Bayesian smoothing with extended second order random walk model: An detailed overview and comparison

Ziang Zhang

aguero.zhang@mail.utoronto.ca

Supervisor(s): James Stafford, Patrick Brown

Department of Statistical Sciences
University of Toronto

June 2021

Contents

Abstract

In this report, we will describe and implement the extended second order random walk model proposed in Lindgren and Rue (2008). This method can be viewed as an extension of the formerly used second order random walk model to irregular spaced locations/knots, derived from discretizing a stochastic differential equation. We will compare this method with other Bayesian smoothing spline methods, both conceptually and practically. This report will provide practitioners with a more thorough understanding of the connection between the second order random walk model and other Bayesian smoothing methods, and a practical guideline on how to choose among these methods.

1 Introduction

Smoothing methods are often used when there is little information on the functional structures of some covariate effects. The main challenging of smoothing is to provide enough flexibility so that the functional form of covariate effect can be accurately inferred without over-fitting the observed data. In smoothing spline method, this trade off is controlled by a smoothing parameter λ, which penalizes the wiggliness of inferred function.

Consider a data set $\{y_i, x_i, i \in [n]\}$, and a nonparametric model $y_i = g(x_i) + \epsilon_i$ where $\epsilon_i \overset{iid}{\sim} N(0, \sigma_\epsilon^2)$ and $x_i \in [a, b]$, then the smoothing spline aims to solve the following problem:

$$\arg \min_{g \in C^2} \left\{ \sum_i \left(y_i - g(x_i) \right)^2 + \lambda \int_a^b g''(x)^2 dx \right\} \tag{1}$$

The sum of square term on the left can be replaced by negative log likelihood, which is also called penalized likelihood method.

In typical frequentist method, the smoothing parameter λ is either taken as fixed value input by the users, or substituted by an optimal value selected from procedure such as REML. Therefore, how to take into account the uncertainty with the unknown hyper-parameter increases the difficulty of frequentist smoothing methods. On the other hand, the hyper-parameter λ will be assigned with a prior distribution in Bayesian smoothing methods, and hence any uncertainty involved with that parameter will be taken into account for the inference. Furthermore, the development of approximate Bayesian inference methods such as Rue et al. (2009) enables Bayesian smoothing to be implemented in a computationally convenient way. Hence, application of Bayesian smoothing method can be advantageous in a lot of settings.

Based on the well known connection between smoothing splines and integrated Wiener processes (Wahba, 1978), Lindgren and Rue (2008) developed a Bayesian smoothing method by assigning a stochastic differential equation (SDE) based prior to the unknown true effect functions. Their method uses a finite element method called Galerkin approximation to the SDE, and then solves for its weak solution. Therefore, the method of Rue et al. (2009) can be viewed as an extension of the second order random walk model (RW2) to irregular spaced locations. The hyper-parameter σ_s which is defined as $\sigma_s \propto 1/\lambda$, represents the standard deviation parameter of the second derivative of the covariate effect function, and will be assigned with a proper prior distribution. Because of the use of numerical approximation, the resulting prior distribution for the effect function will have a sparse precision matrix, and hence will be computationally efficient if used together with approximate Bayesian inference method such as Integrated Nested Laplace Approximation (INLA) (Rue et al., 2009). Both theoretical results and simulation results have been demonstrated for their

Galerkin approximation methods in their original paper (Lindgren and Rue, 2008).

In section 2, we will describe how is smoothing spline typically fitted in Frequentist method, and how it can be reinterpreted as an equivalent Bayesian inference problem with ARIMA prior when locations are equally spaced (Brown and De Jong, 2001). In section 3, we will introduce the extended second order random walk method proposed in Lindgren and Rue (2008), and provide conceptual comparison and connection with the the exact method using ARIMA method. Furthermore, we will write the ARIMA method in the form of a similar SDE specification, and hence generalize the ARIMA method to irregular spaced locations and enhances its computational efficiency. In section 4, we will implement several simulation studies to illustrate the differences between all the mentioned Bayesian smoothing spline methods, in aspects of prior sensitivity, inference accuracy and computational efficiency. We conclude in section 5 with a discussion.

