Exercices sur les lois à queue lourde

Exercice 1

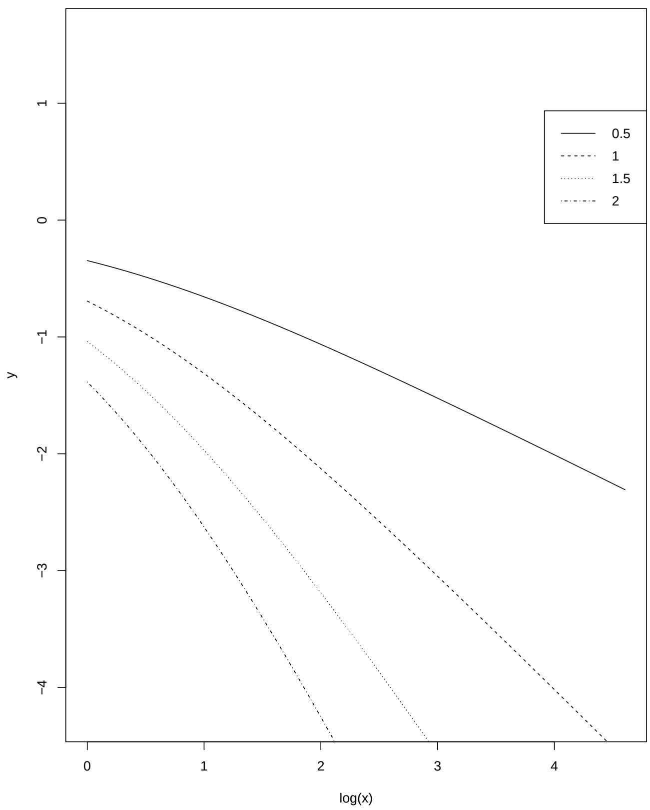
1. Plusieurs fonctions R permettent de calculer les densités des lois de valeurs extrêmes. On peut utiliser les codes ci dessous pour tracer les densités des lois Fréchet, Weibull et Gumbel.

```
library(evd)
x1 = seq(0, 5, 0.01); y = dfrechet(x1);
plot(x1, y, type = "l", xlim = c(-5, 5), ylim = c(0, 1))
x2 = seq(-5, 0, 0.01); y = drweibull(x2); points(x2, y, type = "l", col = "blue")
x3 = seq(-5, 5, 0.01); y = dgumbel(x3); points(x3, y, type = "l", col = "red")

library(extRemes)
y = devd(x1, loc = 1, shape = 1); plot(x1, y, type = "l", xlim = c(-5,5), ylim = c(0,1))
y = devd(x2, loc = -1, shape = -1); points(x2, y, type = "l", col = "blue")
y = devd(x3); points(x3, y, type = "l", col = "red")
```

2. Expliquer le programme suivant. Quelle est la pente de la ligne presque droite obtenue ? Comment obtenir la figure ci dessous en variant le paramètre de loi de Paréto ? (indication : pour ajouter une légende `legend("topright", c("0.5", "1", "1.5", "2"), lty = 1:4)`)

```
library(actuar)
x = seq(1, 100, length.out = 1000); y = log(ppareto(x, 0.5, 1, lower.tail = F))
plot(log(x), y, type = "l", lty = 1, asp = 1)
```



3. Un jeu de données réelles : expliquer le programme ligne à ligne suivant. Qu'affiche le même programme, si l'on remplace `pparetolog` par `ppareto` ? Pourquoi cette différence ? Essayer d'autre méthode pour l'option `measure` et observer les résultats.

```
data(gdental); hist(gdental)
pparetolog <- function(x, shape, scale) ppareto(x, exp(shape), exp(scale))
( p <- mde(gdental, pparetolog, start = list(shape = 1,scale = 1), measure = "CvM") )
x = seq(0,4000,length.out=500); y = dpareto(x,exp(p$estimate)[1],exp(p$estimate)[2])
lines(x,y, col = "red")
```

Exercice 2

Soit $Y_0, Y_1, Y_2, \ldots$ une suite des variables aléatoires i.i.d. avec la fonction de répartition commune

$$F_Y(y) = \exp\left(-\frac{1}{(a+1)y}\right), \quad y > 0,$$

où $a \in [0,1]$ est un paramètre. Définissons le processus $X_0, X_1, X_2, \ldots$ comme suit

$$X_0 = Y_0, \quad X_i = \max(aY_{i-1}, Y_i), \quad i = 1, \ldots, n.$$

1. Montrez que la loi marginale du processus $X_0, X_1, X_2, \ldots$ est Fréchet de paramètre 1, c.a.d. pour $x > 0$,

$$\mathbb{P}(X_i \leq x) = \exp(-1/x), \quad i = 1, \ldots, n.$$

2. Soit $X_0^*, X_1^*, \ldots$ une suite des variables aléatoires i.i.d. de loi Fréchet de paramètre 1. Notons $M_n^* = \max(X_0^*, \ldots, X_{n-1}^*)$. Montrez que pour $z > 0$,

$$\mathbb{P}(M_n^* \leq nz) = \exp(-1/z).$$

3. Notons $M_n = \max(X_0, \ldots, X_{n-1})$. Montrez que pour $z > 0$,

$$\mathbb{P}(M_n \leq nz) = (\exp(-1/z))^{\frac{1}{a+1}}.$$

4. La figure ci-dessous présente les nuages de points du logarithme de la simulation de série $X_1, \ldots, X_{50}$ pour $a = 0, 1/3, 2/3,$ et 1. Le programme utilisé est ci dessous.

