Bayesian Non-Linear Subspace Shrinkage using Horseshoe Priors

Dr. Julia C. Duda

Non-Clinical Statistics Conference

September 25th 2024, Wiesbaden

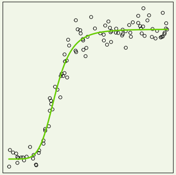
technische universität dortmund

RTG 2624

fakultät statistik

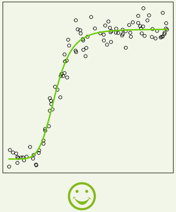
Motivation

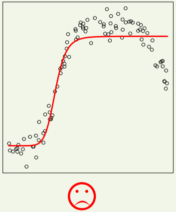
- Particular plausible model in mind
  $\rightarrow$ **Hill**
- Unknown deviations possible



Motivation

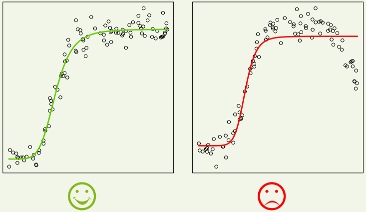
- Particular plausible model in mind
  $\rightarrow$ **Hill**
- Unknown deviations possible

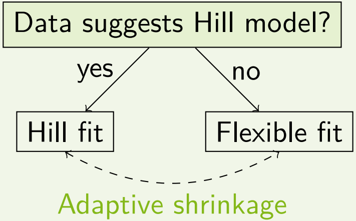




Motivation

- Particular plausible model in mind
  → **Hill**
- Unknown deviations possible





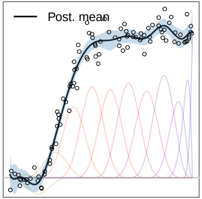
⇒ Non-parametric modeling with **subspace shrinkage**

Non-Linear Functional Shrinkage (NLFS)

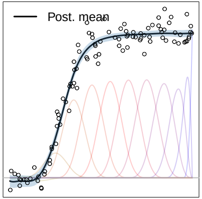
- $Y = \Phi\beta + \varepsilon,\ \varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$
- Φ Design matrix with cubic B-splines bases
- Define prior on β s.t. $\Phi\beta$ shrinks into function space [1]
  Not: Shrink β towards 0

Non-Linear Functional Shrinkage (NLFS)

- $Y = \Phi\beta + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$
- Φ Design matrix with cubic B-splines bases
- Define prior on β s.t.$\Phi\beta$ shrinks into function space [1]
  Not: Shrink β towards 0



(a) Simple B-spline

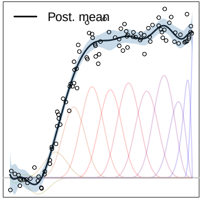


(b) Subspace shrinkage

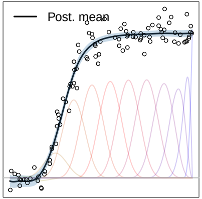
$$g(x|\theta) = \theta_1 + \theta_2 \frac{x^{\theta_4}}{\theta_3{}^{\theta_4} + x^{\theta_4}}$$

Non-Linear Functional Shrinkage (NLFS)

- $Y = \Phi\beta + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$
- Φ Design matrix with cubic B-splines bases
- Define prior on β s.t.$\Phi\beta$ shrinks into function space [1]
  Not: Shrink β towards 0



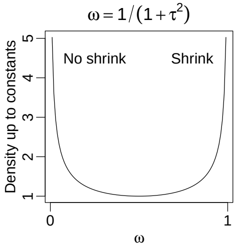
(a) Simple B-spline



(b) Subspace shrinkage

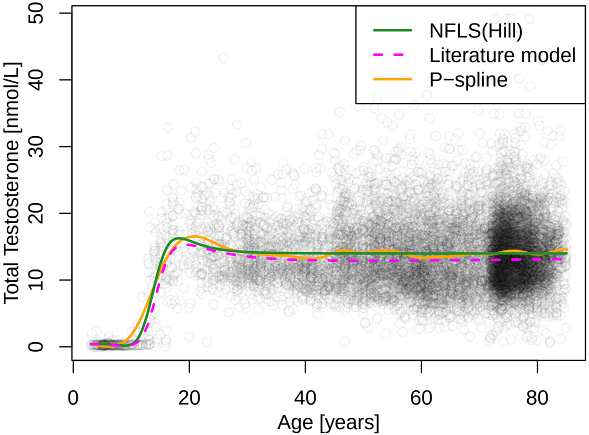
$$g(x|\theta) = \theta_1 + \theta_2 \frac{x^{\theta_4}}{{\theta_3}^{\theta_4} + x^{\theta_4}}$$

