M1 finance: Statistiques non paramétriques
Fiche 3: Tests d'adéquation

Emeline Schmisser, `emeline.schmisser@math.univ-lille1.fr`, bureau 314 (bâtiment M3).

Exercice 1.
On observe les variables suivantes : 3,9 ; 3,6 ; 3,3 ; 3,1 ; 2,7.

1. Est-ce que ces données suivent une loi uniforme sur [2,4] ? (niveau $\alpha = 10\%$).

Exercice 2 : Estimation de la fonction de répartition

1. On veut estimer F_X de telle sorte que pour tout x, $|F_n(x) - F_X(x)| \leq 0.20$ avec probabilité 0.90. Combien d'observations doit-on au minimum avoir ?
2. Même question, mais cette fois-ci on veut un intervalle de confiance à 95%.
3. On observe 13 variables aléatoires d'une fonction de répartition inconnue. Les observations sont 3.5, 4.1, 4.8, 5.0, 6.3, 7.1, 7.2, 7.8, 8.1, 8.4, 8.6, 9. Dessiner F_n et la zone de confiance pour l'estimation de F_X.

Exercice 3 : Test de Lilliefor pour les exponentielles
À partir du test de Kolmogorov-Smirnov, on veut construire un test qui permette de vérifier si des variables aléatoires $(X_1, \ldots, X_n)$ suivent une loi exponentielle de paramètre $1/\lambda$. On calcule donc

$$D_n = \sup_x |F_n(x) - \hat{F}_X(x)|$$

avec $\hat{F}_X(x) = 1 - e^{-x/\hat{\lambda}}$ où $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^n X_i$. Avant de construire une table, il faut vérifier que la loi de D_n ne dépend pas de λ. Posons $Y_i = X_i/\lambda$.

1. Montrer que $Y_i \sim \mathcal{E}(1)$.
2. Calculer D_n^+ en fonction des X_i.
3. Calculer D_n^+ en fonction des Y_i.
   Si la dernière expression ne dépend pas de λ, alors la loi de D_n^+ ne dépend pas de λ. Par symétrie, on conclut que D_n ne dépend pas de λ. On peut donc construire une table pour ce test.

Exercice 4.
On souhaite étudier les log-return de SP500. Le fichier est sur Moodle. On souhaite savoir quelle est la densité de ces variables.

1. Ouvrir le fichier r500 avec `R`. Attention, `R` considère r500 comme une matrice, il faut indiquer à `R` de ne considérer que la première colonne de cette matrice.

On souhaite savoir si la densité peut être gaussienne

2. Utilisez les fonctions `qqnorm` et `qqline`. Commentez les résultats. Comment sont les queues de distribution par rapport à une loi normale ?
3. Faites un test statistique pour savoir si `r500` suit une loi normale. Vous pouvez utiliser `R` comme une super calculatrice, mais le test doit être rédigé sur feuille. Donnez une majoration de la p-valeur.

Puisque les log-return ne suivent pas une loi normale, ils suivent peut-être une loi de Student. On suppose maintenant que

$$R = \sigma\sqrt{\frac{k-2}{k}}T(k)$$

où σ^2 est la variance des log-return.

4. Calculer le coefficient de Kurtosis normalisé des log-return. Quelles valeurs de k pourraient convenir ?
5. À l'aide de `qqplot`, regardez quelle valeur de k pourrait convenir.
6. Faire maintenant le test. On comparera la valeur de la statistique de test avec celles obtenues précédemment.

   Remarque 1 – *Les statistiques de test calculées avec* `R` *sont dans cet exercice toutes de l'ordre de* $c10^{-2}$. *Si ce n'est pas le cas, il y a une erreur dans vos calculs ou vos fonctions.*
   – *La densité des log-return a des queues lourdes, et probablement pas de moments d'ordre 4.*
   – *La plupart de ces tests sont disponibles sous* `R`. *La fonction* `ks.test` *fait le test de Kolmogorov-Smirnov, et les fonctions* `lillie.test` *et* `ad.test`, *disponibles si on installe le paquet* `nortest`, *font les tests de Lilliefor et d'Anderson-Darling pour la normalité.*

Exercice 5 : Fonction de répartion de D_n
On veut construire un programme qui calcule $\mathbb{P}(D_n \leq a)$ pour tout n et tout $a \in [0, 1]$. La loi de D_n n'est pas définie explicitement, on va donc utiliser une méthode de Monte-Carlo.

1. Construire une fonction `Dn` qui étant donné n, simule n variables i.i.d de loi uniforme sur $[0, 1]$ et calcule la statistique D_n.
2. Construire une fonction `simul` qui étant donné n, N et a :
   – Calcule N fois la statistique de test D_n.
   – Calcule $P_{a,n}$ la probabilité empirique que D_n soit plus petite que a.
   – Renvoie $P_{a,n}$.
3. Quelle est la variance de $P_{a,n}$?
4. On veut calculer F_{D_n} à 0.5% près. Quelle doit être la valeur de N minimale pour que $\mathbb{P}(|P_{a,n} - F_{D_n}(a)| \leq 0.005) \geq 0.95$?
5. Cherchez (au centième près) la médiane de F_{D_n} pour $n = 10$.
6. Même question (au millième près) si $n = 100$.