2   Smoothing Spline

2.1   Fitting Smoothing Spline

Consider the smoothing parameter λ in equation 1 is a fixed constant, the solution to equation 1, denoted as $\hat{g}_\lambda(.)$, is well known to be a natural cubic polynomial spline when the response variable $\boldsymbol{y} := (y_1, y_2, ..., y_n)^T \in \mathbb{R}^n$ cannot be perfectly interpolated by a lower order polynomial function. For any function g, let $\boldsymbol{g} := (g(x_1), ..., g(x_n))^T \in \mathbb{R}^n$ denotes the corresponding evaluation vector, then the solution cubic spline $\hat{g}_\lambda(.)$ can be uniquely determined based on its evaluation vector $\hat{\boldsymbol{g}}_\lambda$ (Green and Silverman, 2019).

Using the property of natural cubic spline, the term $\int_a^b g''(x)^2 dx$ for any natural cubic spline $g(.)$ can be written as $\boldsymbol{g}^T K \boldsymbol{g}$, where the matrix K only depends on the covariate locations $\boldsymbol{x} := (x_1, ..., x_n)^T \in \mathbb{R}^n$, not on the response variable $\boldsymbol{y}$. Therefore, the equation 1 in section 1 can be written in the following vector form:

$$\arg \min_{\boldsymbol{g} \in \mathbb{R}^n} (\boldsymbol{y} - \boldsymbol{g})^T (\boldsymbol{y} - \boldsymbol{g}) + \lambda \boldsymbol{g}^T K \boldsymbol{g}. \tag{2}$$

Since this function is convex in $\boldsymbol{g}$, taking derivative and setting it to zero yields the evaluation vector $\hat{\boldsymbol{g}}_\lambda = (I + \lambda K)^{-1} \boldsymbol{y}$. Hence the solution function can be recovered from this evaluation vector.

The above procedures all treat the single smoothing parameter λ as a fixed constant. In practice, there are two common ways to select the value of λ: selecting a constant based on the subjective belief on the required smoothness of the fitted function or estimating its value based on the observed data (Green and Silverman, 2019).

If one decides to estimate the smoothing parameter using the same set of data, methods such as cross-validation (CV), generalized cross-validation(GCV) or restricted maximum likelihood estimation (REML) can be used. When computing quantities such as confidence intervals and standard errors, traditional frequentist approaches will directly plug in the estimate of λ and treat it as a known value. Therefore the traditional frequentist inference methods will tend to underestimate the variability, because they ignore the additional uncertainty from the estimation of λ.

2.2 Bayesian Smoothing Spline with ARIMA Model

Explain why the smoothing spline can be equivalently solved as a Bayesian inference problem with ARIMA prior on the evaluation vector, when knots are equally spaced.

3 Extended Second Order Random Walk Method To Smoothing Spline

This will be the main section in this report. Here I will introduce the methodology in that 2008 rw2 paper in details, and provide conceptual comparison and connection with the exact method using ARIMA method.

3.1 Prior Based On Stochastic Differential Equation

In this section, I will describe the connection between the folded Wiener's process prior defined by stochastic differential equation and the smoothing spline problem. Also, I will explain how is this connection related to Grace Wahba's conclusion in her smoothing spline paper.

3.2 Finite Element Method and Weak Solution

In this section, I will provide the context of the methodology in that 2008 rw2 paper. Specifically, I will introduce the notions such as basis functions, test functions, and finite element methods such as Bubnov-Galerkin approximation and Petrov-Galerkin approximation.

3.3 The Extended Second Order Random Walk Method

In this section, I will describe how the extended second order random walk method is derived using the Bubnov-Galerkin approximation with linear spline basis, and how it is further simplified by applying a diagonal approximation to the covariance matrix. Furthermore, I will describe how to apply the same finite element method with a different choice of basis function to get the generalization of the ARIMA method.

4 Practical Comparison

In this section, I will provide several simulation studies to illustrate the differences between all the mentioned Bayesian smoothing spline methods, in aspects of prior sensitivity, inference accuracy and computational efficiency.

5 Conclusion

Summarizes the conceptual/practical findings from section 3 and 4.

References

Brown, P. and De Jong, P. (2001). Nonparametric smoothing using state space techniques. *Canadian Journal of Statistics*, 29(1):37–50.

Green, P. J. and Silverman, B. W. (2019). *Nonparametric regression and generalized linear models: a roughness penalty approach*. Chapman and Hall/CRC.

Lindgren, F. and Rue, H. (2008). On the second-order random walk model for irregular locations. *Scandinavian Journal of Statistics*, 35(4):691–700.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.

Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 40(3):364–372.