TP 2 : Lois continues et analyse en composantes principales

1 Lois usuelles continues

1.1 Densités

Tracer la densité de la loi $\mathscr{N}(0,1)$. Pour cela, 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 suivantes :

(a) les lois normales $\mathscr{N}(1,2)$ (de moyenne 1 et variance 2) et $\mathscr{N}(1,0.5)$ (sur un même graphique) ;

(b) la loi exponentielle $\mathscr{E}(1)$ (quelle est sa dentité ?) ;

(c) la loi de Cauchy de paramètres 0 et 1 (quelle est sa dentité ?).

1.2 Fonctions de répartition

Tracer la fonction de répartition des lois $\mathscr{N}(0,1)$, $\mathscr{N}(-2,1)$ et $\mathscr{E}(2)$.
Pour X de loi $\mathscr{N}(0,1)$, 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 utilise le jeu de données `iris` inclus dans R. Cet échantillon décrit, pour 150 fleurs, la longueur et la largeur du sépale et du pétale, ainsi que l'espèce parmi trois possibles (lesquelles ?). Un des buts de l'étude est de séparer les trois espèces d'iris.

1. Visualisation préliminaire : **View**(iris) et **pairs**(iris). Calculer le résumé numérique : **summary**(iris). Tracer les *boxplots* par espèce avec : **boxplot**(iris**$**Petal.Length~iris**$**Species).
2. Évaluer les corrélations entre les quatre caractéristiques numériques (avec la fonction **cor**).
3. Évaluer l'échantillon des variables centrées réduites associées aux quatre différentes caractéristiques quantitatives (on pourra convertir en matrice A le **data.frame** et utiliser la commande suivante : **apply**(A,MARGIN=2,FUN=**sd**)).
4. Évaluer la covariance de ces variables centrées réduites (avec la fonction **cov**). Comparer au résultat de la commande **cor**(iris [,1:4]) . Interpréter.
5. Effectuer l'analyse en composantes principales. On rappelle qu'il faut :
   (a) trouver les valeurs propres de la matrices des corrélations **C** des variables. En déduire combien de composantes sont nécessaires pour capturer l'essentiel des données ;
   (b) trouver la matrice de passage M de la base initiale vers la base de diagonalisation de **C** (c'est la matrice dont les colonnes sont les vecteurs propres de **C**) ;
   (c) obtenir les données [1] dans la nouvelle base des composantes principales et tracer le nouveau nuage, si possible en trois couleurs (une par espèce ; option **col**='green' pour tracer en vert) ;
   (d) ceci fait, représenter par un nuage de points les deux composantes principales de l'ACP puis le cercle des corrélations [2] pour tracer le cercle. Quelles sont les variables de départ bien représentées par ces deux composantes ?
6. Calculer de deux façons la variance empirique des deux composantes principales.

---

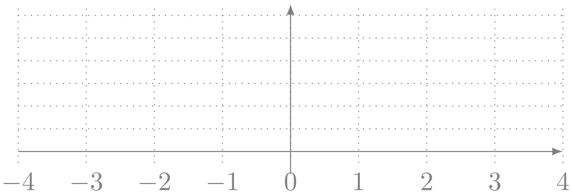
1. Rappel : agir sur les colonnes d'une matrice, c'est multiplier *à droite* par une matrice : par M ou par M^{-1} ?

2. Première méthode : importer une fois pour toutes le paquet plotrix (avec **library**('plotrix') ; si nécessaire, l'installer avec **install.packages**('plotrix')) ; dans le **plot** du nuage, ajouter l'option **asp**=1 ; puis tracer le cercle (avec **draw.circle**(0,0,1)). On pourra aussi utiliser **points**(**cos**(**c**(1:700)/100),**sin**(**c**(1:700)/100),**type**="l").

