

ADAPTIVE SPLINES FOR CONTINUOUS FEATURES IN RISK ASSESSMENT

Ndeye Arame Seck, Michel Denuit

LIDAM Discussion Paper ISBA
2021 / 35

ISBA

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : lidam-library@uclouvain.be

<https://uclouvain.be/en/research-institutes/lidam/isba/publication.html>

ADAPTIVE SPLINES FOR CONTINUOUS FEATURES IN RISK ASSESSMENT

NDEYE ARAME SECK and MICHEL DENUIT

Institute of Statistics, Biostatistics and Actuarial Science - ISBA

Louvain Institute of Data Analysis and Modeling - LIDAM

UCLouvain

B-1348 Louvain-la-Neuve, Belgium

September 15, 2021

Abstract

Number and location of knots strongly impact on fitted values obtained from spline regression methods. P-splines have been proposed to solve this problem by adding a smoothness penalty to the log-likelihood. This paper aims to demonstrate the strong potential of A-splines (for *adaptive splines*) proposed by Goepp et al. (2018) for dealing with continuous risk features in insurance studies. Adaptive ridge is used to remove the un-necessary knots from a large number of candidate knots, yielding a sparse model with high interpretability. Two applications are proposed to illustrate the performances of A-splines. First, death probabilities are graduated in a Binomial regression model. Second, continuous risk factors are included in a Poisson regression model for claim counts in motor insurance. The move from technical to commercial price list can easily be achieved by switching to A-splines of degree 0, i.e. piecewise constant functions.

Keywords: Generalized Additive Models, Penalized Likelihood, Adaptive Ridge, Banding.

1 Introduction and motivation

In insurance studies, splines are often used to account for the effect of continuous features on expected claim frequencies or severities in general insurance, or on mortality or morbidity rates in life and health insurance. However, outputs can drastically change if there are too many or too few knots, or if the knots are not located appropriately. To address this issue, penalties are widely used in spline regression after the seminal paper by Eilers and Marx (1996). With their proposed P-splines, users allow for a large number of knots and then let the penalty control the degree of smoothness of the fit.

P-splines have been successfully applied in insurance studies under Generalized Additive Models (GAMs). While they deliver accurate estimations, they generally fail to provide analysts with a transparent and interpretable model because every single spline resulting from the large number of knots supplied by the user plays a role in the final model: the smoothing parameter in P-splines indeed constraints the coefficients of adjacent splines to be close, reducing model complexity, but no coefficient is set to zero so that all splines are kept as explanatory variables.

Goepp et al. (2018) proposed an alternative approach called adaptive splines, or A-splines to remedy this deficiency. As opposed to P-splines, A-splines consist in iteratively removing the un-necessary knots by adaptive ridge, resulting in a sparse regression model that often appears to be easily interpretable. This short note aims to demonstrate the great potential of A-splines by revisiting two classical applications: the graduation of death probabilities with Binomial GAM and the estimation of expected claim frequencies with Poisson GAM.

Besides classical cubic splines, A-splines also work with splines of degree 0, that is, piecewise constant functions and thus allow to perform automatic detection of breakpoints. In insurance studies, this allows the actuary to transform a continuous feature into a categorical one in order to simplify the tariff structure. The great conceptual advantage is that the analyst remains working with GAMs, substituting splines of degree 0 (piecewise constant functions) to cubic splines (piecewise polynomial functions of third degree) for banding continuous features.

The remainder of the text is organized as follows. Section 2 briefly recalls P-splines, before presenting the new approach based on A-splines. Numerical examples are proposed in Section 3, before the final discussion in Section 4.

2 Adaptive splines in a nutshell

2.1 B-splines

A spline is a piecewise polynomial function that obeys continuity conditions on its value and its first derivatives at the points, called knots, where the pieces join. B-splines of degree q are piecewise polynomials of degree q with local support, with continuous derivatives at a pre-defined set of knots. Cubic splines ($q = 3$) are generally used in insurance studies.

2.2 P-splines

The key idea behind P-splines is to allow for a large number of equally-spaced knots to ensure enough flexibility and to supplement the log-likelihood with a roughness penalty based on the sum of squared second-order differences of the regression coefficients associated to adjacent splines. The role of the penalty is to guarantee sufficient smoothness of the fitted curve. This leads to a penalized likelihood approach where the objective function is a compromise between goodness-of-fit and smoothness.

This approach is very effective in insurance studies and is now widely used in practice. See for instance Chapter 6 in Denuit et al. (2019). The only weakness is its lack of interpretability since all splines are involved in the fitted values which thus remain unstructured.

