Horseshoe processes in partial linear models

Elizabeth Chase, Jeremy Taylor, and Philip Boonstra
Department of Biostatistics
University of Michigan

Thursday 25th August, 2022

Motivation

Biomedical data sometimes exhibit jumps or sharp changes:

- PSA data monitored over prostate cancer course[1]
- Women's basal body temperature monitored over time[2]
- U.S. prostate cancer incidence over time[3]

How should we fit these data?

A good statistical method would be able to fit the sharp jump without introducing excess noise into the smooth portions.

Horseshoe Prior

Suppose we wish to fit the classic linear regression model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} + \epsilon_i, \;\; \epsilon_i \overset{iid}{\sim} N(0, \sigma^2) \quad (1)$$

for $i = 1, \ldots, n$ subjects, each of whom have a vector of p predictors.

If p is large, we might want to impose some shrinkage on the coefficients $\beta_1, \ldots, \beta_p$.

Horseshoe Prior

Within a Bayesian framework, the horseshoe prior is one popular way to do so. It takes the form:

$$\begin{aligned} \beta_j | \tau, \lambda_j &\sim N(0, \tau^2 \lambda_j^2) \\ \tau &\sim C^+(0,1) \\ \lambda_j &\overset{iid}{\sim} C^+(0,1) \end{aligned} \tag{2}$$

for $j = 1, ..., p$.[4]

For each β_j in turn, the horseshoe favors either total shrinkage ($\beta_j = 0$) or minimal shrinkage (leaving β_j close to its MLE).

Gaussian Processes

Suppose we have some outcome y observed over time t. In its simplest form, a Gaussian process can be defined as:

$$y(t) - y(t-s) \sim N(0, s\tau_g^2) \tag{3}$$

A Gaussian process assumes that motion over time is normally-distributed, and it relies on a single parameter, τ_g^2.

If τ_g^2 is large, $y(t)$ will vary a lot; if τ_g^2 is small, $y(t)$ will be fairly constant.

Gaussian Processes

Suppose we have some outcome y observed over time t. In its simplest form, a Gaussian process can be defined as:

$$y(t) - y(t - s) \sim N(0, s\tau_g^2) \tag{3}$$

A Gaussian process assumes that motion over time is normally-distributed, and it relies on a single parameter, τ_g^2.

If τ_g^2 is large, $y(t)$ will vary a lot; if τ_g^2 is small, $y(t)$ will be fairly constant.

What if motion over time were horseshoe-distributed rather than normally-distributed?

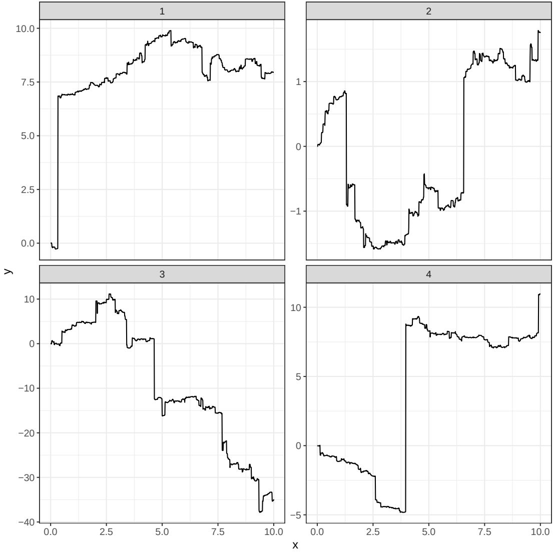
A Horseshoe Process

Using a discrete definition, for discretely observed timepoints t_k, $k = 1, ..., m$, we can define a horseshoe process:

$$\begin{aligned} y_{t_k} - y_{t_{k-1}} | \tau_h, \lambda_k &\sim N(0, \tau_h^2 \lambda_k^2 (t_k - t_{k-1})),\ \ k = 2, ..., m \\ y_{t_1} &= 0 \\ \tau_h &\sim C^+(0, c) \\ \lambda_k &\stackrel{iid}{\sim} C^+(0, 1), k = 2, ..., m \end{aligned} \tag{4}$$

Each increment is horseshoe distributed; each increment has its own local shrinkage parameter λ_k. Variance continues to scale with elapsed time.

