Bayesian Variable Selection and Model Choice for Structured Additive Regression using Spike-and-Slab Priors

Fabian Scheipl

Institut für Statistik
Ludwig-Maximilians-Universität München

Rigorosum
March 17, 2011
Thesis

- Part I: Bayesian Variable Selection and Model Choice for Structured Additive Regression
- Part II: Bayesian Model Choice for Locally Adaptive Splines
  - R-package *negspline*
Motivating example: Insect Venom Allergy

EAACI study

- multi-center study on bee and wasp venom allergy by European Academy of Allergology and Clinical Immunology (EAACI) Interest Group on Insect Venom Hypersensitivity
- outcome: severe, potentially life-threatening systemic reaction after a field sting (Yes/No)
- 962 patients (206 cases) in 14 clinics
- 12 potential predictors (sex, age, culprit insect, biomarkers, medication, ...)

[Ruëff et al.(2009)]
Motivating example: Insect Venom Allergy

Task: Variable selection & model choice

- which *potential* predictors are relevant risk factors for a severe reaction?
  - **Variable Selection**

- influence of predictors different for wasp or bee stings?
  - **Interaction Selection**

- *linear* influence of continuous predictors?
  - **Model Choice**

⇒ use variable selection priors on

- single coefficients

- blocks of coefficients (i.e. spline bases, factor levels, interactions) for non-Gaussian response.
Outline

Model

Prior

Inference

Performance

Application Example: Insect Venom Allergy
Structured additive regression models

\[ \eta_i = \eta_0 + \sum_{k=1}^{m} f_k(z_i), \]
\[ y_i \sim \mathcal{F}(g(\eta_i), \sigma); \quad i = 1, \ldots, n \]

[Fahrmeir et al. (2004)]

- \( \eta_0 \) includes the intercept, offsets & variables not under selection.
- \( f_k(z) \) represents factors, linear, smooth, spatial, random effects and their interactions
- \( \mathcal{F} \) from exponential family with dispersion \( \sigma \)
- response function \( g(\cdot) \)

Goal: Select and estimate relevant summands in \( \sum_{k=1}^{m} f_k(z_i) \)
Generic reparameterization

All terms represented as basis expansions with (conditionally) Gaussian coefficients:

\[ f_k(z) = Z_k \delta_k; \quad \delta_k | s_k^2 \sim \mathcal{N}(0, s_k^2 P_k^-) \]
Generic reparameterization

All terms represented as basis expansions with (conditionally) Gaussian coefficients:

\[ f_k(z) = Z_k \delta_k; \quad \delta_k \mid s_k^2 \sim \mathcal{N}(0, s_k^2 P_k^-) \]

Reparameterize:

- absorb dependency structure \( P \) of coefficients into design matrix
- split into unpenalized and penalized parts if \( P \) improper:

\[ f_k(z) = f_{0k}(z) + f_{\text{pen},k}(z) \]

\[ = X_{0k} \beta_{0k} + X_{\text{pen},k} \beta_{\text{pen},k} \quad \text{with} \quad p(\beta_{0k}) \propto 1, \beta_{\text{pen},k} \sim \mathcal{N}(0, \nu_k^2 I_{d_k}) \]
Generic reparameterization

All terms represented as basis expansions with (conditionally) Gaussian coefficients:

\[ f_k(z) = Z_k \delta_k; \quad \delta_k | s_k^2 \sim \mathcal{N}(0, s_k^2 P_k^-) \]

Reparameterize:

- absorb dependency structure \( P \) of coefficients into design matrix
- split into unpenalized and penalized parts if \( P \) improper:

\[ f_k(z) = f_{0k}(z) + f_{\text{pen},k}(z) \]

\[ = X_{0k} \beta_{0k} + X_{\text{pen},k} \beta_{\text{pen},k} \quad \text{with} \quad p(\beta_{0k}) \propto 1, \beta_{\text{pen},k} \sim \mathcal{N}(0, v_k^2 I_{d_k}) \]

