Slides

nparametric Regression

f Fullscreen

s Speaker View

o Slide Overview

e PDF Export Mode

b Toggle Chalkboard

c Toggle Notes Canvas

d Download Drawings

? Keyboard Help

Mass Spectroscopy

Regression

$$Y_1, \ldots, Y_n \sim \mathsf{N}\left(\mu(\mathbf{x}_i, \boldsymbol{\theta}), \sigma\right)$$

$_i \mid \boldsymbol{\theta}] = \mu(\mathbf{x}_i, \boldsymbol{\theta})$ falls in some class of nonlinear functions

ansion

$$\mu(\mathbf{x}, \boldsymbol{\theta}) = \sum_{j=1}^{J} \beta_j b_j(\mathbf{x})$$

fied set of *basis functions* and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_J)^T$ is a vector of

dinates wrt to the basis

ısion of $\mu(\mathbf{x})$ about point χ

$$\begin{aligned}\mu(x) &= \sum_k \frac{\mu^{(k)}(\chi)}{k!}(x-\chi)^k \\ &= \sum_k \beta_k (x-\chi)^k\end{aligned}$$

number of terms to model globally

r behavior in regions without data

has a "global" impact

ions

$$b_j(x, \chi_j) = (x - \chi_j)^3_+$$

sis

$$b_j(x, \chi_j) = \exp\left(\frac{(x - \chi_j)^2}{l^2}\right)$$

ıctions χ_j

:ontrols the scale at which the mean function dies out as a the center

ıents

ian Kernel g with parameters $\boldsymbol{\omega} = (\boldsymbol{\chi}, \boldsymbol{\Lambda})$

$$= g(\boldsymbol{\Lambda}_j^{1/2}(\mathbf{x} - \boldsymbol{\chi}_j)) = \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\chi}_j)^T \boldsymbol{\Lambda}_j (\mathbf{x} - \boldsymbol{\chi}_j)\right\}$$

Exponential, Double Exponential kernels (can be asymmetric)

ling of wavelet families

ned from a generator function g with location and scaling

metric Model

$$\mu(\mathbf{x}_i) = \sum_{j}^{J} b_j(\mathbf{x}_i, \boldsymbol{\omega}_j)\beta_j$$

basis elements back to our Bayesian regression model

about number of basis elements needed

other shrinkage priors

ients scale as J increases?

ainty in $\boldsymbol{\omega}$ (locations and scales)?

s (J, $\{\beta_j\}$, $\{\boldsymbol{\omega}_j\}$) induces a prior on functions!

sions

$$\mathbf{x}) = \sum_{j=0}^{J} b_j(\mathbf{x}, \boldsymbol{\omega}_j)\beta_j = \sum_{j=0}^{J} g(\boldsymbol{\Lambda}^{1/2}(\mathbf{x} - \boldsymbol{\omega}_j))\beta_j$$

easure $\nu(d\beta, d\boldsymbol{\omega})$

n $J \sim$ Poi(ν_+) where $\nu_+ \equiv \nu(\mathbb{R} \times \boldsymbol{\Omega}) = \iint \nu(\beta, \boldsymbol{\omega}) d\beta \, d\boldsymbol{\omega}$

ı $\beta_j, \boldsymbol{\omega}_j \mid J \overset{\text{iid}}{\sim} \pi(\beta, \boldsymbol{\omega}) \propto \nu(\beta, \boldsymbol{\omega})$

ıd g)

