

The Joys of P-splines

The Craft of Smoothing

Paul Eilers & Brian Marx

.

Preface

In the course “The Joys of P-splines (The Craft of Smoothing),” we describe in detail the basics and use of P-splines, a combination of regression on a B-spline basis and difference penalties (on the B-spline coefficients).

Our approach is practical. We see smoothing as an everyday tool for data analysis and statistics. We emphasize the use of modern software and we provide functions for R. We follow the text:

Paul H.C. Eilers and Brian D. Marx (2021) *Practical Smoothing: The Joys of P-splines* Cambridge University Press.

There will be six main sessions:

- Session 1 presents the idea of bases for regression. It will show why global bases, like power functions or orthogonal polynomials are ineffective and why local bases (gaussian bell- shaped curves or B-splines) are attractive.
- In Session 2 penalties are introduced, as a tool to give complete and easy control over smoothness. The combination of B-splines and difference penalties will be studied for smoothing, interpolation and extrapolation.
- In the first two sessions the data are assumed to be normally distributed around a smooth curve. In session 3 we extend P-splines to non-normal data, like counts or a binomial response. The penalized regression framework makes it straightforward to transplant most ideas from generalized linear models to P-spline smoothing. Important applications are density estimation and variance smoothing.
- Any smoothing method has to balance fidelity to the data and smoothness of the fitted curve. An optimal balance can be found by cross-validation or AIC. This subject is studied in Session 4, as well as the computation of error bands of an estimated curve. We also show how optimal smoothing performs on simulated data, to give you confidence in that it makes the right choices.
- In the first four sessions we only consider one-dimensional smoothing. When there are multiple explanatory variables, we can use generalized additive models, varying-coefficient models, or combinations of them. Tensor products of B-splines and multi-dimensional

difference penalties make an excellent tool for smoothing in two (or more) dimensions. This is the subject of session 5.

- The sixth session looks at the use of P-splines in regression problems with very many variables, which are ordered, like in optical spectra. In the chemometric literature this is known as multivariate calibration.
- Time permitting, an seventh session will be provided, which briefly presents some alternative approaches to P-splines, including mixed model and Bayesian formulations. Some alternatives for optimizing the tuning parameters, as well as some technical details for computation are also given.

In addition there will be two computer lab sessions (with four exercises), in which R software will be used to solve a number of smoothing problems. One session will concentrate on simple functions with limited goals. This will improve your understanding of what is going on “under the hood.” The other session will apply smoothing to the generalized linear model and density estimation. A third lab will be provided that uses the `mgcv` package, written by Simon Wood, a large but powerful tool that can handle a variety of situations. Finally, the fourth lab will explore 2D (generalized) linear smoothing.

We very much hope that you will find this course useful, interesting, and enjoyable.

Brian Marx (bmarx@lsu.edu)

Paul Eilers (p.eilers@erasmusmc.nl)