- \( X_0, X_{\text{pen}} \) from spectral decomposition of implied Cov \((f(z)) \propto ZP^{-}Z'\)
- reduced rank approximation of \( X_{\text{pen}} \) via truncated SVD
  \( \Rightarrow \) fast computation, dimension reduction of posterior
Summary: Setting up the design

Reparametrization from

\[ \eta = \eta_0 + \sum_{k=1}^{m} Z_k \delta_k \text{ with } \delta_k | s_k^2 \sim \mathcal{N}(0, s_k^2 P_k^{-1}) \]

to \[ \eta = \eta_0 + \sum_{j=1}^{p} X_j \beta_j \text{ with } \beta_j | \nu_j^2 \sim \mathcal{N}(0, \nu_j^2 I), \ p \geq m, \]

\[ X_j = X_{0k} \text{ or } X_j = X_{\text{pen},k}. \]

- use proper priors for unpenalized parts in order to do selection
  ⇒ select e.g. linear and non-linear effect of a covariate separately

- \( X_j \) scaled s. t. \( \|X_j\|_F = 0.5 \ \forall j \)
  ⇒ coefficient sizes comparable in terms of effect size ⇒ same prior \( \forall j \)
Model hierarchy

\[ v_0, a_\tau, b_\tau \]

\[ w \sim \text{Beta}(a_w, b_w) \]

\[ \beta_j \sim \text{NMIG}(v_0, w, a_\tau, b_\tau) \]

\[ j = 1, \ldots, p \]

\[ y_i \sim \mathcal{F}(g(\eta_i)) \]

\[ i = 1, \ldots, n \]
Normal-Mixture of inverse Gammas (NMIG)

\[ (\alpha, \beta_j) \]

\[ \gamma_j \sim w I(\gamma_j = 1) + (1 - w) I(\gamma_j = v_0) \]

\[ \tau_j^{-2} \sim \Gamma(\alpha, \beta) \]

\[ \beta_j \sim \mathcal{N}(\mathbf{0}, \gamma_j \tau_j^2 \mathbf{I}_{d_j}) \]

[\text{Ishwaran and Rao(2005)}]
Prior

Problem: Selecting coefficient blocks

\(d_j \gg 1:\)

- mixing of \(\gamma_j\) is very problematic
- very strong dependence between \(\beta_j\) and variance \(\gamma_j \tau_j^2\):
  - large variance \(\Leftrightarrow\) large parameter values,
  - variance \(\to 0\) \(\Leftrightarrow\) parameters \(\to 0\)
- contribution of \(\beta_j\) dominates full conditionals of \(\gamma_j, \tau_j\)
  - \(\Rightarrow\) blockwise sampler very rarely leaves basin of attraction around current state
  - \(\Rightarrow\) weaken dependence by introducing non-identifiable auxiliary parameters

[Gelman et al.(2008)]
Parameter-expanded NMIG: Model Hierarchy

Parameter Expansion: $\beta_j = \alpha_j \xi_j$

Prior

$\text{Prior}$

$(v_0, a_\tau, b_\tau)$

$(a_w, b_w)$

$w \sim \text{Beta}(a_w, b_w)$

$\alpha_j \sim \text{NMIG}(v_0, w, a_\tau, b_\tau)$

$j = 1, \ldots, p$

$\beta = \text{blockdiag}(\xi_1, \ldots, \xi_p)\alpha$

$m_l \sim \frac{1}{2}I(m_l = 1) + \frac{1}{2}I(m_l = -1)$

$l = 1, \ldots, q$

$q = \sum_{j=1}^{p} d_j$

$\xi_l \sim N(m_l, 1)$
Parameter-expanded NMIG

- Parameter Expansion: $\beta_j = \alpha_j \xi_j$
- for every coefficient block $j$:
  - scalar parameter $\alpha_j \sim \text{NMIG}(\nu_0, w, a_\tau, b_\tau)$
    $\Rightarrow$ effective dimension for updating $\gamma_j$ is 1
  - $\alpha_j$ represents the “importance” of the $j$-th coefficient block
  - $\xi_j$ “distributes” $\alpha_j$ across $\beta_j$
Parameter-expanded NMIG

