Penalized Complexity priors for degrees of freedom in P-spline models

StEPhI — Statistics for Environmental Phenomena and their Interactions

www.stephiproject.it

Massimo Ventrucci[1] [massimo.ventrucci@unibo.it], Håvard Rue[2]

[1]Dept. of Statistical Science, University of Bologna; [2] Dept. of Mathematics, NTNU

Introduction

Penalized regression on B-splines (P-spline[1]), is a flexible and stable approach for fitting smooth functional effects in structured additive regression models[2].

$$\boldsymbol{y} = f(\boldsymbol{x}) + \boldsymbol{\epsilon}; \qquad \boldsymbol{\epsilon} \sim N(0, \tau_\epsilon^{-1}\boldsymbol{I})$$
$$f(\boldsymbol{x}) = \boldsymbol{B}\boldsymbol{\beta}$$

$\boldsymbol{B}$ is an n x K matrix of B-splines (equally-spaced over the covariate domain) and $\boldsymbol{\beta}$ a vector of coefficients.

Bayesian P-splines[3]: IGMRF prior

$$\pi(\boldsymbol{\beta}|\tau_\beta) = (2\pi)^{-\text{rank}(\boldsymbol{R})/2}(|\tau_\beta \boldsymbol{R}|^*)^{1/2} \exp\left\{-\frac{\tau_\beta}{2}\boldsymbol{\beta}^\mathsf{T}\boldsymbol{R}\boldsymbol{\beta}\right\}$$
$$\pi(\tau_\beta) = Gamma(a, b)$$

The IGMRF $\pi(\boldsymbol{\beta}|\tau_\beta)$ models deviation from a polynomial of degree $K - \text{rank}(\boldsymbol{R}) - 1$; see [4]. How large the deviation depends on τ_β. Elicitation of a prior for $\pi(\tau_\beta)$ is critical; see the scaling issue in Fig 1.

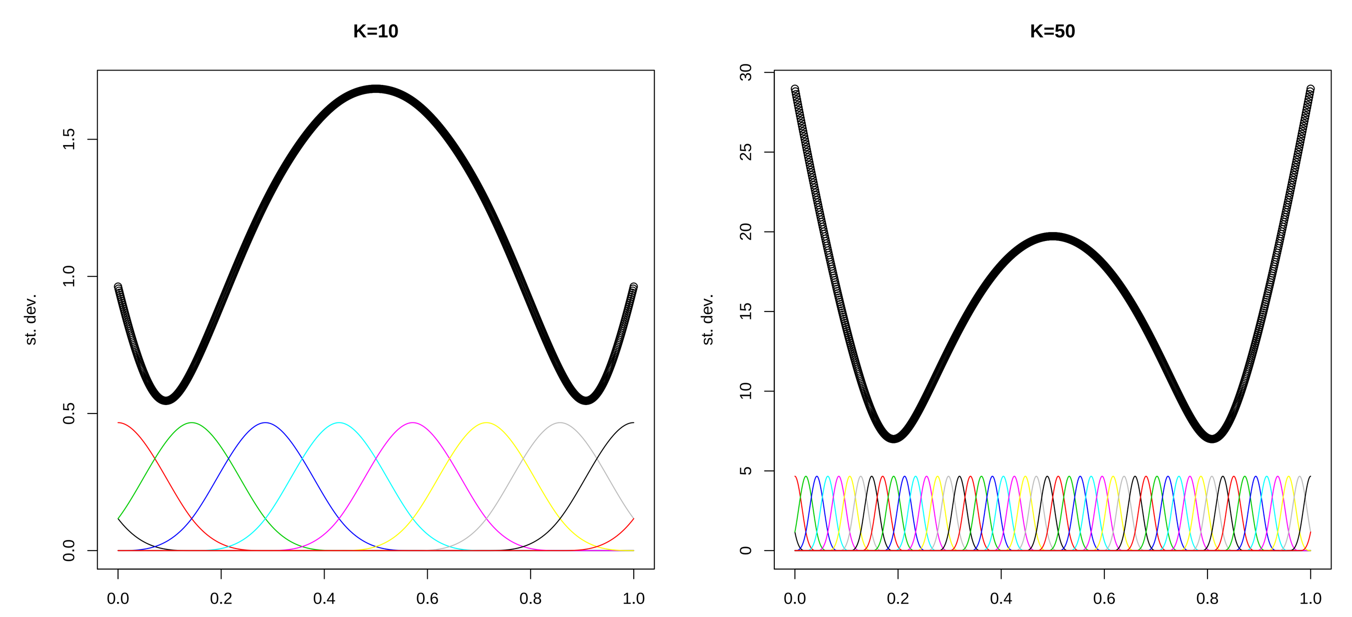
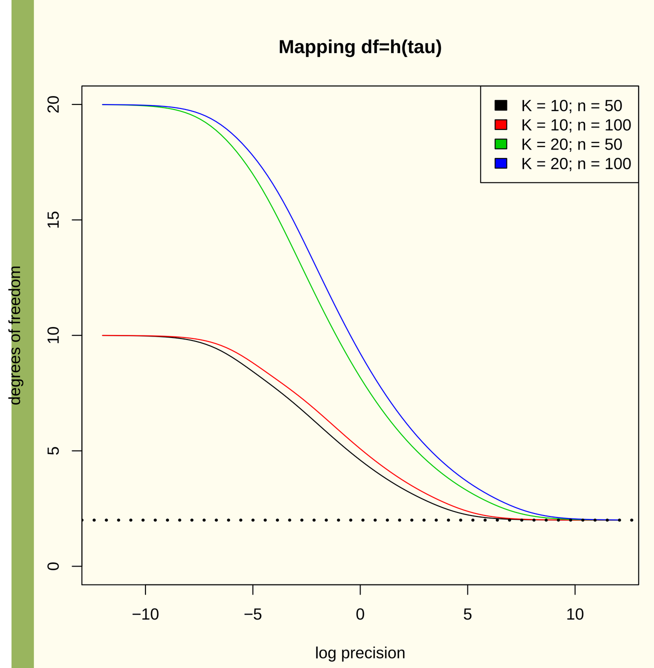


