

# Devoir 2: Rapport

GIF-7005: Introduction à l'apprentissage machine

*Stéphane Caron*

*17 Octobre 2018*

## Contents

<b>Question 1</b>	<b>1</b>
Partie a . . . . .	1
Partie b . . . . .	1
Partie c . . . . .	3
Partie d . . . . .	3
<b>Question 2</b>	<b>3</b>
Partie a . . . . .	3
Partie b . . . . .	3
Partie c . . . . .	3
Partie d . . . . .	3
Partie e . . . . .	3

## Question 1

Dans cette question, l'objectif est d'estimer de manière non-paramétrique la densité d'une loi de mélange entre deux lois normales.

### Partie a

Dans la première partie, nous estimerons la densité de la loi mélange par la méthode de l'histogramme. La figure 1 illustre la densité de la loi en échantillonnant 50 et 10 000 observations de la loi mélange.

Dans la figure 1, la ligne orange correspond à la courbe de densité réelle de la loi mélange. On remarque qu'avec 50 observations, l'estimation de la densité n'est pas très précise alors qu'avec 10 000 observations, on se rapproche beaucoup plus de la vraie densité.

### Partie b

Dans cette partie, nous estimerons encore la densité de la loi de mélange, mais cette fois-ci avec une estimation par noyau *boxcar*. Cette méthode d'estimation n'est pas continue, mais elle évite de devoir poser une origine alors qu'une fenêtre est appliquée à chaque valeur du support. La figure 2 illustre l'estimation de la densité avec 50 et 10 000 observations selon différentes valeurs de *bandwidth* ( $h$ ). Cette dernière valeur, contrôle la largeur de la fenêtre autour de laquelle nous allons considérer les données autour d'un point  $x$  du support.

À partir de la figure 2, on peut remarquer 2 choses. Premièrement, moins il y a d'observations, plus il y a de variations dans les estimations de la densité pour les valeurs du support. Cela se voit par les escaliers plus prononcés dans la figure avec 50 observations seulement. Deuxièmement, plus la fenêtre d'estimation (*bandwidth*) est grande, plus la densité est constante sur le support. Cela fait du sens puisqu'avec une grande fenêtre, on considère beaucoup de données pour estimer la densité en un point. Si la fenêtre est petite, on donne plus d'importance à la densité locale ce qui fait en sorte qu'on remarque davantage les deux cloches normales.

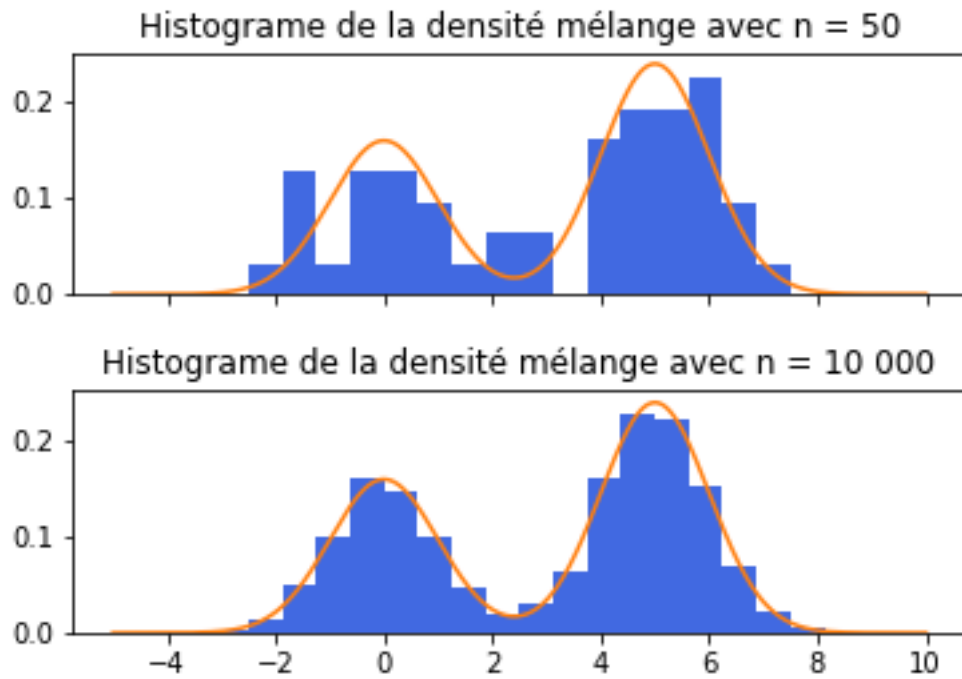


Figure 1: Estimation de la densité par histogramme avec 50 et 10 000 observations.

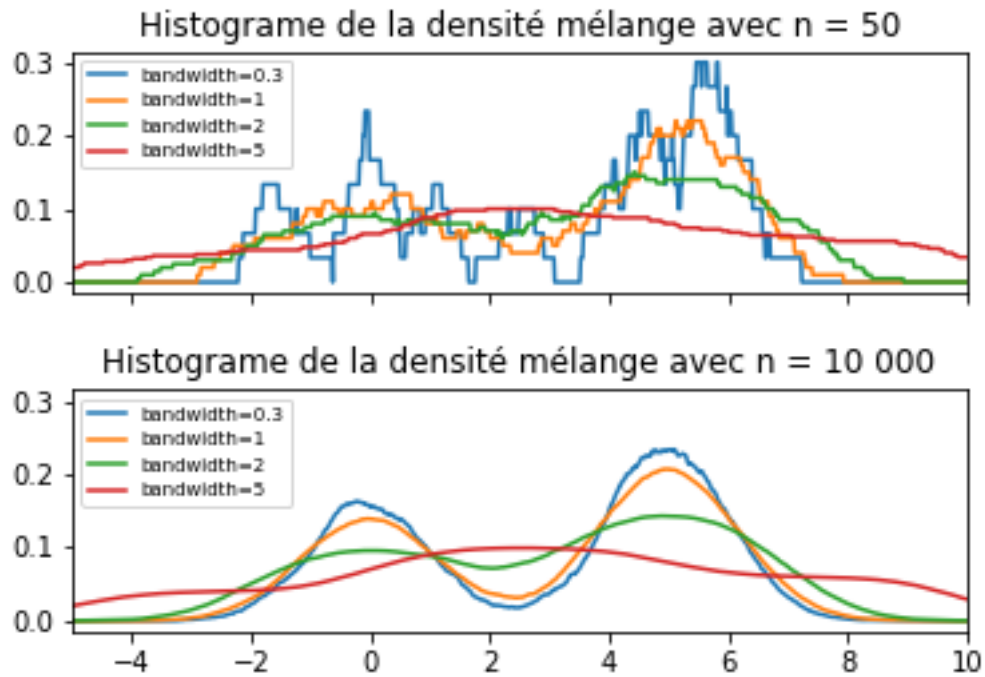


Figure 2: Estimation de la densité par noyau *boxcar* avec 50 et 10 000 observations selon différentes valeurs de *bandwidth*.

Partie c

Partie d

## Question 2

Partie a

Partie b

Partie c

Partie d

Partie e