- Parameter Expansion: $\beta_j = \alpha_j \xi_j$
- for every coefficient block $j$:
  - scalar parameter $\alpha_j \sim \text{NMIG}(v_0, w, a_\tau, b_\tau)$
    $\Rightarrow$ effective dimension for updating $\gamma_j$ is 1
  - $\alpha_j$ represents the “importance” of the $j$-th coefficient block
  - $\xi_j$ “distributes” $\alpha_j$ across $\beta_j$
- $E(\xi|m) = \pm 1$ shrinks $\xi$ towards multiplicative identity $|1|$ & gives heavier tails than $E(\xi) = 0$
- non-identifiability of $\alpha$ and $\xi$ improves mixing for $\beta$
peNMIG properties

Marginal prior $p(\beta = \alpha \xi | a_\tau, b_\tau, a_w, b_w, \nu_0)$:

- monotonously descending score function
- infinite spike at zero & heavy tails
- implied penalty very similar to $L^q$-penalty with $q < 1$

$\Rightarrow$ Adaptive shrinkage, robustness to large $\beta$-values, efficient estimation of sparse coefficient vectors

[Polson and Scott(2010)]
Inference

Sampler

Blockwise Metropolis-within-Gibbs:

- \( w \mid \cdot \sim \text{Beta} \left( a_w + \#(j : \gamma_j = 1), b_w + \#(j : \gamma_j = v_0) \right) \)

- \( \gamma_j \mid \cdot : \quad \frac{P(\gamma_j=1|\cdot)}{P(\gamma_j=v_0|\cdot)} = v_0^{1/2} \exp \left( \frac{(1-v_0) \alpha_j^2}{2v_0 \tau_j^2} \right) \)

- \( \tau_j^{-2} \mid \cdot \sim \Gamma \left( a_\tau + 1/2, b_\tau + \frac{\alpha_j^2}{2\gamma_j} \right) \)

- \( \beta_j \mid \cdot \) via updates of \( \alpha_j \mid \cdot \) and \( \xi_j \mid \cdot \):
  - \( \mathcal{F} = \mathcal{N} \): Gibbs update from Gaussian full conditional
  - \( \mathcal{F} \neq \mathcal{N} \): P-IWLS proposals in a M-H-step [Lang and Brezger(2004)]
  - use rescaled/collapsed design matrices
    \( X_\xi = X \text{diag(blockdiag}(\mathbf{1}_{d_1}, \ldots, \mathbf{1}_{d_p})\alpha) \) and
    \( X_\alpha = X \text{blockdiag}(\xi_1, \ldots, \xi_p) \)
Properties

- indicator $\gamma_j$ on the level of the hypervariance:
  - full conditionals for $w, \gamma, \tau^2$ easy to sample, no marginalization necessary
  - standard update for $\beta$ from full conditionals/ via P-IWLS for $\alpha, \xi$
  - no hard shrinkage (i.e. $(\beta_j|\gamma_j = v_0) \neq 0$)

- hyperparameters difficult to interpret/elicit for function selection

- needs informative prior for $\tau^2$ s.t. “spike” and “slab” are distinct for clear separation between selected and un-selected terms

- parameter expansion increases computational complexity
Empirical evaluation

Investigated sensitivity to hyperparameters, estimation / prediction accuracy, selection properties for

- Simulated univariate random effect and nonlinear models (Gauss/Poisson/binomial)
- Simulated GAMs (Gauss/Poisson)
- Applications: Munich rental guide, forest health, sepsis survival, insect venom allergy
- UCI binary classification benchmark datasets and compare with mboost, mgcv, BayesX, (ACOSSO)
Performance

UCI Benchmark: Selected Datasets

- fit additive logit models
- 20-fold CV on each
- compare predictive deviance, parsimony of fits to mboost

Selected datasets:

