Semiparametric smoothing methods

- **Assumptions:** Only one observation y_i per subject at time point t_i.

- Data are thus of the form

$$(t_i, y_i), \quad i = 1, \ldots, N.$$

- **Goal:** Estimation of the unknown mean curve $\mu(t)$ in the model

$$Y_i = \mu(t_i) + \epsilon_i,$$

where the ϵ_i are independent with mean 0.

Kernel methods: "Sliding window"

- Consider a window around time point t_1.
- Let $\widehat{\mu}(t_1)$ be the average of all y_i corresponding to t_i in that window.
- Analogously for $\widehat{\mu}(t_2), \widehat{\mu}(t_3), \ldots$.

$\longrightarrow$ Sliding window for the estimation.

Kernel methods: “Sliding window”

The width of the window is important:

- If the width is chosen very small, the window can include only one observation at the one extreme $\rightarrow$ interpolation instead of smoothing!
- If the width is chosen very wide, the window can include all observations at the other extreme. This yields a constant:

$$\widehat{\mu}(t) = \frac{1}{N} \sum_{i=1}^{N} y_i.$$

Kernel methods in general

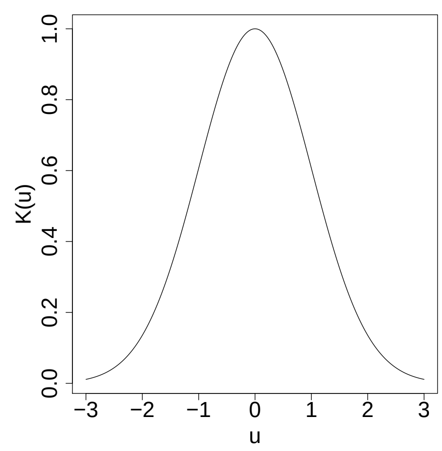
- With the sliding window method, each observation gets the weight 1 ("in the window") oder 0 ("outside the window").

- This method is a special case of kernel smoothing methods.

- More generally, choose a smooth weight function that gives more weight to observations nearer in time than to observations further away.

- Common choice: Gaussian kernel

$$K(u) = \exp(-0.5u^2).$$



Kernel methods in general

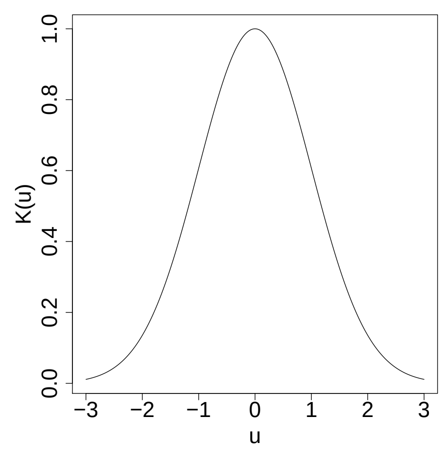
- With the sliding window method, each observation gets the weight 1 ("in the window") oder 0 ("outside the window").

- This method is a special case of kernel smoothing methods.

- More generally, choose a smooth weight function that gives more weight to observations nearer in time than to observations further away.

- Common choice: Gaussian kernel

$$K(u) = \exp(-0.5u^2).$$



Kernel methods in general

- Definition of the kernel estimator:

$$\widehat{\mu}(t) = \sum_{i=1}^{N} \frac{w(t, t_i, h)}{\sum_{i=1}^{N} w(t, t_i, h)} y_i,$$

where $w(t, t_i, h) = K\left((t - t_i)/h\right)$ are the weights and h is the bandwidth.

- Larger values for h yield smoother curves.

- We'll discuss the choice of h in a few slides.

- How is the kernel K defined for the sliding window method?

Smoothing splines (Silverman, 1985)

- If we assume $\mu(t)$ can be well approximated by a twice continuous differentiable function $s(t)$ with second derivative $s''(t)$, consider minimizing

$$J(\lambda) = \sum_{i=1}^{N} (y_i - s(t_i))^2 + \lambda \int \{s''(t)\}^2 dt.$$

- The solution can be shown to be a natural cubic **spline** (a two times differentiable function consisting of piecewise cubic polynomials) with knots at the t_i and can be obtained from (relatively simple) linear equations.

- Penalized splines are an alternative that is computationally less demanding and can be incorporate into more complex models, see Chapter 6.2.

Lo(w)ess smoothing (Cleveland, 1979)

- LOWESS $=$ **LO**cally **WE**ighted regression **S**catterplot **S**moothing
- Function `lowess` in R
- Lo(w)ess can be seen as an extension of kernel methods: at each point t_i, a local polynomial regression is fitted using weighted least squares, giving more weight to observations closer by.
- There is an iterative version that is more robust to outliers, giving them smaller weight.

Choice of smoothing parameters

- In all three approaches (kernel, splines, lowess), the smoothness of the estimated curves is controlled by one **smoothing parameter** (e.g. h, λ). This parameter is typically chosen to optimize a criterion.

- **Goal:** compromise between bias and variance.

- A common criterion that combines bias and variance is the mean squared error, MSE (analogously for h instead of λ):

$$MSE(\lambda) = \frac{1}{N} \sum_{i=1}^{N} \{y_i^* - \widehat{\mu}(t_i; \lambda)\}^2,$$

where y_i^* is a new observation at time point t_i.

Choice of smoothing parameters

$$MSE(\lambda) = \frac{1}{N}\sum_{i=1}^{N}\{y_i^* - \widehat{\mu}(t_i;\lambda)\}^2$$

Observations y_i which were used for estimation of μ should not be compared to $\widehat{\mu}(t_i)$: This would lead to always choosing the smallest band width h or penalty λ and to interpolation instead of a smooth curve (overfitting).

Solution: cross-validation (analogously for h instead of λ)

$$CV(\lambda) = \frac{1}{N}\sum_{i=1}^{N}\{y_i - \widehat{\mu}^{-i}(t_i;\lambda)\}^2,$$

where $\widehat{\mu}^{-i}(t_i;\lambda)$ is obtained without observation i. See Chapter 6.2 for mixed model-based estimation of smoothing parameters.

Note

- Please note that these smoothing methods (and the criterion for the choice of the smoothing parameter) assume independent and identically distributed (i.i.d.) errors.
- Also, dropout and missing values are not taken into account.
- They can still be useful **exploratory** tools.
- Example CD4 data: See lab.
- For how to incorporate smooth mean functions in mixed models accounting for repeated measurements, please see Chapter 6.2.