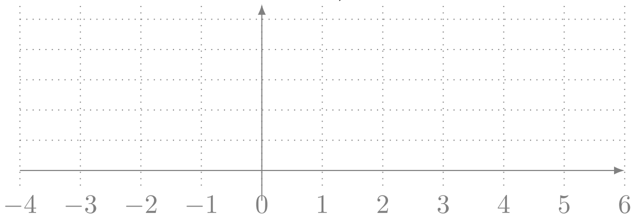
1 Lois usuelles continues

1.1 Densités

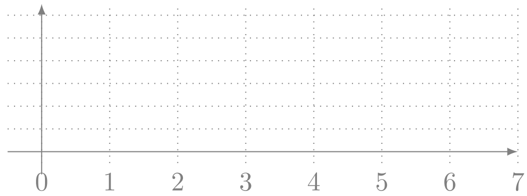
Esquisser les diagrammes (préciser l'échelle verticale ; utiliser deux couleurs).



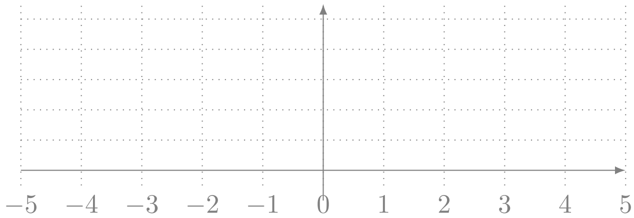
Loi $\mathscr{N}(0,1)$



Lois $\mathscr{N}(1,2)$ et $\mathscr{N}(1,.5)$



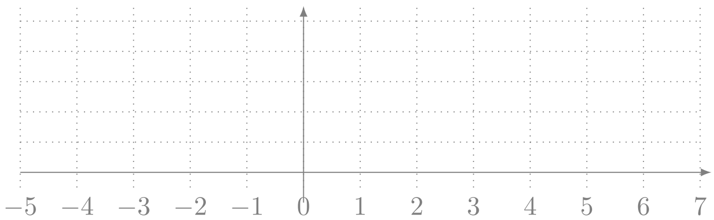
Lois exponentielles $\mathscr{E}(1)$ et $\mathscr{E}(2)$



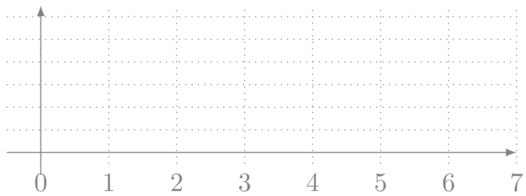
Loi de Cauchy de paramètres 0 et 1 et $\mathscr{N}(0,\sigma_0^2)$

Ci-dessus, σ_0^2 est choisi pour que $\mathscr{N}(0,1)$ ait le même maximum que la loi de Cauchy : $\sigma_0^2 =$ ………

1.2 Fonctions de répartition



Fonctions de répartition de $\mathscr{N}(0,1)$ et $\mathscr{N}(2,1)$



Fonction de répartition de la loi $\mathscr{E}(1)$

Si $X \sim \mathscr{N}(0,1)$, $P(X \leq 2) =$ ……… ; $P(-1 \leq X \leq 0{,}5) =$ ……… ; $P(X \leq u) = 0{,}9$ si $u =$ ……… .

2 Analyse en composantes principales

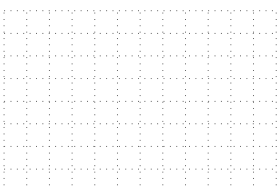
Remarques sur les *boxplots* :

corrélation empirique des variables initiales				
	lng sép	lrg sép	lng pét	lrg pét
lng sép				
lrg sép				
lng pét				
lrg pét				

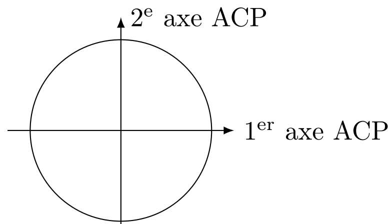
covariance emp. des variables centrées réduites				
	lng sép	lrg sép	lng pét	lrg pét
lng sép				
lrg sép				
lng pét				
lrg pét				

Rôle de la fonction **cor** :

Valeurs propres de la matrice des corrélations : $\lambda_1 \simeq$ ………… ; $\lambda_2 \simeq$ ………… ; $\lambda_3 \simeq$ ………… ; $\lambda_4 \simeq$ ………… .



allure du nouveau nuage



cercle des corrélations