Fig 1: Scaling issue[5]. The marginal standard deviation of the IGMRF, $\boldsymbol{B}(\tau_\beta\boldsymbol{R})^{-1}\boldsymbol{B}^\mathsf{T}$, scales with design (i.e., it changes with the number of knots, i.e. with K).

Main idea

To overcome the scaling issue we consider building priors on an interpretable property of the model, indicating the complexity of the smooth function f, e.g., the **number of effective degrees of freedom**[6].

$$df = \text{tr}\left\{\left(\boldsymbol{B}^\mathsf{T}\boldsymbol{B} + \frac{\tau_\beta}{\tau_\epsilon}\boldsymbol{R}\right)^{-1}\boldsymbol{B}^\mathsf{T}\boldsymbol{B}\right\}$$
$$= \text{tr}\left\{\left(\boldsymbol{I} + \frac{\tau_\beta}{\tau_\epsilon}\boldsymbol{R}(\boldsymbol{B}^\mathsf{T}\boldsymbol{B})^{-1}\right)^{-1}\right\} = \sum_{k=1}^{K}\frac{1}{1+\frac{\tau_\beta}{\tau_\epsilon}v_k}$$



$df \in (K - rank(R), K)$; it is a function of τ_β, given τ_ϵ and **relates to the *degree of an equivalent polynomial*** (equivalent to the smooth function $\boldsymbol{B}\boldsymbol{\beta}$); $\boldsymbol{v} = \text{eigen}(\boldsymbol{R}(\boldsymbol{B}^\mathsf{T}\boldsymbol{B})^{-1})\$\text{val}$

