# Smoothing Splines: introduction

Smooting splines is a type of nonparametric regression in which we fit a smooth curve to our set of data points. This techniques finds a balance between smoothness of the curve and a perfect fit in the extreme case for which we would have (almost) as many piecewise polynomials as observations. Suppose that we observe a set of $n$ pairs $(x_i, y_i), i = 1, ..., n$. The relationship is of the form

$$y_i = f(x_i) + \epsilon_i$$

where $f(.)$ is some 'smooth' function that we can estimate by

$$\hat{f} = \frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda \int [f''(x)]^2 dx$$

It is some kind of penalized least squares regression problem where $\lambda$ is a 'smoothness' parameter and $\int [f''(x)]^2 dx$ is a 'roughness' penalty.

# 'mcycle ' dataset

**mcycle**: A data frame of 133 observations x 2 variables, giving a series of measurements of head acceleration in a simulated motorcycle accident, used to test crash helmets

**times**: in milliseconds after impact
**accel**: in g.

```
1 > head(mcycle)
2   times accel
3 1   2.4   0.0
4 2   2.6  -1.3
5 3   3.2  -2.7
6 4   3.6   0.0
7 5   4.0  -2.7
8 6   6.2  -2.7
```

**Source**: Silverman, B. W. (1985) Some aspects of the spline smoothing approach to non-parametric curve fitting. Journal of the Royal Statistical Society series B 47, 1–52.

**Reference**: Venables, W. N. and Ripley, B. D. (2002) Modern Applied Statistics with S-PLUS. Fourth Edition. Springer

# Summary

Here is the minimal code to run Smooting Splines regression in R.

```
1 > # fit smoothing splines model (ss) with default number of knots
2 > modss = with(mcycle, ss(times, accel))
3 > summary(modss)
4 ..
5 Residuals:
6      Min       1Q   Median       3Q      Max
7 -76.7951 -12.5654  -0.8346  12.5823  50.5576
8
9 Approx. Signif. of Parametric Effects:
10             Estimate Std. Error  t value  Pr(>|t|)
11 (Intercept)  -14.234      2.313   -6.154 1.018e-08 ***
12 x              9.549     21.603    0.442 6.593e-01
13 ..
14 Approx. Signif. of Nonparametric Effects:
15             Df Sum Sq Mean Sq F value Pr(>F)
16 s(x)     10.21 210130 20585.3   40.08      0 ***
17 Residuals 120.79  62035   513.6
18 ..
19
20 > # using smooth.spline function, then plotting
21 > mod1ss = with(mcycle, smooth.spline(times, accel))
22 > fit = data.frame(times = mod1ss$x, accel = mod1ss$y)
23 > head(fit)
24   times     accel
25 1   2.4 -1.373830
26 2   2.6 -1.435130
```

# Plot of the fitted model



Smoothing Splines: Time vs Acceleration
*mcycel dataset*
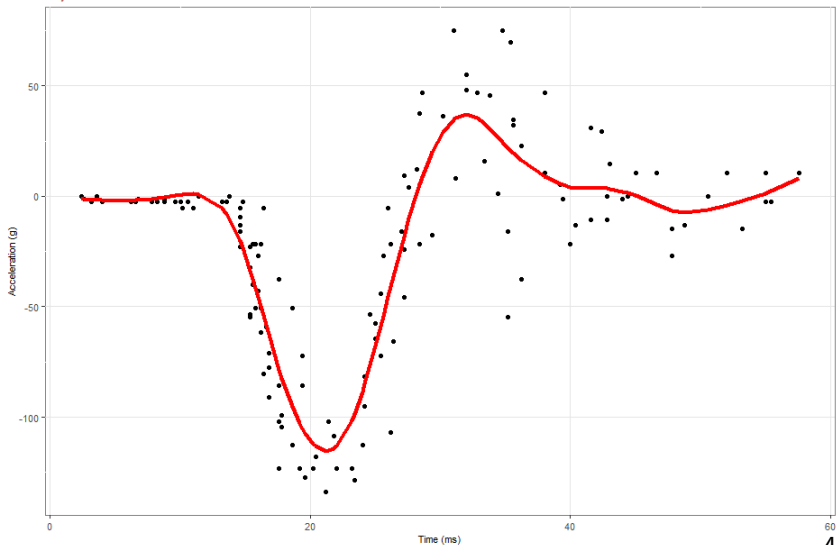
# Main observations

- The smoothing spline reveals a sharp increase in acceleration after 20 milliseconds, indicating a rapid change in speed.

- Following the initial spike, the acceleration decreases and shows a fluctuating pattern over time, suggesting variable speed changes during the motorcycle ride.

- The fitted spline captures several local maxima and minima, highlighting the periods of acceleration and deceleration throughout the observed time span.

- Towards the latter part of the time series, the smoothing spline indicates a gradual stabilization of acceleration, suggesting that the motorcycle's speed changes become less extreme.

# References

Faraway, J. J., Extending the Linear Model (2006), ISBN 0-203-62105-0 (e-book)

Venables, W. N. and Ripley, B. D. (2002) Modern Applied Statistics with S-PLUS. Fourth Edition. Springer

The R Project for Statistical Computing:
https://www.r-project.org/