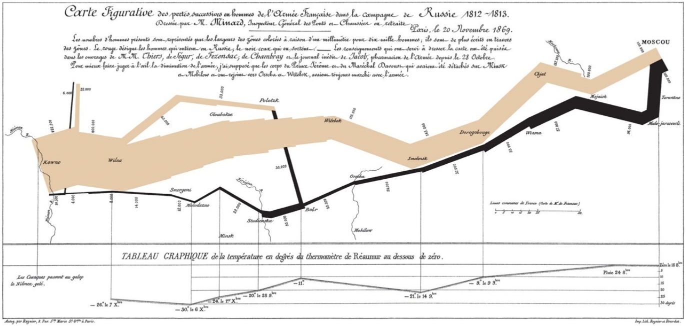
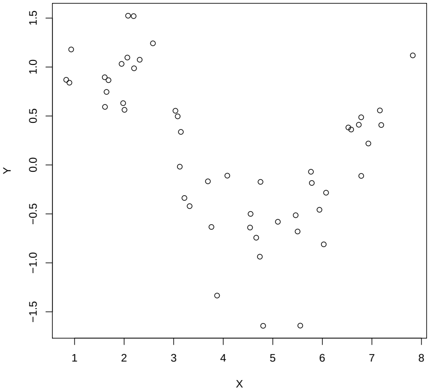
Splines – linear non-linearity

September 9, 2019

“A picture is worth a thousand words”



A picture



Try to sketch a denoised relation between X and Y.

Noisy sine function

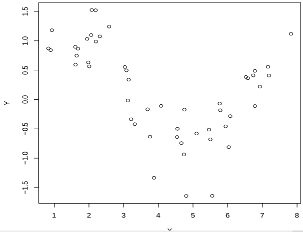
- Let us consider the following non-linear regression model

non-linear regression

$$Y = f(X) + \epsilon$$

where X is an explanatory variable, ϵ is a noisy error and Y is an outcome variable (aka response or dependent variable).

- The model is non-linear when $f(X)$ is not a linear function of X. Consider for example $f(X) = \sin(X)$.
- A sample from such a model



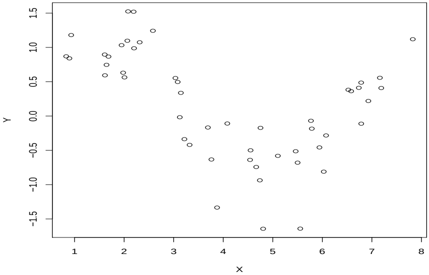
Noisy Sine R-code

```
#Non-linear regression

X=runif(50,0.5,8)
e=rnorm(50,0,0.35)
Y=sin(X)+e

pdf("NoisySine.pdf") #Save a graph to a file
plot(X,Y)
dev.off()             #Closes the graph file
```

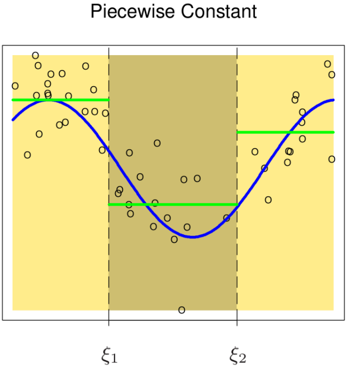
How to (re-)discover a non-linear relation



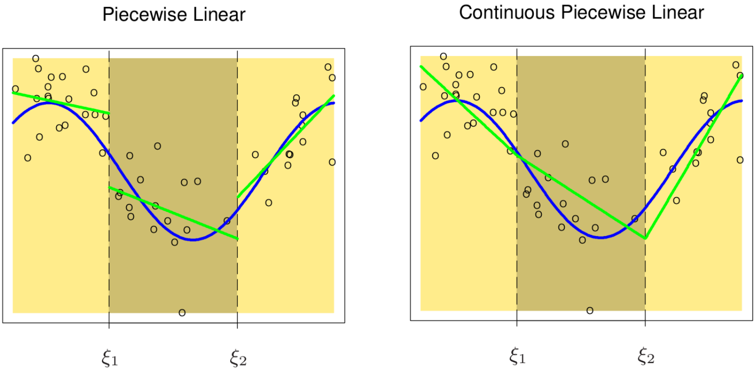
- We are now interested to recover from the above data the relation that stands behind them?
- In practice we do not know that there is any specific function (in this case *sine function*) involved.
- We clearly see that the relation is non-linear.
- We want a standardized and automatic approach.
- **Any ideas?**

Piecewise constant

- We first divide the domain into disjoint regions marked by the **knot points** $\xi_0 < \xi_1 < \cdots < \xi_n < \xi_{n+1}$.
- ξ_0 the begining of the x-interval and ξ_{n+1} its end
- On each interval we can fit independently.
- For example by constant functions



Piecewise linear



- Where the difference between the two pictures lies?
- The second is continuous – a linear **spline**.
- Fit is **no longer independent** between regions.
- How to do it?

Analysis of the problem

- How many parameters there are in the problem?
- **3-intercepts $+$ 3-slopes $-$ 2-knots $=$ 4**
  (we subtract knots because each knot sets one equation to fulfill the continuity assumption)
- The problem should be fitted with four parameters.

