

## SY19

### TP 5: Splines, modèles additifs généralisés

#### Exercice 1

Les données `mcycle` du package `MASS` contiennent des mesures d'accélération de la tête d'un mannequin en fonction du temps, au cours d'une simulation d'accident de moto.

1. Estimer la valeur moyenne de l'accélération en fonction du temps par régression polynomiale d'ordre  $p$ . Représenter graphiquement les données et la fonction de régression estimée pour différentes valeurs de  $p$ . Quelles valeurs semblent visuellement satisfaisantes ?
2. Déterminer la valeur optimale de  $p$  par validation croisée (faire  $K = 5$  groupes).
3. Mêmes questions en utilisant des splines cubiques naturelles. On expérimentera cette fois différents nombres de degrés de liberté (paramètre `df` de la fonction `ns()`), puis on déterminera la valeur optimale par validation croisée.
4. Utiliser cette fois les splines de lissage (fonction `smooth.spline()`). Trouver la valeur optimale du paramètre `df` par la méthode leave-one-out, puis faire varier ce paramètre autour de la valeur optimale, et estimer les erreurs par validation croisée avec le même découpage que dans les deux questions précédentes.
5. Représentez sur un même graphique les fonctions de régression estimées par les trois méthodes. Laquelle donne les meilleurs résultats sur ces données ? Justifiez votre réponse.

#### Exercice 2

On considère à nouveau les données `prostate`, en prenant la variable `lpsa` comme variable à expliquer. Sélectionner un modèle linéaire avec un sous-ensemble de variables explicatives, en utilisant le critère BIC. Pour ces variables, ajuster plusieurs modèles additifs généralisés, en utilisant des splines naturelles ou des splines de lissage. Représenter graphiquement les contributions de chaque variable explicative. Sélectionnez le meilleur modèle par validation croisée.