A Horseshoe Process



Basic Model

Let y_i be some outcome observed for subjects $i = 1, ..., n$ at continuous predictor value x_i. Define $\boldsymbol{t}$ as a length m vector containing the unique, ordered values of $\boldsymbol{x}$. Suppose that $x_i = t_j$.

$$g(E(y_i)) = f_j = \alpha + \sum_{k=1}^{j} h_k \tag{5}$$

$$\begin{aligned} h_k|\tau, \lambda_k &\sim N(0, \tau_h^2 \lambda_k^2 (t_k - t_{k-1})),\ \ k = 2, ..., m \\ h_1 &= 0 \\ \tau_h &\sim C^+(0, c) \\ \lambda_k &\overset{iid}{\sim} C^+(0, 1), k = 2, ..., m \\ \alpha &\sim N(a, b^2) \end{aligned}$$

Extension: Monotonicity Constraints

We may want to constrain our horseshoe process to be monotonic increasing or decreasing. This is easily accommodated in our model through the use of a transformation m:

$$g(E(y_i)) = f_j = \alpha + \sum_{k=1}^{j} m(h_k) \tag{6}$$

m could be exponentiation or the absolute value function. We found that the absolute value function yields better performance.

Extension: Partial Linear Models

Additional covariates are easy to include. Suppose we have an additional length p vector of covariates $\boldsymbol{z}_i$ for each subject, yielding an n x p matrix $\boldsymbol{Z}$ of covariates. Then we can model:

$$g(E(y_i)) = f_i = \alpha + \boldsymbol{\beta} \boldsymbol{z}_i + \sum_{k=1}^{j} h_k \tag{7}$$

$$\boldsymbol{\beta} \sim N(\boldsymbol{0}, \boldsymbol{d}^2)$$

In theory, we could also include multiple horseshoe terms through an additive framework, although the number of parameters would grow rapidly.

Extension: Data Interpolation

There may be values of x at which we wish to obtain predictions, or to obtain a more finely-spaced grid of increments h_k with which to approximate the horseshoe process.[5]

We perform Bayesian imputation to obtain estimates of $\boldsymbol{f}$ at these augmentation points, placing the same prior on $\boldsymbol{f}_{aug}$ as is already on $\boldsymbol{f}_{obs}$.

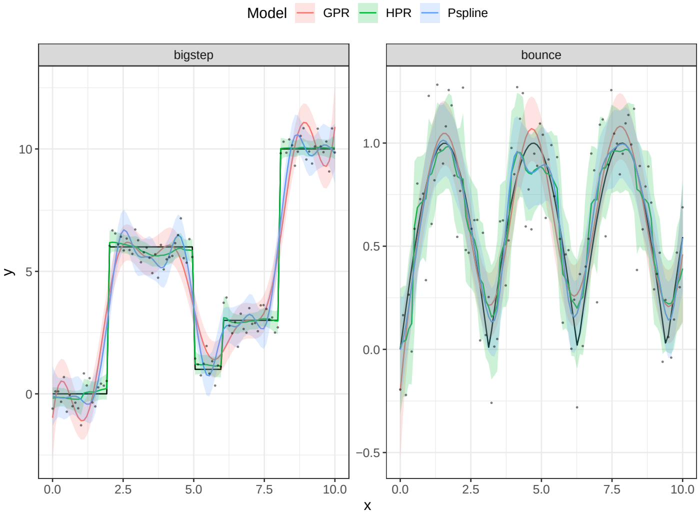
Computation

All models are implemented using Hamiltonian Monte Carlo via Stan and the cmdstanr package in R.[6,7]

For sampling, we use 4 chains, each with a warm-up phase of 1000 samples and a sampling phase of 2000 samples, without thinning, yielding 8000 posterior samples total.

All of the methods described above are implemented in the R package HPR, available on GitHub.

Sample Fits


Model
GPR
HPR
Pspline
bigstep
bounce
y
x
10
5
0
1.0
0.5
0.0
−0.5
0.0
2.5
5.0
7.5
10.0

Performance Overview

- We find that HPR performs well for step functions and piecewise linear functions, and outperforms the comparison methods at fitting step functions.
- Its performance for functions with constant variability is adequate, but not as good as methods like Gaussian process regression or penalized splines.
- We find that our data interpolation scheme returns sensible results across varying numbers of augmentation points.
- We find that our HPR partial linear model performs well at estimating the coefficients of the linear predictors in all settings, and outperforms other methods when fitting step function nonlinear components.