- Given θ: Linearize non-linear function (1. Order Taylor approx.)
- Construct projection matrix into approximated, linear space
- Include in prior to penalize deviations from this space
- Update θ in Bayesian sense
- Add scaling parameter τ^2 for shrinkage

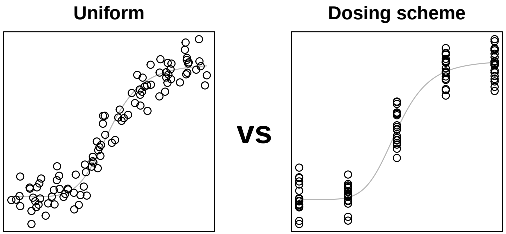


Application: Testosterone levels

- [2] compared >300 parametric models to model total testosterone levels in men
- NLFS as a comfortable alternative
- P-spline leads to artefactual bumps



Limitations and Extensions



Accounting for a dosing scheme + Add smoothing

- Including a shrinkage grid $\tilde{x}_1, \ldots, \tilde{x}_{\tilde{n}}$ independent of $x_1, \ldots, x_n$
- Gaussian processes (GPs) approach
  - Incorporate shrinkage prior in GP covariance matrix
- Add smoothness penalty [4]

Literature I

[1] J. C. Duda and M. Wheeler. Bayesian non-linear subspace shrinkage using horseshoe priors. *arXiv preprint arXiv:2407.17113*, 2024.

[2] T. W. Kelsey, L. Q. Li, R. T. Mitchell, A. Whelan, R. A. Anderson, and W. H. B. Wallace. A validated age-related normative model for male total testosterone shows increasing variance but no decline after age 40 years. *PloS one*, 9(10):e109346, 2014.

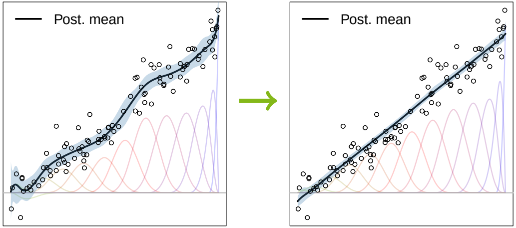
[3] E. Makalic and D. F. Schmidt. A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters*, 23(1):179–182, 2015.

[4] P. Wiemann and T. Kneib. Adaptive shrinkage of smooth functional effects towards a predefined functional subspace. *arXiv preprint arXiv:2101.05630*, 2021.

Appendix

Shrinkage Paper: Existing Theory

- $Y = \Phi\beta + \varepsilon,\ \varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$
- $\Phi = \Phi(x) = \begin{pmatrix} \phi_1(x_1) & \dots & \phi_k(x_1) \\ \vdots & \ddots & \vdots \\ \phi_1(x_n) & \dots & \phi_k(x_n) \end{pmatrix}$
- $\beta \sim$ s.t. $\Phi\beta$ shrink into column space of Φ_0
  e.g. $\Phi_0 = (\mathbf{1}_n\, \boldsymbol{x})$
  Not: Shrink β towards 0



(a) Simple B-spline (b) Subspace shrinkage

$$p(\beta|\sigma^2, \tau^2) \propto (\tau^2)^{-k/2} \exp\left(-\frac{1}{2\sigma^2\tau^2}\beta^\top\Phi^\top \underbrace{(I - P)\Phi\beta}_{\text{Subsp. deviation}}\right),\ P = \Phi_0(\Phi_0^\top\Phi_0)^{-1}\Phi_0^\top$$

$\Rightarrow$ **Prior penalizes deviations from subspace**

Shrinkage Paper: Computational overview (Hill)

$$Y = \alpha + \Phi\beta + \varepsilon$$
$$\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$$
$$\sigma^2 \sim IG(\sigma_a, \sigma_b)$$
$$\alpha \sim \mathcal{N}(\mu_\alpha, \sigma^2_\alpha)$$
$$\beta \sim \mathcal{N}(0, \sigma^2\tau^2(\Phi^\top(I - P_{\dot{F}_\theta})\Phi)^{-1})$$
$$\theta_3 \sim \mathcal{N}_+(\mu_{\theta_3}, \sigma^2_{\theta_3})$$
$$\theta_4 \sim \mathcal{LN}(\mu_{\theta_4}, \sigma^2_{\theta_4})$$
$$\omega = 1/(1+\tau^2) \sim \text{Beta}(a, b)$$
$$a = 0.5,\ b = \exp(-\log(n)/2)$$

