

Quantile sheets

Sabine K. Schnabel^{1,2} and Paul H.C. Eilers^{1,3}

¹ Biometris, Wageningen University and Research Centre, The Netherlands

² Centre for BioSystems Genomics, The Netherlands

³ Department of Biostatistics, Erasmus Medical Center Rotterdam, The Netherlands

Email: sabine.schnabel@wur.nl

Abstract

Quantile regression is a popular tool for modelling the conditional quantiles of a response as a function of one (or more) independent variables. Quantile curves for different probabilities often cross. In theory this is not possible, but it is frequently encountered in small data sets. We propose to estimate all quantile curves simultaneously with a surface, a so-called quantile sheet. This sheet is fitted with a sum of tensor products of B -splines with two penalties to control for smoothness.

Keywords: Quantiles, smoothing, tensor product, P -splines

AMS subject classifications: 65D10, 93E14

1. Introduction and Motivation

Smoothing has become an important topic in the field of quantile regression [3]. This method can also be used to study the behaviour of the conditional distribution over a covariate such as time or age. In recent years many approaches and software packages have been published for quantile regression. Implementations include the R packages `quantreg`, `cobs` and `VGAM`.

By definition, quantile curves of different probabilities cannot cross, but in practice we encounter this phenomenon frequently. It appears more often in small data sets, but is not limited to these situations. The crossing of curves may influence further analysis, e.g. when it comes to studying the conditional distribution. In the literature one can find a number of different proposals to prevent these crossings. Restricted regression quantiles were one of the first suggestions to avoid crossing curves. They form a flexible framework with good computational properties. Later, different kernel estimators were proposed to estimate non-crossing quantile curves. Other non-parametric techniques were also suggested in recent publications. Further approaches to avoid crossing of quantile curves include the combination of additional constraints with the problem as well as natural monotoneization.

We propose an alternative approach. The basic idea introduces a surface over a two-dimensional domain of the covariate x and the probability τ . This surface is called a *quantile sheet*. Cutting the surface at a fixed probability results in a smooth quantile curve. All quantile curves are estimated

simultaneously. Therefore the crossing problem disappears if the sheet is monotonically increasing in probability. The sheet is constructed as a sum of tensor product B -splines and we use penalties to control the smoothness in both x - and τ -direction. For the estimation iteratively weighted least squares is used in combination with computationally advantageous array methods.

2. Estimation of quantile curves using a quantile sheet

A quantile sheet is based on a sum of tensor product B -splines [2]. Following the philosophy of P -splines, a large number of tensor products is used. This may lead to over-fitting, but we control the smoothness of the surface by two difference penalties on the model coefficients. One smoothing parameter will control the smoothness in x -direction while the second parameter governs the smoothness (and monotonicity) in direction of the probability τ . Both smoothing parameters are chosen via an adapted cross-validation criterion. While most quantile estimation methods are based on linear programming, we are using iteratively reweighted least squares [4]. In addition we apply array algorithms for multi-dimensional P -spline fitting [1] for a fast computational implementation. Details on the theoretical derivation of our proposed methods can be found in [5].

3. Applications and outlook

Quantile regression is used in almost any area of application when one aims at describing properties of the underlying (unknown) distribution of the data. In public health, quantiles are often used as references e.g. for the growth of children. As recommendations to parents and more importantly decisions of the practitioners are often based on these reference charts, reliable estimation of the quantile curves is of special importance.

Quantile sheets are a powerful tool to estimate non-crossing smooth quantile curves. Derivatives of the smooth surface are smooth, too, and piecewise quadratic. These properties can be exploited for further analysis. More covariates can be included in an additive variant of the model. This will be investigated in future research.

References

1. Currie, I.D., Durban M. and P.H.C. Eilers (2006). Generalized linear array models with applications to multi-dimensional smoothing. *Journal of the Royal Statistical Society, Series B* 68, 259–280.
2. Eilers, P.H.C., Currie I.D. and Durban M. (2006). Fast and compact smoothing on large multidimensional grids. *Computational Statistics & Data Analysis* 50, 61–76.
3. Koenker R. (2005). *Quantile Regression*. Cambridge University Press.
4. Schlossmacher, E. J. (1973). An iterative technique for absolute deviations curve fitting. *Journal of the American Statistical Association* 68, 857–859.
5. Schnabel S.K. and Eilers P.H.C. (2012). Simultaneous estimation of quantile curves using quantile sheets. *Under review*.