Full Results

Motivating Data

We use HPR to model women's basal body temperature (BBT) over the course of the menstrual cycle.

In healthy women, BBT usually starts low at the beginning of the menstrual cycle, jumping sharply immediately after ovulation, and then returning to pre-ovulation temperatures with the start of the next period.

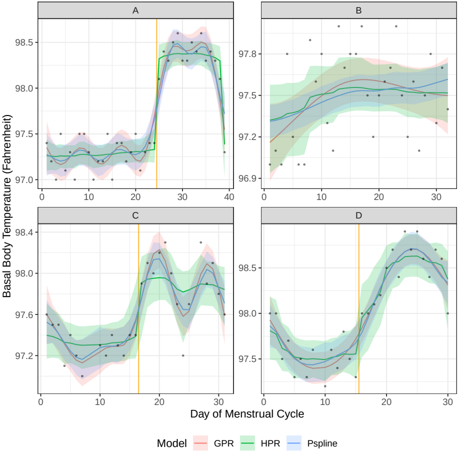
Tracking this pattern over several months may give insight into reproductive health challenges or assist with family planning.

Here, we use example BBT charts abstracted from Weschler (2015).[8]

Motivating Data

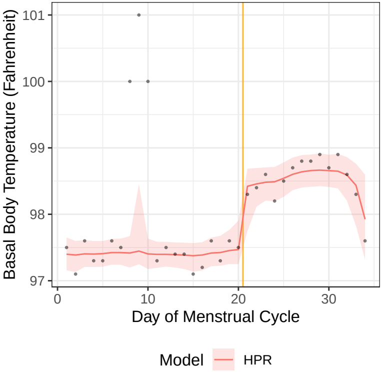
We fit a horseshoe process regression (HPR), Gaussian process regression (GPR),[9,10,11] and zero-degree penalized spline model (Pspline)[12] separately for each woman's chart.

Figure: Fitted basal body temperature trajectory and 95% credible/confidence intervals for four women.



Results

Figure: Fitted basal body temperature trajectory and 95% credible/confidence intervals from a horseshoe process regression (HPR) adjusted for the presence of fever for a woman who was ill during days 8-10 of her menstrual cycle.



Unanswered Questions

- Alternative approaches for computation (variational inference?)
- Regularized horseshoe?
- Alternative approaches for data interpolation
- Other non-Gaussian outcomes (negative binomial, ordinal)

References I

[1] JS Kim, JG Ryu, JW Kim, EC Hwang, SI Jung, TW Kang, D Kwon, and K Park. Prostate-specific antigen fluctuation: what does it mean in diagnosis of prostate cancer? *International Brazilian Journal of Urology*, 41(2):258–264, 2015.

[2] JP Royston and RM Abrams. An objective method for detecting the shift in basal body temperature in women. *Biometrics*, 36(2):217–224, 1980.

[3] AL Potosky, EJ Feuer, and DL Levin. Impact of screening on incidence and mortality of prostate cancer in the United States. *Epidemiologic Reviews*, 23(1):181–186, 2001.

[4] CM Carvalho, NG Polson, and JG Scott. The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480, 2010.

References II

[5] B Zhu, PXK Song, and JMG Taylor. Stochastic functional data analysis: a diffusion model-based approach. *Biometrics*, 67(4):1295–1304, 2011.

[6] Stan Development Team. *Stan Reference Manual, Version 2.28*. Stan (mc-stan.org), 2021.

[7] J Gabry and R Cesnovar. CmdStanR: the R interface to CmdStan. `https://mc-stan.org/cmdstanr/reference/cmdstanr-package.html`, 2021.

[8] T Weschler. *Taking Charge of Your Fertility: The Definitive Guide to Natural Birth Control, Pregnancy Achievement, and Reproductive Health*. HarperCollins, 20th anniversary edition edition, 2015.

[9] SN Wood. mgcv: GAMs and generalized ridge regression for R. *R News*, 1/2:20–25, 2001.

