l'indice de réfraction.

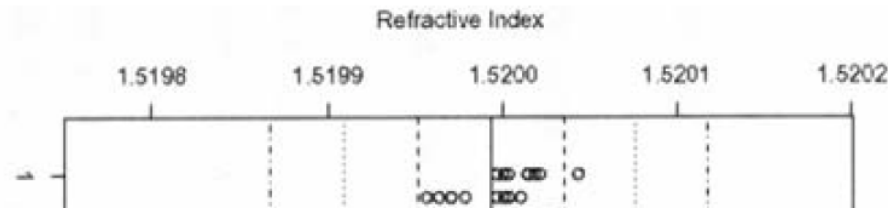


FIG. 1 – Copie d'écran d'une partie de la figure 6 de [3]. Les lignes verticales délimitent 6 écart-types. En translatant par la pensée vers la gauche pour que la dernière soit à 1.5201, on voit que l'étendue couverte par 6 écart-types n'est pas loin de  $1.5201 - 1.51985 = 0.00025$ , et en multipliant par 1000 à 0.25 dans nos unités. Un écart type vaut donc 6 fois moins, soit environ 0.042.

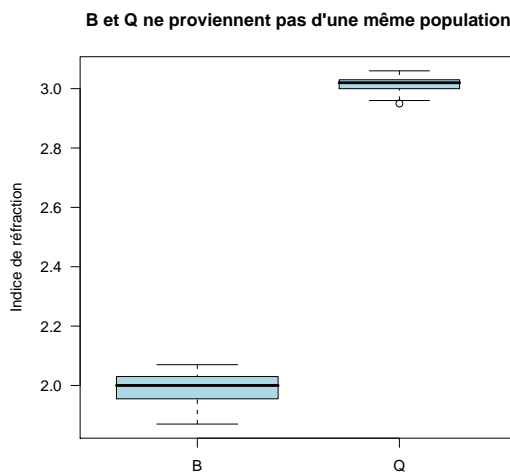
### 3.4 Rapprochement des échantillons Q et B

Il n'y a pas besoin de faire un test ici pour rejeter l'hypothèse nulle selon laquelle Q et B proviennent d'une même population. Faisons le quand même pour voir ce qu'il se passe dans ce cas de figure :

```
t.test(refverre$B, echaver$Q)
      Welch Two Sample t-test
data:  refverre$B and echaver$Q
t = -74.805, df = 39.929, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.0506028 -0.9953231
sample estimates:
mean of x mean of y
 1.990370  3.013333
```

La très faible p-value ici nous indique que nous sommes en train d'énoncer une trivialité que nous pouvons représenter graphiquement avec une boîte à moustaches ainsi :

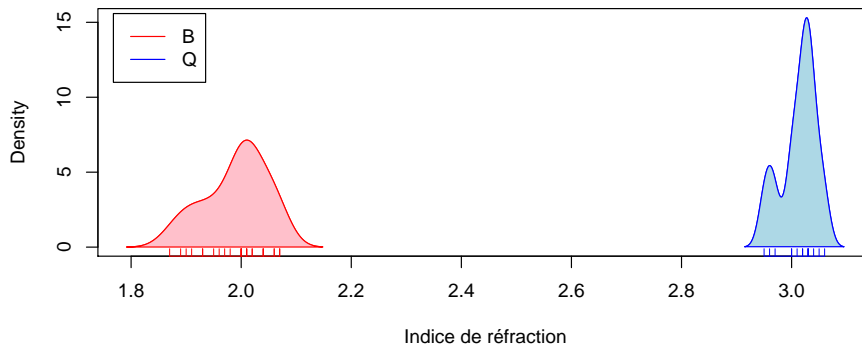
```
boxplot(list(B = refverre$B, Q = echaver$Q), varwidth = TRUE, col = "lightblue",
        las = 1, ylab = "Indice de réfraction", main = "B et Q ne proviennent pas d'une même population")
```



Une autre représentation possible est la suivante :

```
dstB <- density(refverre$B)
dstQ <- density(echaver$Q)
plot(dstB$x, dstB$y, type = "l", xlim = range(dstB$x, dstQ$x), ylim = range(dstB$y,
  dstQ$y), ylab = "Density", col = "red", xlab = "Indice de réfraction",
  main = "B et Q ne proviennent pas d'une même population")
lines(dstQ$x, dstQ$y, col = "blue")
polycurve <- function(x, y, base.y = min(y), ...) {
  polygon(x = c(min(x), x, max(x)), y = c(base.y, y, base.y),
    ...)
}
polycurve(dstB$x, dstB$y, col = "pink", border = "red")
polycurve(dstQ$x, dstQ$y, col = "lightblue", border = "blue")
rug(refverre$B, col = "red")
rug(echaver$Q, col = "blue")
legend("topleft", inset = 0.02, legend = c("B", "Q"), col = c("red",
  "blue"), lty = 1)
```