at $|\beta_j|$ are absolutely summable

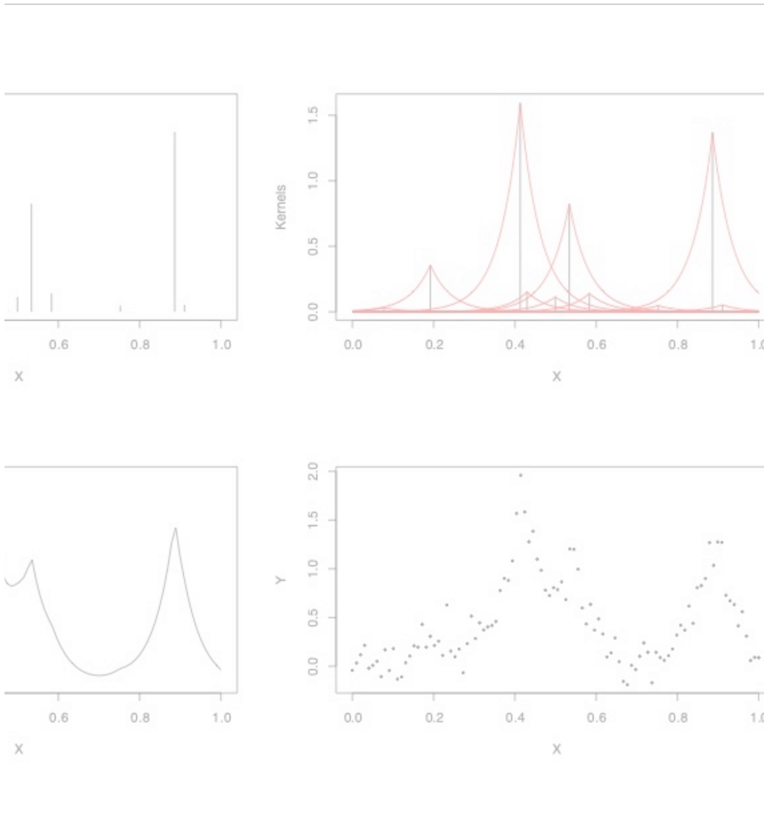
f large coefficients (in absolute value)

e number of small $\beta_j \in [-\epsilon, \epsilon]$

e and Tu (2011) AoS

:xample

$^{\eta}\gamma(\chi) d\beta \, d\chi$



l Representation

$$\ _j(\mathbf{x}, \boldsymbol{\omega}_j)\beta_j = \sum_{j=0}^{J} g(\boldsymbol{\Lambda}^{1/2}(\mathbf{x} - \boldsymbol{\omega}_j))\beta_j = \int_{\boldsymbol{\Omega}} b(\mathbf{x}, \boldsymbol{\omega})\mathcal{L}(d\boldsymbol{\omega})$$

ed measure (generalization of Completely Random Measures)

$\sim$ Lévy(ν) $\qquad \mathcal{L}(d\boldsymbol{\omega}) = \sum_{j \le J} \beta_j \delta_{\boldsymbol{\omega}_j}(d\boldsymbol{\omega})$

isson Representation of $\mathcal{L}$

support points (possibly infinite!)

ints of discrete measure $\{\boldsymbol{\omega}_j\}$

k of a random measure as stochastic process where $\mathcal{L}$ assigns
o sets $A \in \boldsymbol{\Omega}$

$$\nu(\beta, \boldsymbol{\omega}) = \beta^{-1} e^{-\beta\eta} \pi(\boldsymbol{\omega}) d\beta \, d\boldsymbol{\omega}$$
$$\mathcal{L}(A) \sim \textsf{Gamma}(\pi(A), \eta)$$

ıcients plus non-negative basis functions allows priors on non-
without transformations

:auchy process is $\alpha = 1$)

$$\nu(\beta, \boldsymbol{\omega}) = c_\alpha |\beta|^{-(\alpha+1)} \, \pi(\boldsymbol{\omega}) \qquad 0 < \alpha < 2$$

r both the Gamma and α-Stable processes

›roblematic for MCMC!

on I

ν to obtain a finite expansion:

ipport points $\boldsymbol{\omega}$ with β in $[-\epsilon, \epsilon]^c$

r approximation error)

évy measure $\nu_\epsilon(\beta, \boldsymbol{\omega}) \equiv \nu(\beta, \boldsymbol{\omega})\mathbf{1}(|\beta| > \epsilon)$

