

# Manifold Learning

## Lab x001 : Simulation et estimation de la densité

Jairo Cugliari

S1 2016–2017

### 1 Simulation par la méthode de rejet

1. Écrire sur **R** la fonction `f_tri(x)` qui calcule la valeur de la densité  $f(x)$  d'une variable aléatoire triangulaire sur  $[0,2]$  pour l'argument  $x$  donné en entrée.
2. Écrire sur **R** la fonction `rejection(f, a, b, M)` qui simule une valeur de la densité  $f$  en utilisant la méthode de rejet. Nous supposons que  $f$  est une densité à support sur le segment  $[a,b]$  et qu'elle est majorée uniformément sur son support par  $M$ .
3. Simuler 1000 réalisations d'une variable aléatoire triangulaire en utilisant la fonction écrite dans le point précédent. Sauvegarder les valeurs simulées dans un fichier de texte (on les utilisera par la suite).

### 2 Estimation de la densité par histogramme

1. Générer un échantillon  $X_1, \dots, X_n$  de taille  $n=100$  où  $X_i \sim \text{Triangulaire}(0,2)$ .
2. Utiliser la fonction `hist` de **R** pour estimer la densité de l'échantillon. Comparer le résultat de l'estimation quand l'on fait varier le nombre de classes  $m$ . En particulier utiliser les valeurs heuristiques :  $m=1+10\log(n)/3$  et  $m=\sqrt{n}$ .

### 3 Estimation de la densité par noyau

1. Générer un échantillon  $X_1, \dots, X_n$  de taille  $n=100$  où  $X_i \sim \text{Triangulaire}(0,2)$ .
2. Utiliser la fonction `density` de **R** pour estimer la densité de l'échantillon. Comparer le résultat de l'estimation quand l'on fait varier la largeur de fenêtre  $h$ . En particulier utiliser les valeurs heuristiques :  $h=1.06\hat{\sigma}n^{-1/5}$  où  $\hat{\sigma}=\min\{\hat{\sigma}, \text{IQR}/1.34\}$ .
3. Écrire sur **R** la fonction `riskkde(obs, h)` qui estime le risque  $J$  de l'estimateur à noyau de la densité avec une largeur de fenêtre  $h$ .
4. Estimer la courbe de risque  $J(h)$ . Comparer la valeur de  $h^*=\text{argmin}J(h)$  avec la valeurs heuristiques pour  $h$ .

### 4 Exercices additionnels

**Old Faithful geiser** Utiliser les données `faithful` du paquet `datasets`. Nous voulons estimer la densité des variables `waiting` et `eruptions`.

1. Utiliser l'estimateur à noyau pour estimer chacune des densités marginales (i.e. de manière individuelle)
2. Il est possible d'estimer la densité conjointe des deux variables. Explorer les options données par les packages `ks` et `hexbin`.

### 5 Lectures recommandés.

- All of Nonparametric Statistics (2006) L. Wasserman, Chapitre 6