2.3 A-splines

As P-splines, A-splines also allow for a large number of initial knots. A penalized likelihood procedure called adaptive ridge is then applied to get a sparse model with a small number of knots. Sparsity is often obtained with the help of Lasso-type, or L_0 penalties setting some regression coefficients to 0. This results in a reduced set of knots in the final estimate. Goepf et al. (2018) noticed that if some finite difference of order $q + 1$ of the spline coefficients is equal to 0 then the corresponding knot is not needed and the same fit can be obtained with a reduced set of knots. These authors thus proposed a L_0 penalty on these differences. However, such a penalty is not differentiable and Goepf et al. (2018) approximated it with the help of an iterative procedure called adaptive ridge. The penalty term then becomes a weighted sum of squared finite differences of order $q + 1$ where the weights are iteratively computed from the previous values of the regression coefficient.

3 Numerical illustrations

3.1 Graduation of death rates

Mortality data used here come from the Belgian Institute of Statistics; they can be freely downloaded from `statbel.fgov.be`. Responses under consideration are observed yearly numbers of deaths D_x among males aged x in 0-102 together with initial exposures L_x . We work with Binomial regression, assuming that $D_x \sim \text{Binomial}(L_x, q_x)$. Death probabilities q_x are left unspecified and estimated with cubic splines ($q = 3$). We start with 93 internal knots, i.e one knot at each age from 1 to 93 years.

The resulting fit is displayed in the upper panel of Figure 1 for Belgian males in calendar year 2019. The R package **Aspline** contributed by Goepf et al. (2018) has been used to produce the results presented here. To tune the smoothing parameter and for the sake of comparisons with P-splines, we favored cross validation over the criteria included in **Aspline** (BIC, AIC and extended BIC). Adaptive ridge results in 5 knots at ages 3, 18, 32, 66, and 90 that are easy to interpret. The first knot at age 3 accounts for perinatal mortality. The second and third knots at ages 18 and 32 capture the accident hump. The fourth knot at age 66 recognizes the mortality just after retirement while the last knot at age 90 creates the inflection point where death probabilities switch from convex to concave pattern.

Compared to the fit using P-splines (performed with the function `gam` in the package `mgcv`) that is displayed in the lower panel of Figure 1, A-splines and P-splines closely agree about the underlying mortality curve except at ages 6-10. The former only uses 5 knots whereas the latter produces a fit involving all 93 knots with an effective degree of freedom equal to 18.6.

3.2 Frequency modeling in motor insurance

Here, we use the dataset `freMTPLfreq` from the package `CASdatasets`. The response is the observed number of claims. In addition to categorical features, we consider policyholder's age (`DriverAge` ranging from 18 to 92) and vehicle's age (`CarAge` ranging from 0 to 22) as continuous features. For the sake of illustration, we restrict to main effects and do not explore interactions (such as the classical age-gender one). Since `Aspline` currently works only with a single continuous feature, we have implemented it at each step of the backfitting algorithm (with the added advantage of data compression that allowed us to deal with the large data set under consideration) under the Poisson counting distributions.

To smooth with respect to the continuous risk factor `DriverAge`, we start with 73 initial internal knots (at each integer age) with $q = 3$. The upper left panel in Figure 2 displays the estimated effect of `DriverAge` obtained with A-splines. P-splines and A-splines closely agree on this data set (producing almost identical estimated age effects). Only 4 knots are selected by adaptive ridge, located at ages 30, 42, 54, and 74. The first one recognizes the extra riskiness of young drivers. The second and third knots account for the temporary increase in risk in age range 40-50, attributed to the presence of young drivers in the household, who borrow their parents' car. The last one is associated to increasing risk at older ages.

Let us now decrease q from 3 to 0 in order to categorize `DriverAge`. Starting with the same set of initial knots, we obtain the estimated piecewise function displayed in the lower left panel of Figure 2. The vertical dashed lines correspond to the breakpoints selected by adaptive ridge from the 73 initial knots, located at ages 21, 24, 27, 43, 54, 78, and 86. The resulting tariff classes for `DriverAge` are 20 and younger, $[21; 24[$, $[24; 27[$, $[27; 43[$, $[43; 54[$, $[54; 78[$, $[78; 86[$, and 86 and older.