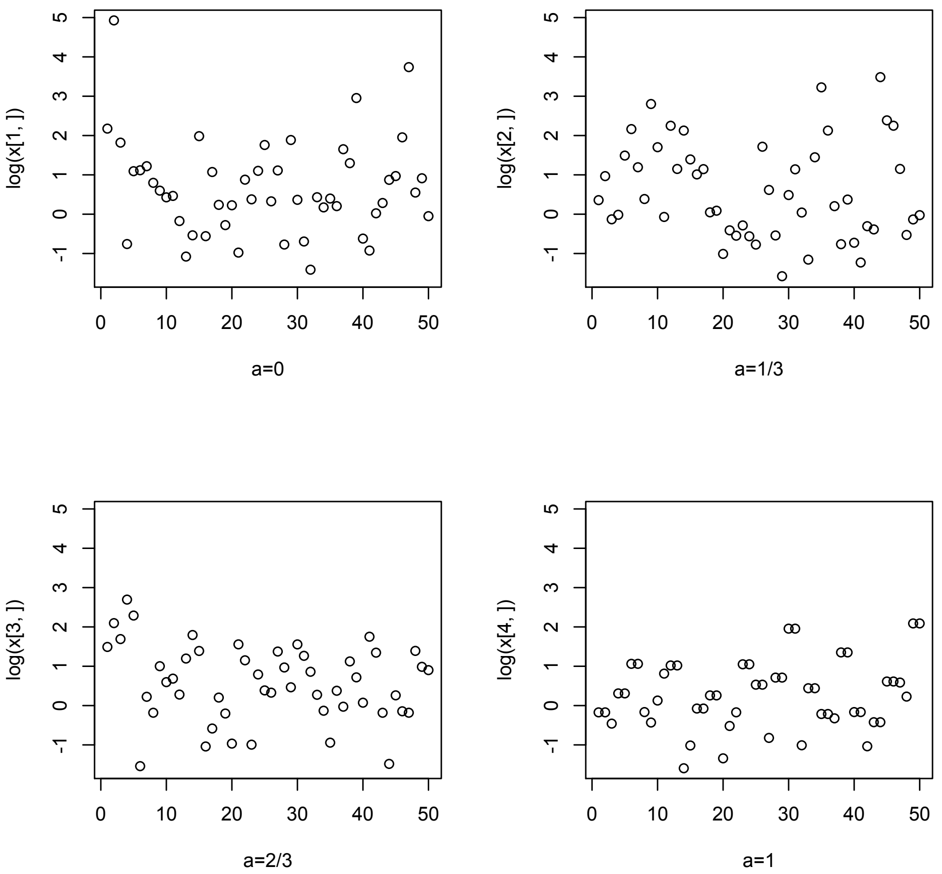
```
n = 51; a = c(0, 1/3, 2/3, 1); x = matrix(0, nrow = 4, ncol = n-1)
for(j in 1:4){
   u = runif(n); y = -1/(a[j]+1)/log(u)
   for(i in 1:(n-1)){x[j,i] = max(a[j]*y[i], y[i+1])}
}

par(mfrow=c(2,2))
plot(x[1,],type="p",xlab=c("a=0"))
plot(x[2,],type="p",xlab=c("a=1/3"))
plot(x[3,],type="p",xlab=c("a=2/3"))
plot(x[4,],type="p",xlab=c("a=1"))

par(mfrow=c(2,2))
plot(log(x[1,]),type="p",ylim=c(min(log(x)),max(log(x))),xlab=c("a=0"))
plot(log(x[2,]),type="p",ylim=c(min(log(x)),max(log(x))),xlab=c("a=1/3"))
plot(log(x[3,]),type="p",ylim=c(min(log(x)),max(log(x))),xlab=c("a=2/3"))
plot(log(x[4,]),type="p",ylim=c(min(log(x)),max(log(x))),xlab=c("a=1"))
```

Pouvez vous expliquer pourquoi les valeurs extrêmes apparaissent en groupe quand a augmente ? En particulier, si $a = 1$, les observations les plus grandes apparaissent toujours en couple.

5. En utilisant les résultats obtenus pour les questions 1-3, pouvez vous expliquer pourquoi le maximum des 50 observations diminue quand a augmente ?


log(x[1,])
a=0
log(x[2,])
a=1/3
log(x[3,])
a=2/3
log(x[4,])
a=1