**B et Q ne proviennent pas d'une même population**

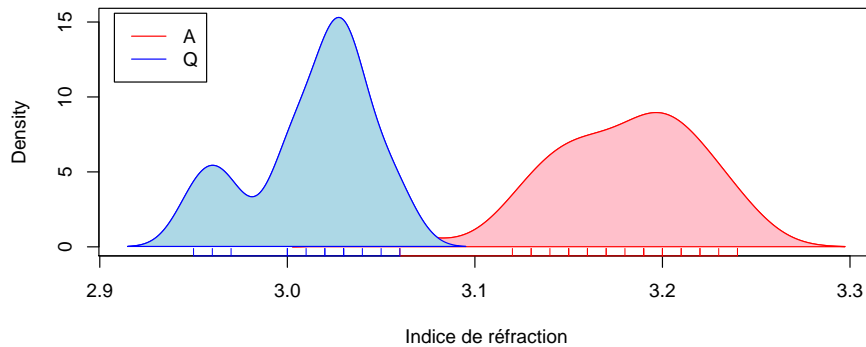


### 3.5 Rapprochement des échantillons Q et A

La réponse à la question est bien moins évidente ici. Commençons par faire une représentation graphique.

```
dstA <- density(refverre$A)
dstQ <- density(echaver$Q)
plot(dstA$x, dstA$y, type = "l", xlim = range(dstA$x, dstQ$x), ylim = range(dstA$y,
  dstQ$y), ylab = "Density", col = "red", xlab = "Indice de réfraction",
  main = "A et Q proviennent-ils d'une même population ?")
lines(dstQ$x, dstQ$y, col = "blue")
polycurve(dstA$x, dstA$y, col = "pink", border = "red")
polycurve(dstQ$x, dstQ$y, col = "lightblue", border = "blue")
rug(refverre$A, col = "red")
rug(echaver$Q, col = "blue")
legend("topleft", inset = 0.02, legend = c("A", "Q"), col = c("red",
  "blue"), lty = 1)
```

## A et Q proviennent-ils d'une même population ?



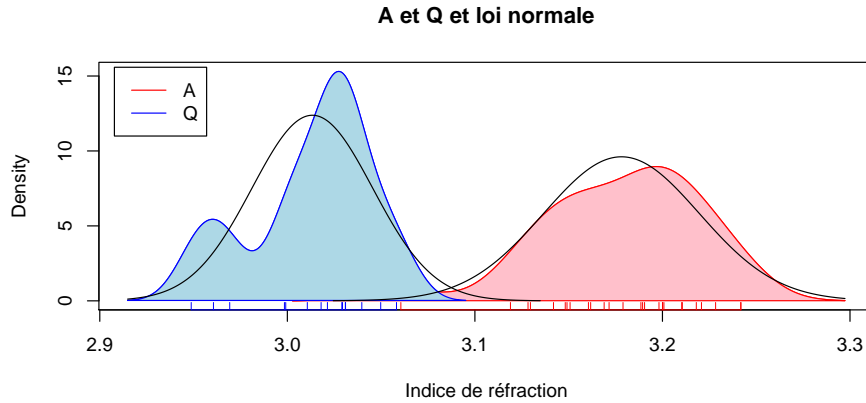
On serait donc tenté de rejeter l'hypothèse nulle ici. Ce qui correspond au résultat du test t classique.

```
t.test(refverre$A, echaver$Q)
      Welch Two Sample t-test
data:  refverre$A and echaver$Q
t = 14.2894, df = 35.481, p-value = 2.747e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.1414108 0.1882189
sample estimates:
mean of x mean of y
 3.178148  3.013333
```

Le problème du test t est qu'il suppose une distribution normale du type de la courbe en noir ci-dessous.