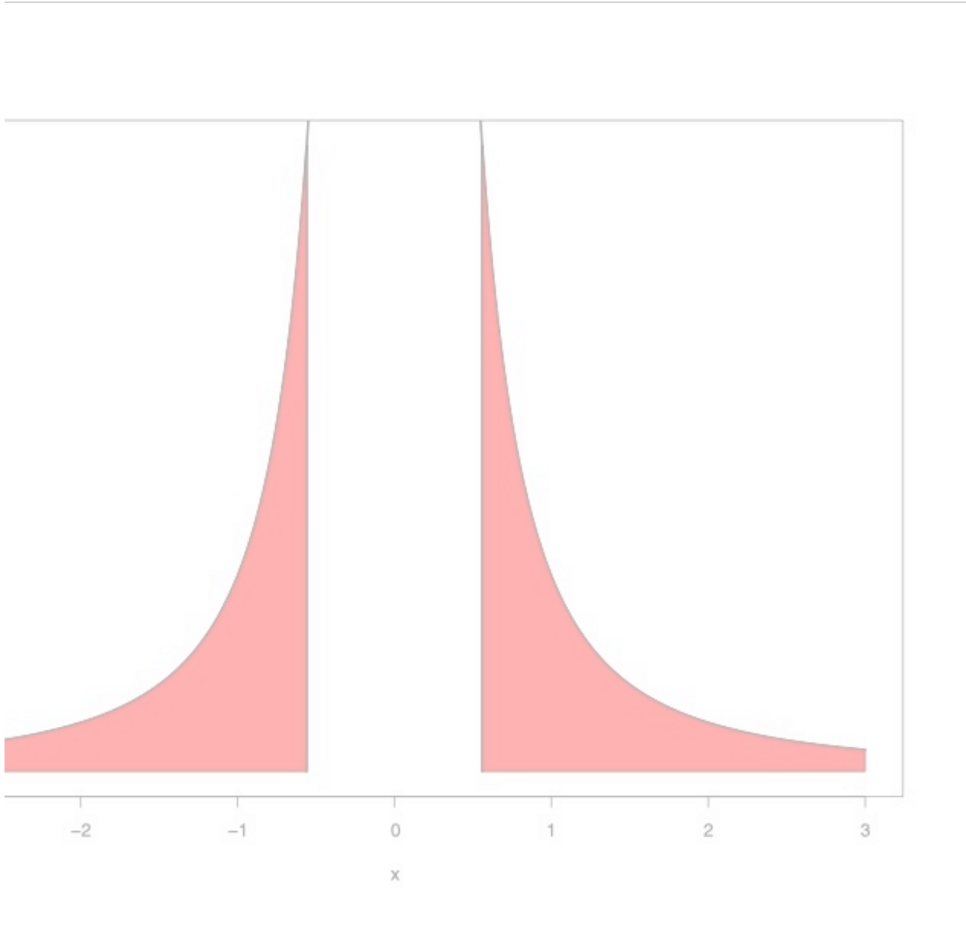
here $\nu_\epsilon^+ = \nu([-\epsilon, \epsilon]^c, \boldsymbol{\Omega})$

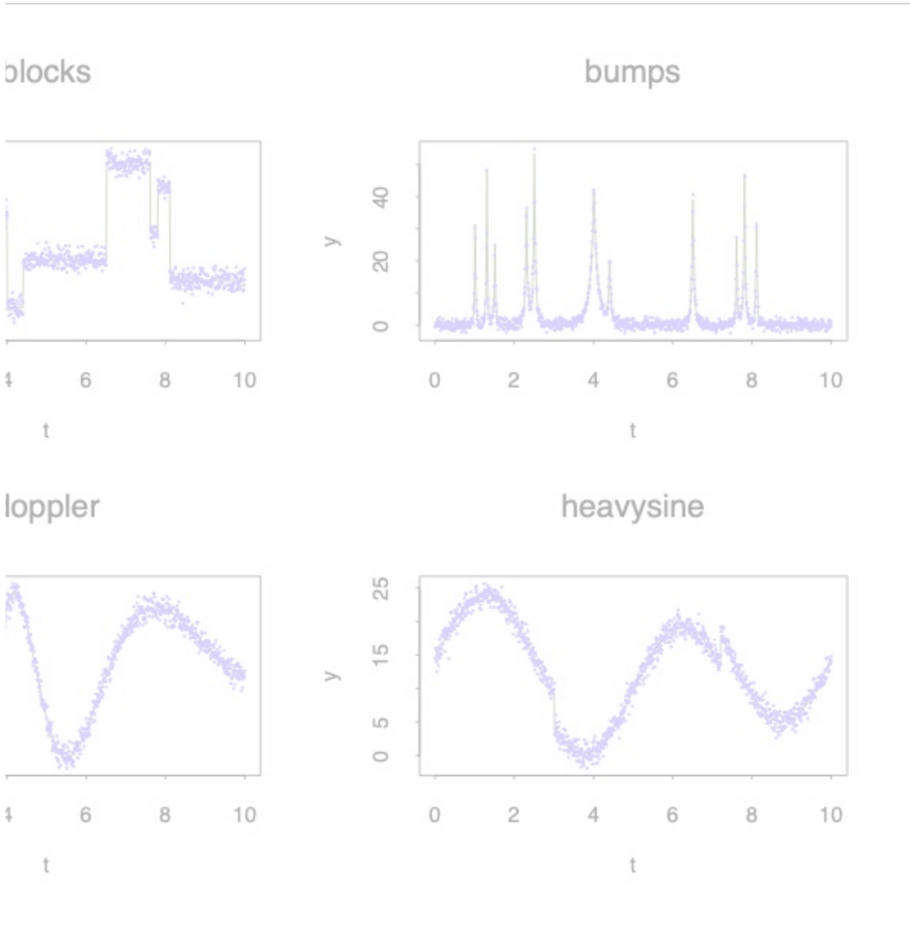
$d\boldsymbol{\omega}) \equiv \nu_\epsilon(d\beta, d\boldsymbol{\omega})/\nu_\epsilon^+$