From now on we assume the knots locations are decided for and not changing.

Making non-linear linear

- What is the minimal number of vectors needed to express linearly any vector in 4 dimensions? **4**
- Such vectors are (linearly) independent (none is linearly expressed by the remaining ones)
- Find 4 piecewise linear continuous functions that are 'independent', say $h_1(X)$, $h_2(X)$, $h_3(X)$, $h_4(X)$.
- Then any function piecewise linear with the given knots can be written linearly by them

$$f(X) = \beta_1 h_1(X) + \beta_2 h_2(X) + \beta_3 h_3(X) + \beta_4 h_4(X) = \sum_{j=1}^{4} \beta_j h_j(X).$$

- $f(X)$ is continuous in X because each of $h_j(X)$ is.
- There are four parameters, so that any continuous piecewise linear function should be fitted by proper choice of β_j's.

Basis functions

- There many choices for h_j, $j = 1, \ldots, 4$.
- The following is a natural one

$$h_1(X) = 1,\ h_2(X) = X,\ h_3(X) = (X - \xi_1)_+,\ h_4(X) = (X - \xi_2)_+,$$

  where t_+ is a positive part of a real number t.
- The model for the data

$$Y_i = \beta_1 h_{i1} + \cdots + \beta_r h_{ir} + \varepsilon_i,$$

  $i = 1, 2, \ldots, n$, where $h_{ij} = h_j(X_i)$.
- The model in the matrix notation

$$\mathbf{Y} = \mathbf{H}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

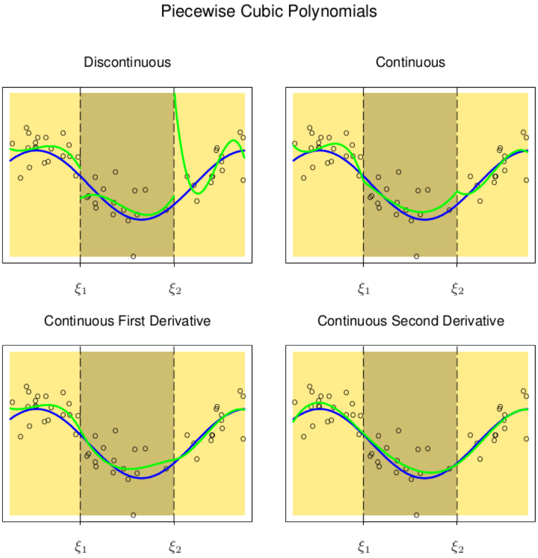
where **H** is the matrix of h_{ij}'s.

- Fitting problem is solved by fitting the **linear regression problem** (the least squares method).

Extension to smoother version – cubic splines

- The piecewise linear splines have discontinuous derivatives at knots. **Why?**
- We can increase the order of **smoothness at the knots** by increasing the degree of polynomial that is fitted in each region and then imposing the **continuity constraints at each knot**.
- The cubic splines are quite popular for this purpose.

Illustration – cubic splines



Basis

- Let us count the number of parameters needed.
  - Number of parameter of a cubic polynomial is: **4**
  - Number of knots is **4** so we have **3** polynomials (we count the right and the left point of the abscissa's range)
  - The number of knots where the smoothness constraints are imposed: **2**
  - The number of constraints at a knot to have smooth second derivative: **3** (the equations for continuity of the functions and their two derivatives)

  Number of the parameters:

$$3 * 4 - 2 * 3 = 6$$

- Example of the (functional) spline basis

$$h_1(X) = 1,\ h_2(X) = X,\ h_3(X) = X^2,\ h_4(X) = X^3,$$
$$h_5(X) = (X - \xi_1)^3_+,\ h_6(X) = (X - \xi_2)^3_+$$

Another Basis – B-splines

- There are convenient splines that can be defined recursively called B-splines.
- We consider only the special case of cubic B-splines (see the textbooks for more general discussion, notation here is slightly changed).

Cubic spline = piecewise cubic with the derivative up to the second order are continuous

- Assume $\xi_1, \dots, \xi_K$ internal knots and two endpoints ξ_0 and ξ_{K+1}.
- Add three *artificial* knots that are equal to ξ_0 and similarly additional three knots that are equal to ξ_{K+1} for the total of $K+8$ knots that from now on are denoted by τ_i, $i = 1, \dots, K+8$.
- Define recursively functions $B_{i,m}$ that are splines of the mth order of smoothness, $i = 1, \dots, K+8$, $m = 0, \dots, 3$
- the 0-order of smoothness is discontinuity at the knots, the first order is continuity of function, the second order is continuity of the first derivative, etc

Recursion

- For the knots τ_i, $i = 1, \dots, K+8$ we define $B_{i,m}$, $i = 1, \dots, K+8$, $m = 0, \dots, 3$
- The piecewise constant (0-smooth), $i = 1, \dots, K+7$,

$$B_{i,0}(x) = \begin{cases} 1 & \text{if } \tau_i \leq x < \tau_{i+1} \\ 0 & \text{otherwise} \end{cases}$$

- Higher (m) order of smoothness , $i = 1, \dots, K+8-m$,

$$B_{i,m}(x) = \frac{x - \tau_i}{\tau_{i+m-1} - \tau_i} B_{i,m-1}(x) + \frac{\tau_{i+m} - x}{\tau_{i+m} - \tau_{i+1}} B_{i+1,m-1}(x).$$

- $B_{i,3}$ are cubic order splines that constitutes basis for all cubic splines.