<table>
<thead>
<tr>
<th>data set</th>
<th>n</th>
<th>covariates</th>
<th>of which factors</th>
<th>balance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cards</td>
<td>653</td>
<td>15</td>
<td>5</td>
<td>0.83</td>
</tr>
<tr>
<td>credit</td>
<td>1000</td>
<td>24</td>
<td>9</td>
<td>0.43</td>
</tr>
<tr>
<td>Heart1</td>
<td>296</td>
<td>13</td>
<td>5</td>
<td>0.85</td>
</tr>
<tr>
<td>hepatitis</td>
<td>80</td>
<td>19</td>
<td>0</td>
<td>0.19</td>
</tr>
<tr>
<td>Ionosphere</td>
<td>351</td>
<td>33</td>
<td>0</td>
<td>0.56</td>
</tr>
<tr>
<td>musk</td>
<td>476</td>
<td>166</td>
<td>0</td>
<td>0.77</td>
</tr>
<tr>
<td>Sonar</td>
<td>208</td>
<td>60</td>
<td>0</td>
<td>0.87</td>
</tr>
</tbody>
</table>
UCI Benchmark: Performance

Prediction Deviance and Inclusion Differences

\[
\frac{\text{predDev}(\text{Boost}) - \text{predDev}(\text{SpikeSlab})}{\text{no. of terms}}
\]

credit

Cards

Heart1

Ionosphere

hepatitis

Sonar

musk

prior

- \text{c(1, 1):0.005} \\
- \text{c(1, 1):0.00025} \\
- \text{c(20, 40):0.005} \\
- \text{c(20, 40):0.00025}

(selected(Boost) - selected(spikeSlab))/\text{no. of terms}
UCI Benchmark: Performance

Prediction Deviance and Exclusion Differences

(excluded(SpikeSlab)-excluded(Boost))/no. of terms

credit

Cards

Heart1

Ionosphere

hepatitis

Sonar

musk

prior

- c(1, 1): 0.005
- c(20, 40): 0.005
- c(1, 1): 0.00025
- c(20, 40): 0.00025

predDev(Boost)-predDev(SpikeSlab)

0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

Back
Conclusions

- Sensitivity to hyperparameters
  - posterior means of \( \eta \) very robust
  - selection sensitive to \( v_0 \), robust against \((a_\tau, b_\tau), (a_w, b_w)\)
  - good, stable results with \emph{identical} default hyperparameters on binomial, (standardized) Gaussian, Poisson, piecewise-exponential responses

- Predictive performance / estimation accuracy very competitive to mboost, mgcv, ACOSSO, BayesX
Conclusions

- Sensitivity to hyperparameters
  - posterior means of $\eta$ very robust
  - selection sensitive to $v_0$, robust against $(a_\tau, b_\tau), (a_w, b_w)$
  - good, stable results with *identical* default hyperparameters on binomial, (standardized) Gaussian, Poisson, piecewise-exponential responses

- Predictive performance / estimation accuracy very competitive to mboost, mgcv, ACOSSO, BayesX

- Selection
  - parsimony of fitted models competitive to mboost
  - default priors have long-run selection properties similar to RLRT on simple simulated data
  - severe overestimation of inclusion probabilities for large coefficient blocks for non-Gaussian responses (!)
  - often unstable for $p \gg n$
Conclusions

- Sensitivity to hyperparameters
  - posterior means of $\eta$ very robust
  - selection sensitive to $v_0$, robust against $(a_\tau, b_\tau), (a_w, b_w)$
  - good, stable results with *identical* default hyperparameters on binomial, (standardized) Gaussian, Poisson, piecewise-exponential responses

- Predictive performance / estimation accuracy very competitive to mboost, mgcv, ACOSSO, BayesX

- Selection
  - parsimony of fitted models competitive to mboost
  - default priors have long-run selection properties similar to RLRT on simple simulated data
  - severe overestimation of inclusion probabilities for large coefficient blocks for non-Gaussian responses (!)
  - often unstable for $p \gg n$

- CV performance in case studies shows that estimation & selection are robust against large numbers of “noise” terms
Specifying a Model with spikeSlabGAM