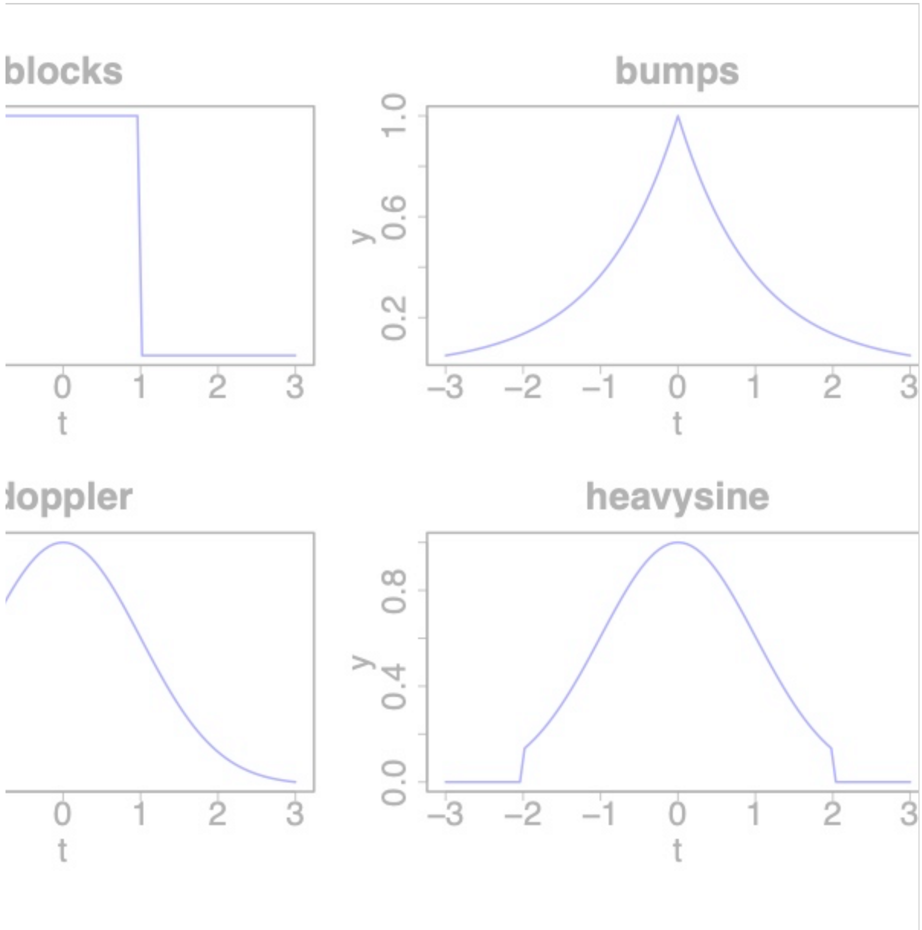
pproximation leads to double Pareto distributions for β

$$\pi(\beta_j) = \frac{\epsilon}{2\eta}|\beta|^{-\alpha-1}\mathbf{1}_{|\beta|>\frac{\epsilon}{\eta}}$$

