M1 finance: Statistiques non paramétriques
Fiche 8: Estimation de la densité

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

Exercice 1 : Estimation de la dérivée d'une densité
Soient $X_1, X_2, \ldots, X_n$ des variables aléatoires i.i.d de densité f dérivable. On veut estimer f'. On suppose que K' et $f' \in L^2$.

1. Montrer que K et f admettent une limite en $\pm\infty$. Quelle est cette limite ?
2. Exprimer $f'_h = f' * K_h$ en fonction de f et de K'.
3. Proposer un estimateur de f pour h fixé.
4. Majorer le risque L^2 de cet estimateur.
5. Si $f' \in \mathcal{H}(\beta)$, quel est l'ordre de grandeur de la fenêtre h optimale ? Quel est le risque de l'estimateur dans ce cas ? Comment est le risque par rapport à l'estimation d'une densité ?
6. On veut minimiser le risque avec une méthode adaptative. Montrer que l'on veut minimiser une formule du type
$$F - 2G$$
7. Par quoi remplace-t-on F ?
8. Calculer l'espérance de G. On fera les modifications nécessaires pour que l'espérance dépende de f, mais pas de f'.
9. Proposer un estimateur sans biais de G.
10. Comment peut-on choisir h ?

Exercice 2 : Estimation d'une densité, méthodes à noyaux, R
On veut estimer la densité des log-return de SP500 et comparer cette densité à la densité d'une loi normale.

1. Stocker le vecteur r500 dans un vecteur X. Ce vecteur code les log-return de SP500. Représenter graphiquement les variations du log-return.
2. Choisir un noyau K parmi ceux vus en cours et construire une fonction K.
3. Choisir un intervalle I où vous aller construire un estimateur de la densité. Cet intervalle est évidemment fonction des données ! Construire un vecteur x qui varie doucement sur cet intervalle.
4. Construire une fonction est_h qui prenne en argument h et :
   – qui construise les trois matrices $m \times n$ (m étant la longueur de x, et n celle de X).
$$\begin{pmatrix} X_1 & \ldots & X_n \\ \ldots & \ldots & \ldots \\ X_1 & \ldots & X_n \end{pmatrix}, \quad \begin{pmatrix} x_1 & \ldots & x_1 \\ \ldots & \ldots & \ldots \\ x_m & \ldots & x_m \end{pmatrix}, \quad \begin{pmatrix} X_1 - x_1 & \ldots & X_n - x_1 \\ \ldots & \ldots & \ldots \\ X_1 - x_m & \ldots & X_n - x_m \end{pmatrix}$$
   – qui renvoie le vecteur $\hat{f}_h(x)$.
5. Tracer l'histogramme des fréquences de SP500 (on prendra un histogramme de 50 barres). Faire varier h. Superposer les courbes des estimateurs $\hat{f}_h$.
6. Comment évolue la courbe quand h est trop petit ? Trop grand ?

7. Trouver le meilleur h possible. D'après les résultats théoriques, on sait que le meilleur h est $n^{-1/(2\beta+1)}$. Il varie donc entre n^{-1} si la densité est très irrégulière et 1 si la densité est très régulière. Si $f \in \mathcal{H}(1)$, on a alors que $h_{opt} = n^{-1/3}$. Est-ce que cet h est proche de $n^{-1/3}$?

On va maintenant comparer cette densité à une densité d'une loi normale

8. Calculer la moyenne et la variance empirique des données. Tracer la densité normale correspondant à ces valeurs sur le graphe précédent. Est-ce que les deux densités sont proches ?
9. Calculer la médiane et la déviation médiane à la médiane. Tracer la densité normale correspondant à ces valeurs (toujours sur le même graphe). Comparer l'estimateur à noyau et cette courbe. Comment expliquer la différence avec l'estimateur obtenu précédemment ? (il ne suffit pas de dire que les estimateurs choisis ne sont pas les mêmes, il faut expliquer pourquoi ils sont différents).

Exercice 3 : Validation croisée, sous R

1. Construire une fonction *risque* à partir de est_h qui prenne en argument h et calcule $CV(h)$. (Discrétiser l'intégrale et utiliser une boucle pour calculer G). La fonction *risque* prend en argument h, X et y.
2. Trouver h qui minimise $CV(h)$ à 0.01 près. Pour cela :
   - Poser $h_{\hat{m}} = 1$ et $CV = 100$.
   - Faire varier h de 0.1 à 1 par pas de 0.1 et calculer $CV(h)$. Si $CV(h) < CV$, remplacer CV par $CV(h)$ et $h_{\hat{m}}$ par h.
   - Refaire varier h par pas de 0.01 entre $h_{min} - 0.1$ et $h_{min} + 0.1$.
3. Construire l'estimateur $\hat{f}_{\hat{m}}$. Superposer son graphe à celui de la densité théorique.