Logit model for severe allergic reaction with

- smooth terms for age, tryptase
- random effect for study center
- all two-way interactions between terms

```r
> library(spikeSlabGAM)
> formula <- severe ~ (insect + stings + sex + betablocker +
aceinhibitor + heartmeds + cap + age + logtryp)^2 +
rnd(studcent)
```
Specifying a Model with spikeSlabGAM

Logit model for severe allergic reaction with

- smooth terms for age, tryptase
- random effect for study center
- all two-way interactions between terms

```r
> library(spikeSlabGAM)
> formula <- severe ~ (insect + stings + sex + betablocker +
  aceinhibitor + heartmeds + cap + age + logtryp)^2 +
  rnd(studcent)
> hyperparams <- list(tau = c(5, 25), gamma = c(v0=0.00025))
> mcmc <- list(nChains = 20, chainLength = 2000, burnin = 1000, thin = 20)
```
Specifying a Model with `spikeSlabGAM`

Logit model for severe allergic reaction with

- smooth terms for age, tryptase
- random effect for study center
- all two-way interactions between terms

```r
> library(spikeSlabGAM)
> formula <- severe ~ (insect + stings + sex + betablocker +
  aceinhibitor + heartmeds + cap + age + logtryp)^2 +
  rnd(studcent)

> hyperparams <- list(tau = c(5, 25), gamma = c(v0=0.00025))
> mcmc <- list(nChains = 20, chainLength = 2000, burnin = 1000, thin = 20)

> mAll <- spikeSlabGAM(formula = formula, data = severity,
  hyperparameters = hyperparams, mcmc = mcmc, family = "binomial")
```
Application Example: Results

> summary(mAll)

Spike-and-Slab STAR for Binomial data
Model:
severe ~ (fct(insect, contr = "contr.treatment") + fct(stings, contr = "contr.treatment") + fct(sex, contr = "contr.treatment") + fct(betablocker, contr = "contr.treatment") + fct(aceinhibitor, contr = "contr.treatment") + fct(heartmeds, contr = "contr.treatment") + fct(cap, contr = "contr.treatment") + (lin(age) + sm(age)) + (lin(logtryp) + sm(logtryp)))^2 + rnd(studcent) - lin(age):sm(age) - lin(logtryp):sm(logtryp)

962 observations; 267 coefficients in 66 model terms.

Prior:
  a[tau]  b[tau]  v[0]  a[w]  b[w]
5.0e+00  2.5e+01  2.5e-04  1.0e+00  1.0e+00

MCMC:
Saved 40000 samples from 20 chain(s), each ran 40000 iterations after a burn
P-IWLS acceptance rates: 0.97 for alpha; 0.61 for xi.

Null deviance:  999
Mean posterior deviance:  841
### Application Example: Results

Marginal posterior inclusion probabilities and term importance:

<table>
<thead>
<tr>
<th>Term</th>
<th>P(γ=1)</th>
<th>π</th>
<th>dim</th>
</tr>
</thead>
<tbody>
<tr>
<td>fct(insect)</td>
<td>0.157</td>
<td>0.004</td>
<td>1</td>
</tr>
</tbody>
</table>
| fct(stings)                 | 1.000    | 0.248 | 1   ***
| fct(sex)                    | 0.701    | 0.058 | 1   **
| lin(age)                    | 0.998    | 0.329 | 1   ***
| sm(age)                     | 0.002    | 0.000 | 8   |
| lin(logtryp)                | 1.000    | 0.187 | 1   ***
| sm(logtryp)                 | 0.030    | 0.001 | 9   |
| rnd(studcent)               | 0.715    | 0.140 | 14  **
| fct(insect):fct(stings)     | 0.002    | 0.000 | 1   |
| fct(insect):sm(logtryp)     | 0.464    | 0.015 | 7   * |
| fct(cap):lin(logtryp)       | 0.142    | 0.002 | 4   |
| sm(age):lin(logtryp)        | 0.008    | 0.000 | 7   |
| sm(age):sm(logtryp)         | 0.003    | 0.000 | 32  |

*: P(γ=1) > .25 **: P(γ=1) > .5 ***: P(γ=1) > .9
Application Example: Insect Venom Allergy

Application Example: Visualization

```r
> plot(mAll, "fct(insect):sm(logtryp)", cumulative=TRUE)
```

![Graph showing the interaction of logtryp and insect categories]

- **logtryp**: Logarithmic transformation of the tryptic activity.
- **insect**: Categories such as bee and wasp.

