

## TP 2 : Lois continues et analyse en composantes principales

### 1 Lois usuelles continues

#### 1.1 Densités

Pour tracer la densité de la loi  $\mathcal{N}(0, 1)$ , on peut utiliser les commandes suivantes :

```
x <- seq(from=-4, to=4, by=0.01)
y <- dnorm(x, mean=0, sd=1)
plot(x, y, type='l')
```

Tracer de même la densité des lois  $\mathcal{N}(1, 2)$  (de moyenne 1 et variance 2),  $\mathcal{N}(1, 0.5)$ ,  $\mathcal{E}(1)$  et de la loi de Cauchy de paramètres 0 et 1.

#### 1.2 Fonctions de Répartition

Tracer la fonction de répartition des lois  $\mathcal{N}(0, 1)$ ,  $\mathcal{N}(-2, 1)$  et  $\mathcal{E}(2)$ .

Pour  $X$  de loi  $\mathcal{N}(0, 1)$ , calculer  $P(X \leq 2)$ ,  $P(-1 \leq X \leq 0.5)$  et trouver  $u$  tel que  $P(X \leq u) = 0.9$ .

### 2 Analyse en composantes principales

On continue avec le jeu de données du TP1. On rappelle qu'on le charge avec les commandes :

```
Don <- read.delim(
  "http://math.univ-lyon1.fr/homes-www/dabrowski/enseignement/ProbaL2/Donnees.csv");
attach(Don)
```

- Évaluer les corrélations entre les 4 différentes caractéristiques quantitatives : age, poids, taille, alcool (en utilisant la fonction `cor`).
- Évaluer l'échantillon des variables centrées réduites associées aux 4 différentes caractéristiques quantitatives. On pourra utiliser la commande `apply(A, MARGIN = 2, FUN = sd)` après avoir converti en matrice  $A$  le data.frame.
- Évaluer la covariance de ces variables centrées réduites (en utilisant la fonction `cov`).
- Effectuez l'analyse en composantes principales.

On rappelle qu'il faut

1. Trouver les valeurs propres de la matrice des corrélations  $C$  des variables. En déduire combien de composantes sont nécessaires pour capturer l'essentiel des données.
2. Trouver la matrice de passage de la base initiale vers la base de diagonalisation de  $C$  (c'est la matrice dont les colonnes sont les vecteurs propres de  $C$ ).
3. Obtenir les données dans la nouvelle base des composantes principales.

Représentez par un nuage de points les 2 composantes principales de l'ACP puis le cercle des corrélations, on pourra utiliser

```
install.packages("plotrix");
library("plotrix")
```

et la commande `draw.circle(0, 0, 1)` pour représenter le cercle. Quelles sont les variables de départ bien représentées par ces 2 composantes ?

- Quelle est la variance empirique des deux composantes principales (obtenir cette variance de deux façons) ?