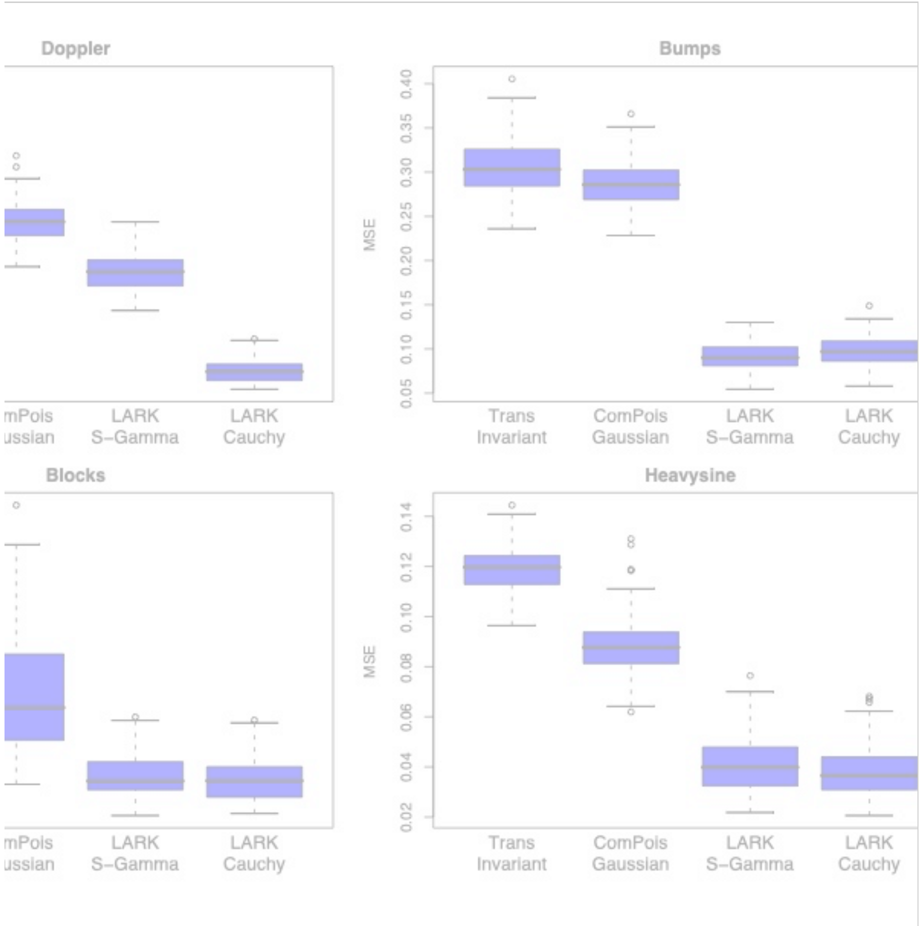
' Process Prior




blocks
bumps
doppler
heavysine
y
t
0
2
4
6
8
10
20
40
5
15
25


blocks
0
1
2
3
t
bumps
1.0
0.6
0.2
y
−3
−2
−1
0
1
2
3
t
doppler
0
1
2
3
t
heavysine
0.8
0.4
0.0
y
−3
−2
−1
0
1
2
3
t

vy Adaptive Regression Kernels



ersible Jump MCMC

MCMC

points J varies from iteration to iteration

: (birth)