This visualization helps in understanding the relationship between the logarithm of tryptic activity and the types of insects, highlighting how different insects affect the tryptic activity differently.

*Fabian Scheipl (LMU)*

*Bayesian Variable Selection and Model Choice*

*Rigorosum 2011*
Summary

- implemented in spikeSlabGAM: established formula syntax, hassle-free visualization of fits
- fits (geo-)additive mixed models for Gaussian, binomial and Poisson responses
- principled way of accounting for model uncertainty
- fast & stable parallelized MCMC inference
- well-tested default priors broadly applicable
- unreliable for $p \gg n$
- anti-conservative for selection of large coefficient blocks in non-Gaussian case
Outlook

- add “always included” semiparametric terms
- more general response types via data augmentation: categorical responses, AFT models, PGM error distributions (joint work with Susanne Konrath)
- diagnosing & dealing with concurvity
References


Thanks for listening.
Comments? Questions?
Centered Designs

Center $\mathbf{X}$ against lower-order effects:

$$\mathbf{X} \rightarrow \mathbf{X} (\mathbf{I} - \mathbf{Z}(\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top)$$

- $\mathbf{Z} = \mathbf{1}$ for simple terms (linear, factors)
- $\mathbf{Z} = [\mathbf{1} \ \mathbf{X}_0]$ for penalized terms (uni-/multivariate smooths, GMRFs)
  - $\Rightarrow$ enforce sum-to-zero constraint on the estimated effect
  - $\Rightarrow$ uncertainty about mean response level shifted into global intercept
- $\mathbf{Z} = [\mathbf{1} \ \mathbf{X}_{0,1} \ \mathbf{X}_{0,2} \ldots \mathbf{X}_{0,d} \ \mathbf{X}_{\text{pen},1} \ \mathbf{X}_{\text{pen},2} \ldots \mathbf{X}_{\text{pen},d}]$ for interactions between $d$ main effects
  - $\Rightarrow$ removes overlap of respective column spaces
  - $\Rightarrow$ improves separability between main effects and their interactions
Problem: Selecting coefficient blocks

For $d_j >> 1$, sampler for $\gamma_j$ gets stuck - why?
Start in equilibrium:

- iteration (0) with $v_0, w, \tau^2_0, \beta(0)$ s. t. $P(\gamma(0) = 1 | \cdot) = .5$
- look at $P(\gamma(1) = 1 | \tau^2_1, \sum^d \beta^2_1)$ as a function of $\sum^d \beta^2_1 / \sum^d \beta^2_0$
  - for various values of $d$
  - $\gamma(0) = 1$ and $\gamma(0) = v_0$
  - $\tau^2_1 = .1, .5, .9$-quantiles of its full conditional (given $\beta(0), \gamma(0)$)
Problem: Selecting coefficient blocks

\[ P(\gamma(1) = 1) \text{ as a function of the relative change in } \sum^d \beta^2 \]

for varying \( d, \gamma(0), \tau^2_{(1)} \):

\[ \frac{\sum \beta^2_{(1)}}{\sum \beta^2_{(0)}} \]
Marginal density: 
\[ p(\beta = \alpha \xi | a_\tau, b_\tau, a_w, b_w, v_0) = \int p(\alpha | \gamma, \tau^2) p\left(\frac{\beta}{\alpha}\right) \frac{1}{|\alpha|} p(\tau^2 | a_\tau, b_\tau) p(\gamma | w, v_0) p(w | a_w, b_w) d\alpha d\tau^2 d\gamma dw \]

\[ = \xi \]
peNMIG properties: Constraint regions

NMIG: \((\beta_1, \beta_2) = (\alpha_1 \xi_1, \alpha_2 \xi_2)\)

peNMIG: \((\beta_1, \beta_2) = (\alpha_1 \xi_1, \alpha_2 \xi_2)\)

peNMIG: \((\beta_1, \beta_2) = \alpha \cdot (\xi_1, \xi_2)\)
Adaptive Shrinkage

Adaptive shrinkage for simple linear model ($n = 50, p = 10, SNR = 2$):