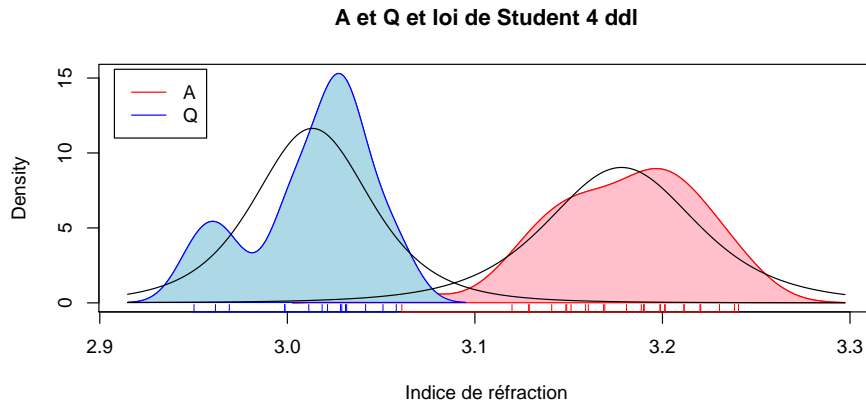
```
dstQ <- density(echaver$Q)
plot(dstA$x, dstA$y, type = "l", xlim = range(dstA$x, dstQ$x), ylim = range(dstA$y,
  dstQ$y), ylab = "Density", col = "red", xlab = "Indice de réfraction",
  main = "A et Q et loi normale")
lines(dstQ$x, dstQ$y, col = "blue")
polycurve(dstA$x, dstA$y, col = "pink", border = "red")
polycurve(dstQ$x, dstQ$y, col = "lightblue", border = "blue")
rug(jitter(refverre$A), col = "red")
rug(jitter(echaver$Q), col = "blue")
legend("topleft", inset = 0.02, legend = c("A", "Q"), col = c("red",
  "blue"), lty = 1)
tot <- c(dstA$x, dstQ$x)
x <- seq(from = min(tot), to = max(tot), length = 500)
norm.A <- dnorm(x, mean(refverre$A), sd(refverre$A))
norm.Q <- dnorm(x, mean(echaver$Q), sd(echaver$Q))
qui <- norm.A > 0.01
lines(x[qui], norm.A[qui])
qui <- norm.Q > 0.01
lines(x[qui], norm.Q[qui])
```





Toute la question est de savoir si cette loi normale est réaliste ou non. En utilisant une distribution de student à 4 ddl comme dans [3] on aurait :

```
plot(dstA$x, dstA$y, type = "l", xlim = range(dstA$x, dstQ$x), ylim = range(dstA$y,
  dstQ$y), ylab = "Density", col = "red", xlab = "Indice de réfraction",
  main = "A et Q et loi de Student 4 ddl")
lines(dstQ$x, dstQ$y, col = "blue")
polycurve(dstA$x, dstA$y, col = "pink", border = "red")
polycurve(dstQ$x, dstQ$y, col = "lightblue", border = "blue")
rug(jitter(refverre$A), col = "red")
rug(jitter(echaver$Q), col = "blue")
legend("topleft", inset = 0.02, legend = c("A", "Q"), col = c("red",
  "blue"), lty = 1)
t.A <- dt((x - mean(refverre$A))/sd(refverre$A), df = 4)/sd(refverre$A)
lines(x, t.A)
t.Q <- dt((x - mean(echaver$Q))/sd(echaver$Q), df = 4)/sd(echaver$Q)
lines(x, t.Q)
```



On a une moins bonne résolution entre les deux groupes parce que la loi de Student a des extrémités de la distribution plus épaisses.

## Références

- [1] A.G. Alvarez. *FKBL : fuzzy knowledge base learning*, 2007. R package version 0.50-4.

- [2] A. Asuncion and D.J. Newman. UCI machine learning repository, 2007.
- [3] R.L. Bennett, J.M. Curran, N.D. Kim, S.A. Coulson, and A.W.N. Awn Newton. Spatial variation of refractive index in a pane of float glass. *Science & Justice*, 43 :71–76, 2003.
- [4] J.M. Curran. The statistical interpretation of forensic glass evidence. *International Statistical Review*, 71 :497–520, 2003.
- [5] M.D.G. Dabbs and E.F. Pearson. The variation in refractive index and density across two sheets of window glass. *Journal of the Forensic Science Society*, 10 :139–148, 1970.
- [6] F. Leisch and E. Dimitriadou. *mlbench : Machine Learning Benchmark Problems*, 2008. R package version 1.1-4.
- [7] F. Leisch, K. Hornik, and B.D. Ripley. *mda : Mixture and flexible discriminant analysis*, 2006. S original by Trevor Hastie & Robert Tibshirani. R port by Friedrich Leisch and Kurt Hornik and Brian D. Ripley. R package version 0.3-2.
- [8] A.I. Priven and O.V. Mazurin. Comparison of methods used for the calculation of density, refractive index and thermal expansion of oxide glasses. *Glass Technol.*, 44 :156–166, 2003.
- [9] K. Schliep and K. Hechenbichler. *kknn : Weighted k-Nearest Neighbors*, 2007. R package version 1.0-5.
- [10] G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6 :461–464, 1978.
- [11] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer, New York, fourth edition, 2002. ISBN 0-387-95457-0.
- [12] A. Winkelmann and O. Schott. Über die Elastizität und über die Zugund Druckfestigkeit verschiedener neuer Gläser in ihrer Abhängigkeit von der chemischen Zusammensetzung. *Ann. Physik Chemie*, 51 :S697–S730, 1894.