Let us now compare the results obtained with A-splines from the classes obtained with the help of the `Insurancerating` package that is often used in practice to simplify tariff structure. On the data set under consideration, this approach results in age classes 24 and younger, $[25; 32[$, $[32; 61[$, $[61; 85[$, and 85 and older. In order to decide which partitioning is more appropriate, let us compare the respective performances of the two GLMs with the categorized versions of the two continuous features. To this end, we apply the same treatment to `CarAge`. The result is visible in the right panels of Figure 2. Again, A-splines and P-splines agree on the estimation of the smooth effect of `CarAge` which is captured with a single knot selected by adaptive ridge. A-splines with $q = 0$ produce categories $[0; 6[$, $[6; 13[$, $[13; 18[$, $[18; 21[$, and 21 and above whereas `Insurancerating` ends up with only two classes $[0; 16[$ and 16 and above. The GLM with `DriverAge` and `CarAge` categorized with the help of A-splines of degree 0 has deviance 10'389.91 and AIC 53'630. The corresponding values when `DriverAge` and `CarAge` are categorized with the help of `Insurancerating` are 10'407.87 and 53'790. We thus see the the banding produced by A-splines outperforms the one obtained from `Insurancerating` on both criteria. This shows the good performances of

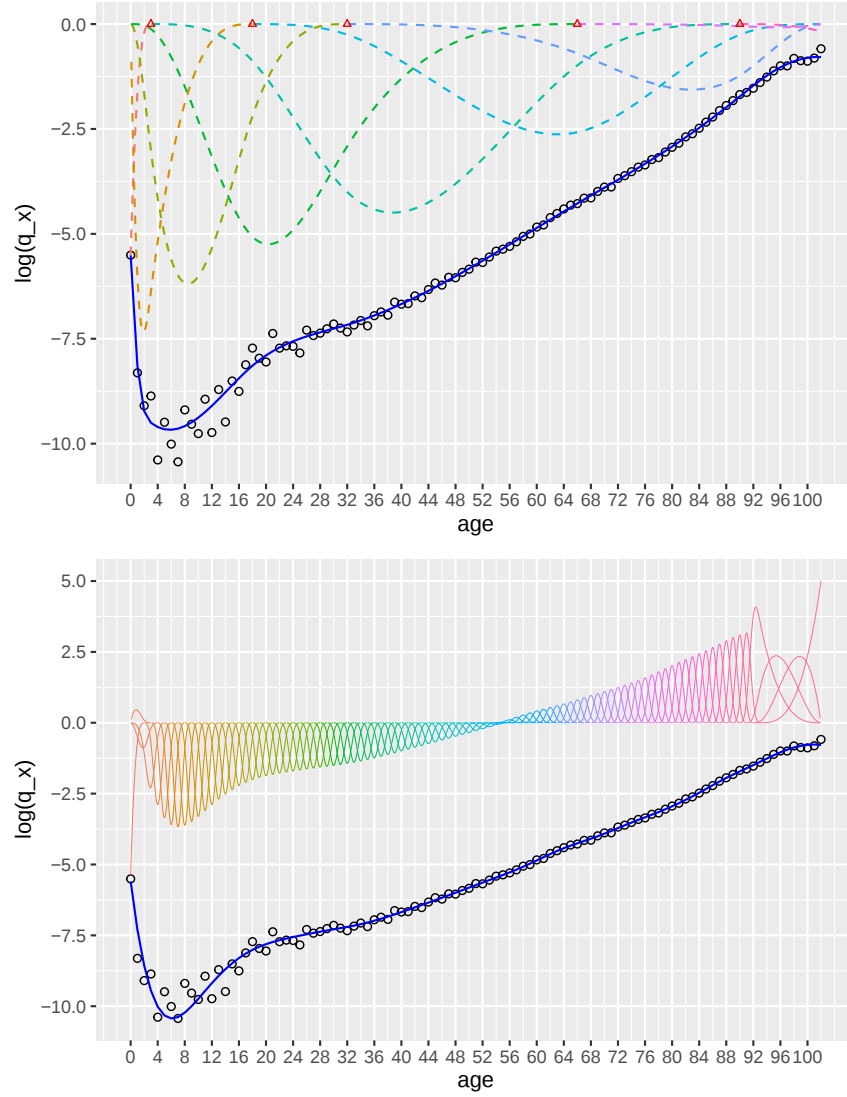


Figure 1: Knots (red triangles) and corresponding cubic splines selected by adaptive ridge with A-splines, and resulting fit (continuous line) with crude data (circles) in the upper panel. P-splines and resulting fit in the lower panel.