![](chart.png)

Algorithm
- peNMIG
- NMIG
- Lasso
Function Selection: Univariate nonlinear logit-model

Data generating process: \( P(y_i = 1) = \text{logit}(\eta(x_i)) \)
Function Selection: Univariate nonlinear logit-model

### Mean Type I Error

<table>
<thead>
<tr>
<th>Type I Error Rate</th>
<th>quadratic</th>
<th>sinus</th>
<th>bumpy</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>[Graph]</td>
<td>[Graph]</td>
<td>[Graph]</td>
</tr>
<tr>
<td>0.04</td>
<td>[Graph]</td>
<td>[Graph]</td>
<td>[Graph]</td>
</tr>
<tr>
<td>0.16</td>
<td>[Graph]</td>
<td>[Graph]</td>
<td>[Graph]</td>
</tr>
<tr>
<td>0.36</td>
<td>[Graph]</td>
<td>[Graph]</td>
<td>[Graph]</td>
</tr>
<tr>
<td>0.64</td>
<td>[Graph]</td>
<td>[Graph]</td>
<td>[Graph]</td>
</tr>
<tr>
<td>1.00</td>
<td>[Graph]</td>
<td>[Graph]</td>
<td>[Graph]</td>
</tr>
</tbody>
</table>

### Mean Type II Error

<table>
<thead>
<tr>
<th>Type II Error Rate</th>
<th>quadratic</th>
<th>sinus</th>
<th>bumpy</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>[Graph]</td>
<td>[Graph]</td>
<td>[Graph]</td>
</tr>
<tr>
<td>0.25</td>
<td>[Graph]</td>
<td>[Graph]</td>
<td>[Graph]</td>
</tr>
<tr>
<td>0.50</td>
<td>[Graph]</td>
<td>[Graph]</td>
<td>[Graph]</td>
</tr>
<tr>
<td>0.75</td>
<td>[Graph]</td>
<td>[Graph]</td>
<td>[Graph]</td>
</tr>
<tr>
<td>1.00</td>
<td>[Graph]</td>
<td>[Graph]</td>
<td>[Graph]</td>
</tr>
</tbody>
</table>

### Prior

<table>
<thead>
<tr>
<th>Prior</th>
<th>(5, 50):0.00025</th>
<th>(5, 50):0.005</th>
<th>(5, 25):0.00025</th>
<th>(5, 25):0.005</th>
</tr>
</thead>
</table>

### Back
Simulation Study: Generalized Additive Model

Settings:

▷ $N = 500; 2000$
▷ $y_i \sim Po(s_i \exp(x_i' \beta)); s_i \sim U[0.66, 1.5]$
▷ covariates uncorrelated or AR(1) with correlation 0.7
▷ model structure:
  ▷ low sparsity: 16 covariates, 8 associated with smooth functions, 4 linear, 4 without influence
  ▷ high sparsity: 20 covariates, 3 associated with smooth functions, 1 linear, 16 without influence
▷ analysis:
  ▷ compare MSE, complexity recovery to `mboost::gamboost()`
  ▷ compare with MSE from oracle GAM with true formula
  ▷ look at sensitivity to hyperparameters $a_\tau, b_\tau, v_0$
Poisson GAM: Complexity recovery

Accuracy: Proportion of true positives and true negatives

- $n = 500$
- $n = 2000$
- $ho = 0$
  - sparse
  - prior $c(5, 25):0.00025$
  - prior $c(10, 30):0.00025$
  - prior $c(5, 50):0.00025$
- $ho = 0$
  - not sparse
  - prior $c(5, 25):0.005$
  - prior $c(10, 30):0.005$
  - prior $c(5, 50):0.005$