References III

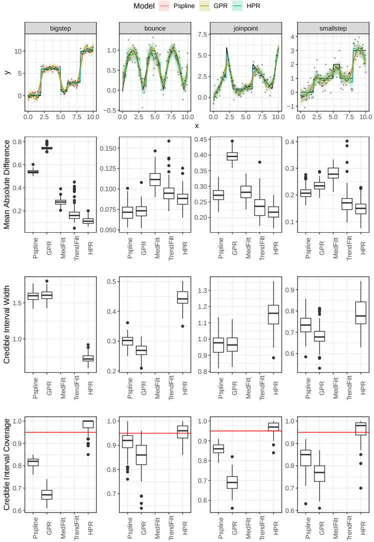
[10] SN Wood, N Pya, and B Safken. Smoothing parameter and model selection for general smooth models (with discussion). *JASA*, 111:1548–1575, 2016.

[11] CB Erickson, BE Ankenman, and SM Sanchez. Comparison of Gaussian process modeling software. *European Journal of Operational Research*, 266(1):179–192, 2018.

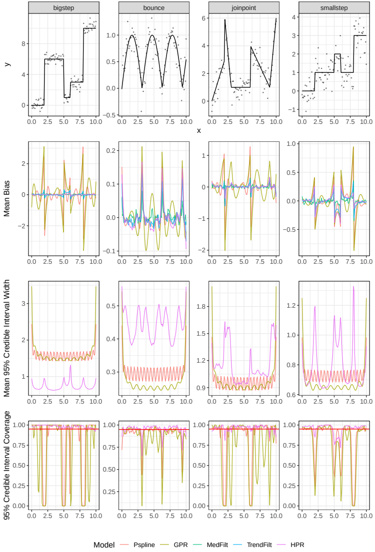
[12] PHC Eilers, BD Marx, and M Durbán. Twenty years of p-splines. *Statistics and Operations Research Transactions*, 39 (2):149–186, 2015.

Thank you! Any questions?