Penalized Complexity (PC) priors

The four principles of PC priors[7] applied to P-splines

1) **Parsimony:** the *base model* is the best unless there is enough evidence for more a *flexible model*. We denote with $\boldsymbol{f}_0 = \boldsymbol{B}\boldsymbol{\beta}_0$ the base model and with $\boldsymbol{f} = \boldsymbol{B}\boldsymbol{\beta}$ the flexible one. Model f_0 is a polynomial of degree equal to the rank deficiency of $\boldsymbol{R}$ minus 1.
2) **Measure of complexity:** the *distance* between two competing models is measured by $d = \sqrt{2\text{KLD}}$; KLD is the Kullback Leibler Divergence. From[7]: for $\tau_\beta \ll \tau_{\beta_0}$ and $\tau_{\beta_0} \to \infty$, the quantity KLD$(\boldsymbol{\beta}||\boldsymbol{\beta}_0)$ goes to $\frac{\tau_{\beta_0}K}{2\tau_\beta}$, then $d = \sqrt{2\text{KLD}(\boldsymbol{\beta}||\boldsymbol{\beta}_0)} = \sqrt{\tau_{\beta_0}K/\tau_\beta}$.
3) **Constant rate penalization:** for increasing d, flexible models are penalized by a constant decay rate. The PC prior is an exponential distribution on the distance scale, $\pi_{PC}(d) = \lambda\exp(-\lambda d)$, with rate λ. By a change of variable and setting the rate $\lambda = \theta/\sqrt{K\tau_{\beta_0}}$, [7] obtain the PC prior for τ_β as a Gumbel$(1/2, \theta)$ type 2 distribution.
4) **User-defined scaling:** define θ by eliciting an upper bound for the degree of an equivalent polynomial for $\boldsymbol{f}$. Let us denote the mapping between τ_β and df (conditional on τ_ϵ) as $h(\tau_\beta|\tau_\epsilon) = \sum_{k=1}^{K}\frac{1}{1+\frac{\tau_\beta}{\tau_\epsilon}v_k}$. Let U be an upper bound for df and α the tail event probability, we need θ such that $\alpha = Pr(h(\tau_\beta) > U) = F(h^{-1}(U))$, where F is the Gumbel c.d.f,

PC prior conditional on τ_ϵ : $\pi_{PC}(df|U,\alpha) = \pi_{PC}(\tau_\beta|\theta = -\log(\alpha)\sqrt{(h^{-1}(U))})$

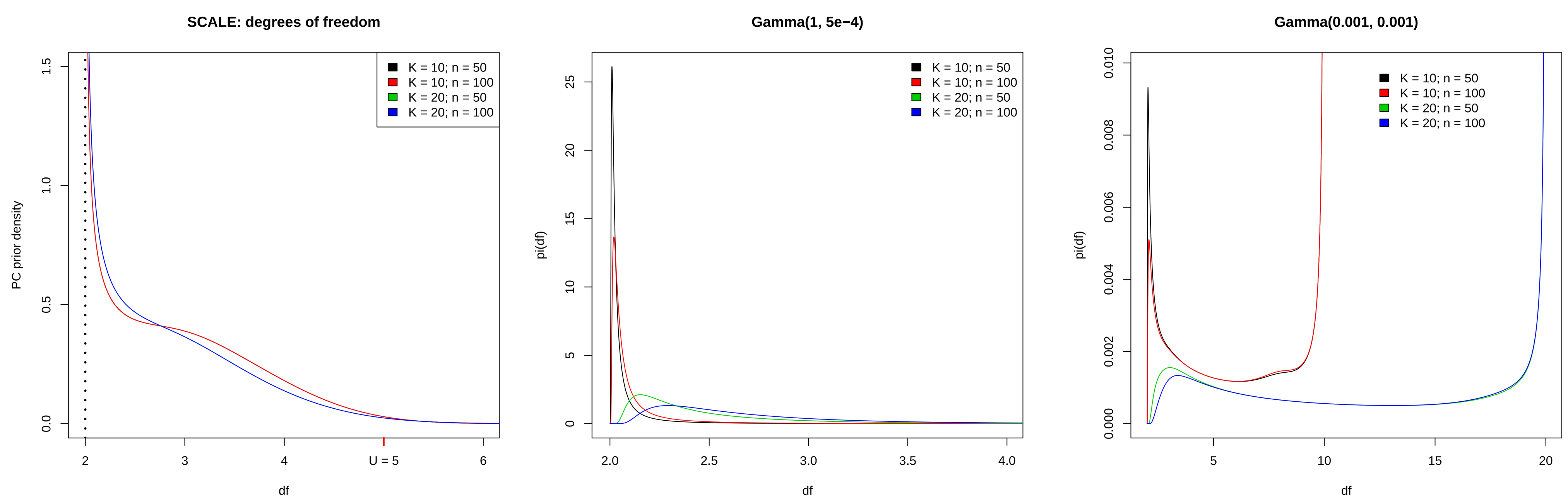


Fig 2: Implied degrees of freedom for the PC prior $\pi_{PC}(U = 5, \alpha = 0.01)$ (left), the Gamma$(1,5e-4)$ and Gamma$(0.001,0.001)$; τ_ϵ set equal to 1.