Illustration – evenly distributed knots

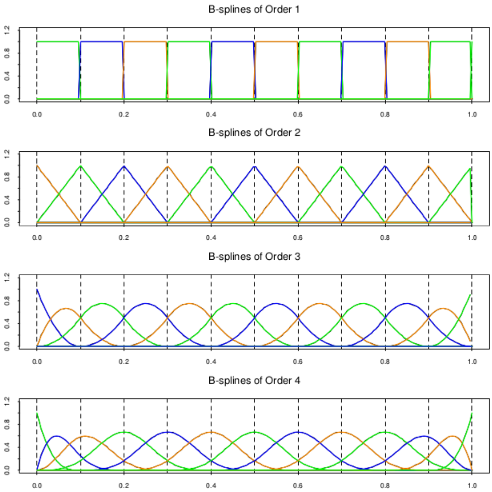
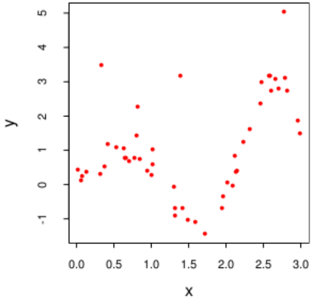
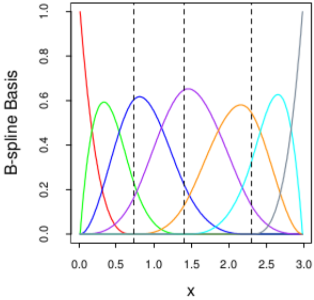


Illustration – non-evenly distributed knots

- Another data set and B-spline basis





Splines without knot selection

The regression problem with one predictor

$$y = \alpha + f(x) + \epsilon.$$

- The **maximal set of knots**: a knot is located at each abscissa location in the data.
- Clearly, without additional restrictions this leads to overfitting and non-identifiability. **Why?**
- These issues are taken care of since irregularity is penalized.
- Outside the range of predictors it is estimated by a linear function (smoothing on the boundaries).

Penalty for being non-smooth

- Minimize the penalized residual sum of squares

$$PRSS(f,\lambda) = \sum_{i=1}^{N} (y_i - f(x_i))^2 + \lambda \int f''(t)^2 dt$$

- $\lambda = 0$: any fit that interpolates data exactly.
- $\lambda = \infty$: the least square fit (second derivative is zero)
- We fit by the cubic splines with knots set at all the values of x's and the solution has the form

$$f(x) = \sum_{j=1}^{N+4} \gamma_j B_j(x), \tag{1}$$

where γ_j's have to be found.

B-spline basis

- The splines $B_j(x)$, $j = 1, \ldots, N+4$, are used in the smoothing splines, where the initial x_i, $i = 1, \ldots, N$ are augmented by 2 end points defining the range of interest for the total of $N+2$ knots.
- We have seen that if there is N internal points, then there have to be $N+4$ of the third order splines that are independent in order for them to constitute a basis. **Do the count!**
- One can compute explicitly the coefficients of the following matrix

$$\boldsymbol{\Omega}_B = \left[\int B_i''(t) B_j''(t) \, dt \right]$$

Solution

- The solution has the following explicit form

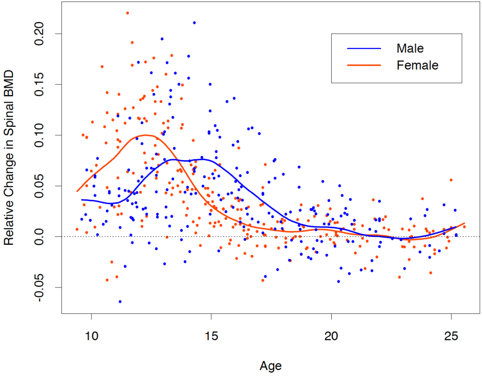
$$\hat{\gamma} = \left(\mathbf{B}^T\mathbf{B} + \lambda\mathbf{\Omega}_B\right)^{-1}\mathbf{B}^T\mathbf{y},$$

  where

$$\mathbf{\Omega}_B = \left[\int B_i''(t)B_j''(t)\ dt\right]$$

- To see this substitute (1) to the PRSS – it becomes a regular least squares problem that is solved by $\hat{\gamma}$.
- Further details in Assignment 2.

Example – bone mineral density



The response is the relative change in bone mineral density measured at the spine in adolescents, as a function of age. A separate smoothing spline was fit to the males and females, with $\lambda = 0.00022$. It can be argued that this choice of λ corresponds to about 12 degrees of freedom (the number of parameters in a comparable standard spline fit of the solution). See the textbook for the discussion of transformation from the degrees of freedom to λ and vice versa.