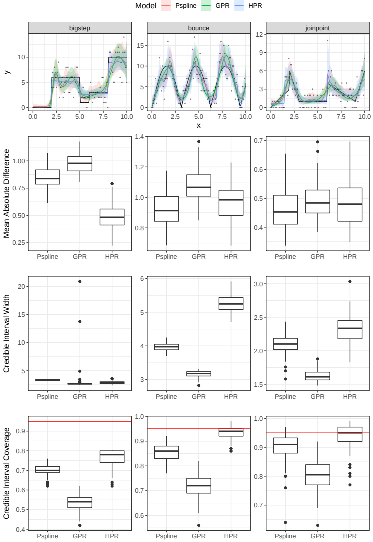
Gaussian Results



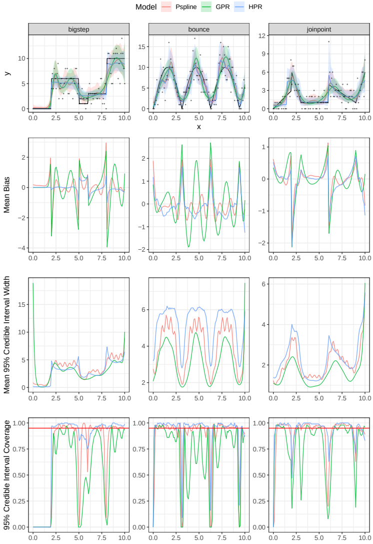
Gaussian Results



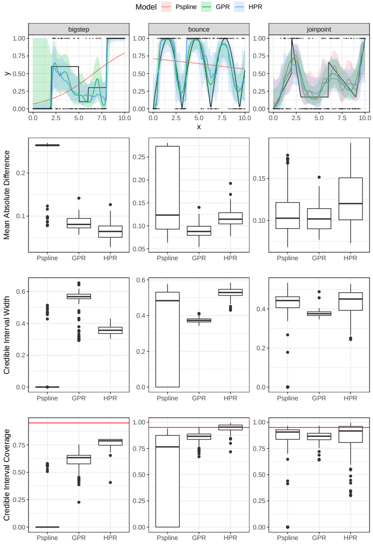
Poisson Results



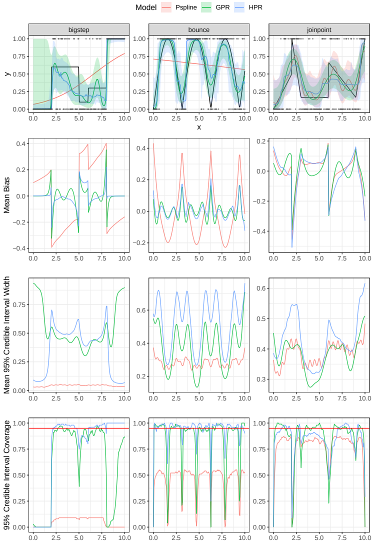
Poisson Results



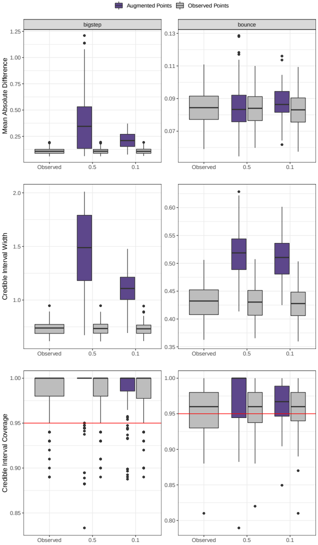
Binomial Results



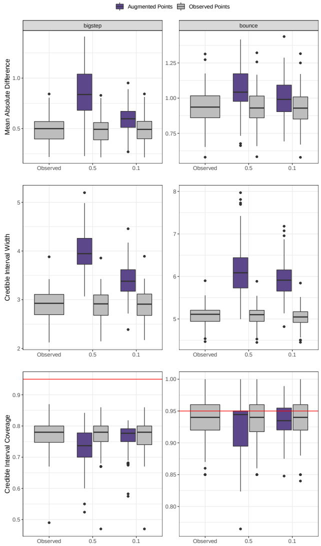
Binomial Results



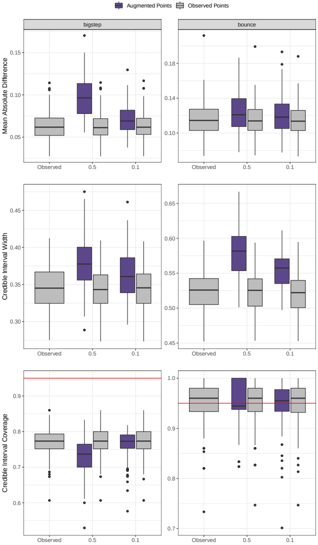
Gaussian Interpolation



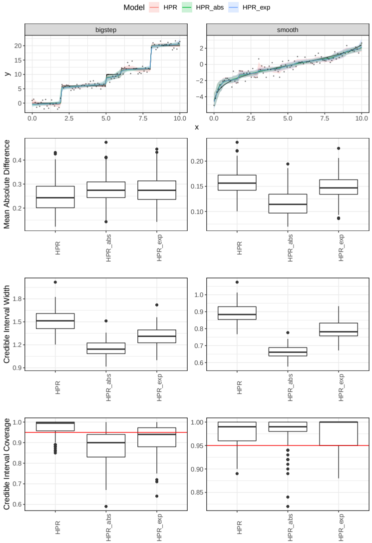
Poisson Interpolation



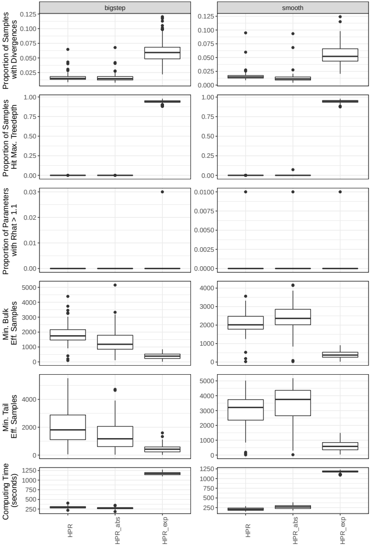
Binomial Interpolation



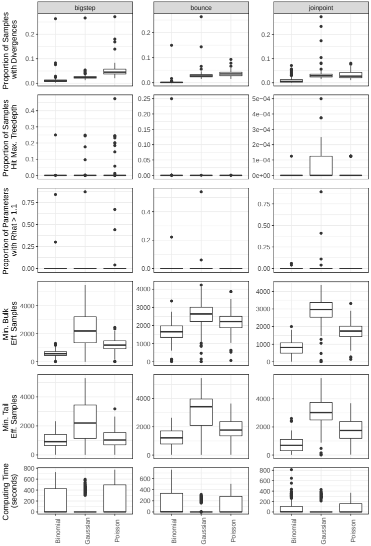
Monotonic Results



Monotonic Results



Computational Assessment



Computational Assessment - Interpolation