P-splines with joint prior $\pi(\tau_\beta, \tau_\epsilon) = \pi(\tau_\beta|\tau_\epsilon)\pi(\tau_\epsilon)$

$$\pi(\tau_\beta|\tau_\epsilon) = \text{Gumbel}(1/2, \theta(\tau_\epsilon)) \qquad \pi(\tau_\epsilon) \propto 1/\tau_\epsilon \tag{1}$$

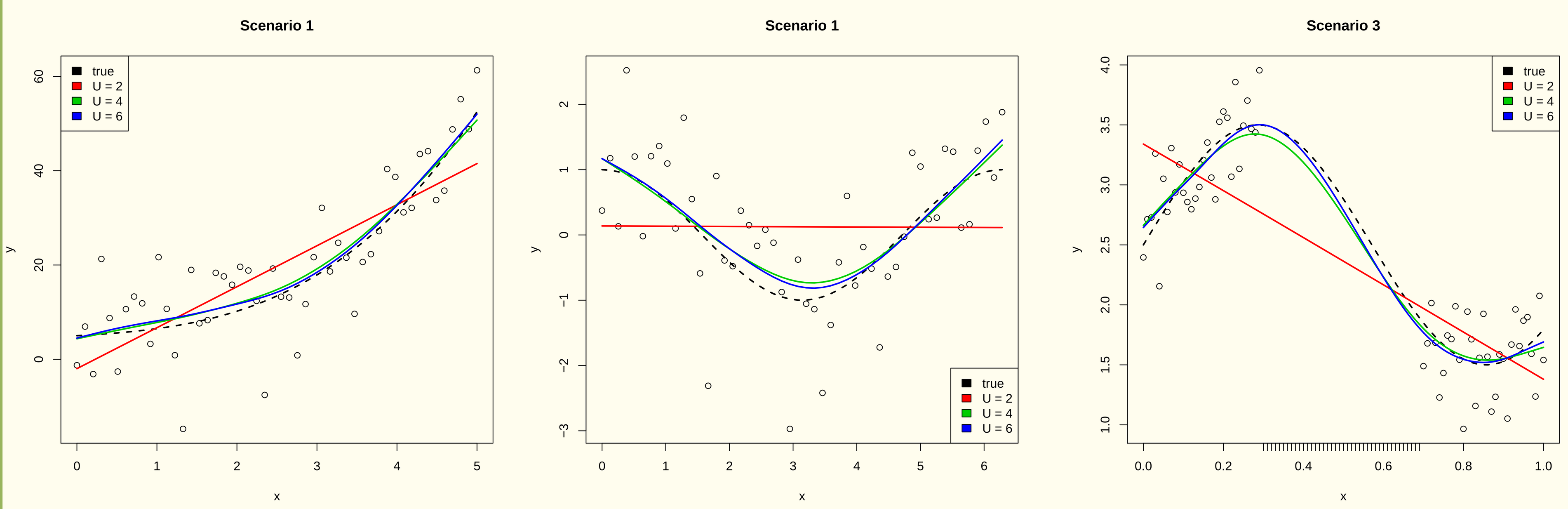


Fig 3: The joint prior in action, in three examples with simulated data.

Concluding remarks

- The PC prior $\pi_{PC}(df|U,\alpha)$ does not force overfitting (non zero probability at the base model) and has the same interpretation under different designs.
- To fit P-splines with joint prior (1) we use MCMC: $\theta(\tau_\epsilon)$ has to be recomputed at each iteration, conditional on the current τ_ϵ.
- Future work will include generalization to non Gaussian data.

[1] Eilers, P, Marx, B (1996). Flexible Smoothing with B-splines and Penalties, *Statistical Science* **11**, 89-121. [2] Fahrmeir L, Kneib T, Lang S (2004). Penalized structured additive regression for space-time data: a Bayesian perspective, *Statistica Sinica* **14**, 715-745. [3] Lang S, Brezger A (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics* **13**, 183-212. [4] Rue H, Held, L (2005). Gaussian Markov Random Fields. *Chapman and Hall*. [5] Sørbye SH, Rue H (2013). Scaling intrinsic Gaussian Markov random field priors in spatial modelling , *Spatial Statistics*. [6] Fong, Y, Rue, H, Wakefield, J (2010). Bayesian inference for generalized linear mixed models, *Biostatistics*. [7] Simpson D, Rue H, Martins TG, Riebler A, Sørbye SH (2015). Penalising model component complexity: A principled, practical approach to constructing priors, *ArXiv e-prints*.