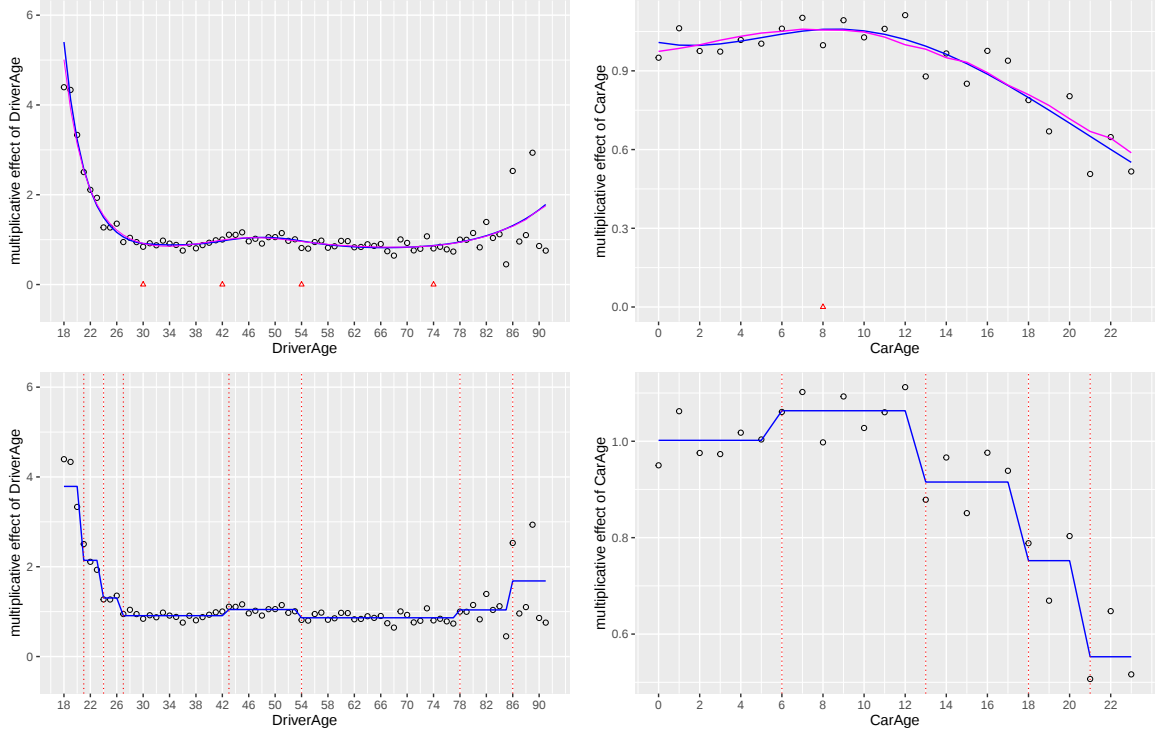


Figure 2: Knots (red triangles) selected by adaptive ridge and resulting fit with $q = 3$ (continuous line in blue for A-splines and in purple for P-splines) together with age effects obtained by treating **DriverAge** as a categorical feature (circles) in the upper left panel. Banding of **DriverAge** obtained from A-splines with $q = 0$ in the lower left panel. Analogs for **CarAge** in the right panels.

A-splines on this particular data set.

4 Discussion

A-splines are very attractive since they only differ from the widely-used P-splines by the choice of the penalty term. Also, decreasing the spline degree to 0 provides the actuary with an easy way to approximate continuous features with categorical ones (which is especially useful to simplify the price list when moving to commercial premiums) staying within the GAM setting.

The numerical illustrations provided in Section 3 suggest that A-splines are indeed a valuable tool for insurance studies. For $q = 3$, the resulting fit is essentially the same as the one produced by P-splines but with the added value that it can be interpreted: each knot reveals a particular facet of the underlying risk. This makes the model transparent, which is important when actuaries are asked to justify premium variations. This can also be exploited for a variety of other analyses. For instance, the location of knots appears to remain stable over time so that mortality projections can be obtained by extrapolating the trends visible in spline coefficients (restricting to ages 60 and above, it turns out that a single knot is enough to capture the mortality schedule). Mortality projection can thus be obtained as Hainaut and Denuit (2021) did with wavelet coefficients.

References

- Denuit, M., Hainaut, D., Trufin, J. (2019). Effective Statistical Learning Methods for Actuaries Volume 1: GLM and Extensions Springer Actuarial Lecture Notes Series.
- Eilers, P.H.C., Marx, B.D. (1996). Flexible smoothing with splines and penalties (with discussion). Statistical Science 11, 89-121.
- Goepp, V., Bouaziz, O., Nuel, G. (2018). Spline regression with automatic knot selection. arXiv preprint arXiv:1808.01770.
- Hainaut, D., Denuit, M. (2020). Wavelet-based feature-extraction for mortality projection. ASTIN Bulletin 50, 675-707.