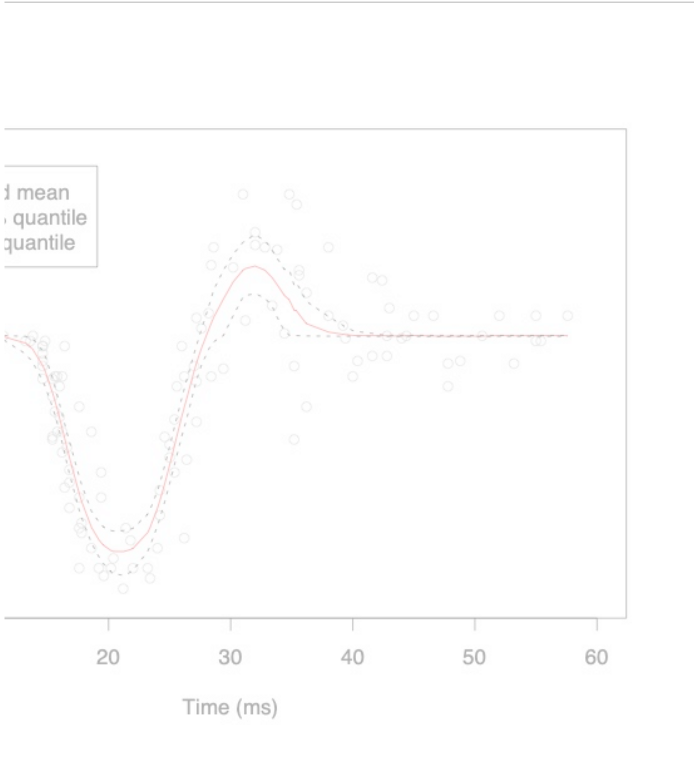
ng point (death)

oints (merge)

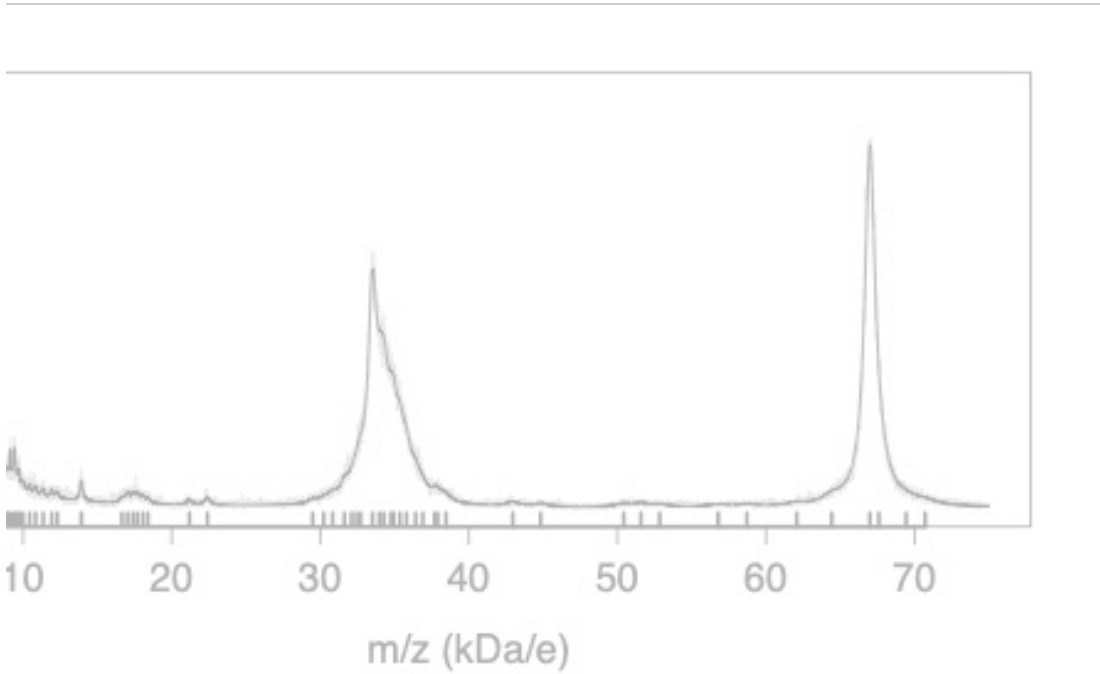
o two

nt(s)

eration



py



; than “shrinkage” priors or SVM

ıg number of support points as n increases

i through choice of ϵ

ɔn-normal data, non-negative functions or even discontinuous

:tion intervals

:o Gaussian Process Priors

ıdom scales, locations as dimension of $\mathbf{x}$ increases

:imation II

×

;ithub.io/website/