Non-linear functional shrinkage

```
Initialize: α^(1), β^(1), σ^2(1), τ^(1), ω^(1), θ^(1)
for i : 2 → B do
    Calculate Ḟ_θ(i−1)
    Sample β^(i) ∼ p(β|·)              ▷ Conjugate
    Sample α^(i) ∼ p(α|·)              ▷ Conjugate
    Sample σ^2(i) ∼ p(σ^2|·)           ▷ Conjugate
    Sample ω^(i) ∼ p(ω|·)              ▷ Slice Sampler
    Sample θ^(i) ∼ p(θ|·)              ▷ MH Sampler
end for
return All samples
```

Overview NLFS overview

Shrinkage Paper: Multiple Function Spaces

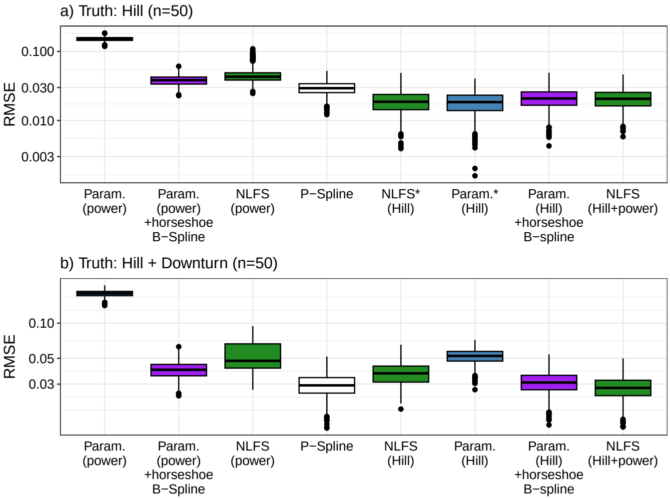
- $r \in \{1, \ldots, R\} = \mathcal{R}$ function spaces $\Omega_0^{(r)} = \{h_\theta^{(r)} | \theta \in \Theta^{(r)}\}$ of interest
- For each $\Omega_0^{(r)}$, calculate the Jacobian, $\dot{\mathsf{F}}_\theta^{(r)}$ yielding

$$\dot{\mathsf{F}}_\theta^{(\mathcal{R})} = (\dot{\mathsf{F}}_\theta^{(1)} \ldots \dot{\mathsf{F}}_\theta^{(R)})$$

- Use $\dot{\mathsf{F}}_\theta^{(\mathcal{R})}$ to construct P_θ
- Note: Full rank and no linear bases other than a single intercept column in $\dot{\mathsf{F}}_\theta^{(\mathcal{R})}$

Overview NLFS overview

Shrinkage Paper: Additional Simulation Results



Overview NLFS overview

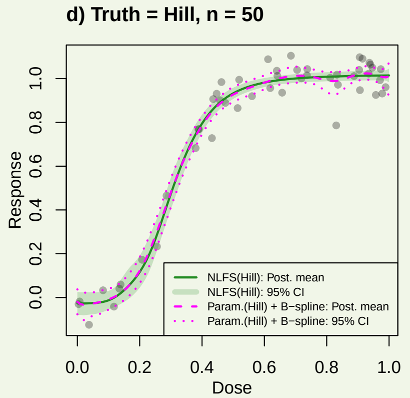
Shrinkage Paper: Additional Simulation Results

Is Param. + Shrinkage Spline the same?

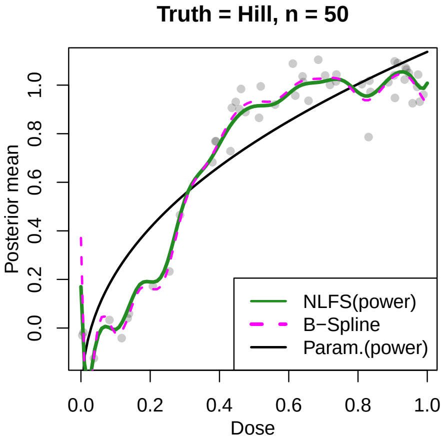
- Param.(Hill) + horseshoe B-spline
- $y = h_\theta(x) + \Phi\beta + \varepsilon$
- $\beta \sim N(0, \sigma^2\tau^2 \text{diag}({\lambda_1}^2, \dots, {\lambda_k}^2))$
- $\tau \sim C^+(0,1)$ and $\lambda_j \overset{\text{iid}}{\sim} C^+(0,1)$ [3]

Overview NLFS overview

NLFS shrinks more uniformly



Shrinkage Paper: Limitations and Extensions



Smoothing

- Add smoothing penalty similar to Wieman & Kneib 2021 [4]
- $\mathrm{Cov}(\beta|\sigma^2, \tau^2, \theta) = \sigma^2\tau^2(\Phi^\top(I - P_\theta)\Phi + \lambda^2 R^\top R)^{-1}$
- $R^\top R$ penalty matrix for smoothing splines
- New computational challenges

Overview NLFS overview