TP machines n° 2 et 3 (groupes MASS 3 et Math Info 5)

Manipuler des vecteurs. Comparer des distributions

1. Introduction
On se propose ici de comparer l'évolution des indices boursiers allemand (DAX), français (CAC) et anglais (FTSE) entre 1992 et 1998 à partir du package `EuStockMarkets` (Remarque : on peut avoir accès à la liste des packages disponibles avec la commande `library(help="datasets")`). `EuStockMarkets` est une matrice à 4 colonnes contenant dans chacune les valuers quasi-journalières des 4 principales valeurs européennes

2. Isoler les vecteurs DAX, FTSE et CAC. Les comparer Dans un premier temps, nous allons construire 3 vecteurs `DAX`,`CAC` et `FTSE` contenant chacun l'évolution de l'indice normalisé de chacune des trois valeurs : `DAX=EuStockMarkets[,1]/EuStockMarkets[1,1]` ....
On veut ensuite comparer les différentes évolutions en traçant les 3 vecteurs (`plot(DAX)`), en traçant les diagrammes quantile-quantile (`qqplot(DAX,CAC)`), en calculant la distance de Kolmogorov-Smirnov (`ks.test(DAX,FTSE)`) et en regardant les coefficients de corrélation (`cor(DAX,CAC)`). Conclure. Peut-on affirmer d'après vous que l'évolution du DAX est plus proche de celle du FTSE que du CAC ?

3. Isoler le comosante stochastique/aléatoire de lévolution des stocks (enlever le taux d'intérêt sans risque) On peut remarquer que les différents indices de marchés croissent en moyenne au cours du temps (ils sont environ multipliés par 3 entre 1992 et 1998). En fait, chaque stock croit d'un certain montant chaque jour de façon quasi sûre (c'est ce que les économistes appellent le taux d'intérêt sans risque) et à cette croissance s'ajoute alors un aléa (qui peut induire une décroissance de l'indice d'un jour à l'autre mais en moyenne sur beaucoup de jours, l'indice croit). On veut ici enlever cette croissance pour comparer les évolutions purement aléatoires des indices. En premières approximation, on considère que le taux d'intérêt sans risque r est égal à :

$$r = \frac{1}{n}\sum_{i=1}^{n-1} \frac{\mathrm{S}_{i+1} - \mathrm{S}_i}{\mathrm{S}_i},$$

où S_i est la valeur de l'indice au jour i et n le nombre total de jours considérés. Calculer le taux d'intérêt sans risque r pour chaque indice (correction : `rDAX=mean(DAX[-1]/DAX[-length(DAX)])-1` (où `DAX[-1]` représente le vecteur DAX sans la première valeur et `DAX[-length(DAX)]` représente le vecteur DAX sans sa dernière valeur.).

4. Tester le modèle de Black-Scholes D'après le modèle de Black-Scholes $w_i = \frac{\mathrm{S}_{i+1}-\mathrm{S}_i}{\mathrm{S}_i} - r$ est un incrément purement aléatoire suivant une loi normale. On se propose de tester cette hypothèse avec un test du Khi-deux. Dans un premier temps, nous construisons donc le vecteur `wDAX` des incréments aléatoires supposés suivre une loi normale (correction `wDAX=(DAX[-1]-DAX[-length(DAX)])/DAX[-length(DAX)] - rDAX`). Ensuite on regarde si ce vecteur suit une loi normale (correction : `ks.test(wDAX,"pnorm",mean(wDAX),sd(wDAX))`). Conclure.

Écrire des scripts contenant des fonctions

1. Introduction On se propose ici d'écrire des fichiers sources (.R) dans un éditeur de texte contenant des fonctions plus ou moins complexes que l'on chargera ensuite directement sous R.

Première fonction Lancer l'éditeur. On cherche à écrire une fonction `rapport <- function(x)...` qui retourne le rapport entre le 9^{ime} décile et le 1^{er} décile ($\hat{\mathrm{F}}(0.9)/\hat{\mathrm{F}}(0.9)$) d'une distribution donnée x. Charger ensuite le programme dans R et tester votre fonction sur `x=seq(1,10)` par exemple.

Tracer la courbe de Lorentz La courbe de Lorentz $\mathrm{L}(x)$ représente graphiquement la diversité d'une distribution (différences de niveau de vie..). Pour une statistique d'ordre $x_{(i)}, 1 \leqslant i \leqslant n$, pour $\frac{k}{n} \leqslant x < \frac{k+1}{n}$ (*i.e* k est la partie entière de $x * n$ avec $x \in [0;1]$) on a :

$$\mathrm{L}(x) = \frac{\sum_{i=1}^{k} x_{(i)} + x * n - k}{n * \bar{x}_n}.$$

Compléter la fonction `lorentz <- function(x, r)  lo <- numeric(r) y<-numeric(length(x)) y=sort(x) y contient les la statistique d'ordre de x for(i in 1 :r)  ... plot(seq(1/r,1,by=1/r),lo) ;` qui renvoie la courbe de Lorentz d'une distribution `x` avec `r` le nombre de valeurs uniformément réparties dans $[0;1]$ pour lesquelles est calculée la valeur de la courbe de Lorentz `lo`. On utilisera la fonction `floor(w)` qui renvoie la partie entière d'un nombre `w` : `floor(5.6)=5`. Tester la fonction sur une distribution constante `x=rep(1,30` et sur deux distributions normales avec des écart types différents. Conclure.

Calculer l'indice de Gini d'une distribution L'indice de Gini est un indice très connu permettant de mesurer la variabilité d'une série statistique (et donc comparer les inégalités de salaire entre différents pays..) et est égal à deux fois l'aire comprise entre la courbe de Lorentz et la diagonale d'égalité parfaite. Une formule facile pour calculer l'indide de Gini g_n d'une série statistique $x_i, 1 \leqslant i \leqslant n$ est :

$$g_n = \frac{2}{n^2} \sum_{i=1}^{n} i * \frac{x_{(i)}}{\bar{x}_n} - \frac{n+1}{n}.$$

À partir de la définition précédente, écrire une fonction `gini <- function(x)...` renvoyant l'indice de Gini d'une série statistique `x`.