STAT 740: B-splines & Additive Models

Timothy Hanson

Department of Statistics, University of South Carolina

Outline

1. Generalized linear models
2. Additive model for normal data
3. Generalized additive mixed models

Includes many common models

The linear model (LM) encompasses many common models, including

- Multiple regression
- Multi-factor, unbalanced ANOVA
- ANCOVA models
- Interaction models
- Polynomial (i.e. response surface) models
- etc.

Benefits

- Easy interpretation of regression coefficients.
- Easy to fit, get tests.
- Linear model is first-order approximation to general "regression surface."
- Higher order models also "approximations."

The linear model

For regression data $\{(\mathbf{x}_i, Y_i)\}_{i=1}^n$ with J predictors the LM incorporating linear effects is written

$$\underbrace{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}}_{n\times 1} = \underbrace{\begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1J} \\ 1 & x_{21} & x_{22} & \cdots & x_{2J} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nJ} \end{bmatrix}}_{n\times(J+1)} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_J \end{bmatrix}}_{(J+1)\times 1} + \underbrace{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}}_{n\times 1},$$

or succinctly as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

The error vector is assumed

$$\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \tfrac{1}{\tau}\mathbf{I}_n),$$

and so the model parameters are $\boldsymbol{\beta}$ and τ.

Bayesian adds priors for β and τ

An informative prior is often

$$\beta \sim N_{J+1}(\mathbf{m}, \mathbf{S}) \text{ independent of } \tau \sim \Gamma(a, b).$$

Let $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. Then full conditional distributions for Gibbs sampling take the form

- $\beta|\tau \sim N_{J+1}(\mathbf{V}[\tau\mathbf{X}'\mathbf{Y} + \mathbf{S}^{-1}\mathbf{m}], \mathbf{V})$ where $\mathbf{V} = [\mathbf{X}'\mathbf{X}\tau + \mathbf{S}^{-1}]^{-1}$
- $\tau|\beta \sim \Gamma\left(a + 0.5n, b + 0.5||\mathbf{Y} - \mathbf{X}\beta||^2\right)$
- Easy to set up in R, JAGS, SAS (in GENMOD).
- A flat prior corresponds to $\mathbf{S}^{-1} = \mathbf{0}$ and $a = b = 0$.
- **Note**: If $\text{cov}(\epsilon) = \mathbf{R}$ then $\beta|\mathbf{R} \sim N_{J+1}(\mathbf{V}[\mathbf{X}'\mathbf{R}^{-1}\mathbf{Y} + \mathbf{S}^{-1}\mathbf{m}], \mathbf{V})$ where $\mathbf{V} = [\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} + \mathbf{S}^{-1}]^{-1}$.

Eliciting priors for β and τ

- Historical prior (aka "power prior").
- Ibrahim, J. and Chen, M.-H. (2000). Power prior distributions for regression models. *Statistical Science*, 15, 46–60.
- Data augmentation prior.
- Bedrick, E., Christensen, R., and Johnson, W. (1996). A new perspective on priors for generalized linear models. *Journal of the American Statistical Association*, 91, 1450–1460.
- *g*-prior, "default" prior.
- Zellner, A. (1983). Applications of Bayesian analysis in econometrics. *The Statistician*, 32, 23–34.

Transformations of predictors

- Scatterplot shows marginal relationship between predictors and y_i. Can lead to adding quadratic terms or simple transformations, e.g. $x_{i1}^* = \sqrt{x_{i1}}$, $x_{i1}^* = \log(x_{i1})$, etc.
- Problem: can be deceptive. (Example?)
- Added variable (aka partial regression) plots are more refined, but assume remaining predictors don't need to be transformed.
- Solution: consider J transformations simultaneously: additive model.

Ethanol data, R help file

Ethanol fuel was burned in a single-cylinder engine. For various settings of the engine compression and equivalence ratio, the emission of nitrogen oxide was recorded. Specifically, $n = 88$ observations on

- NOx: Concentration of nitrogen oxide (NO and NO2) in micrograms/J.
- C: Compression ratio of the engine.
- E: Equivalence ratio – a measure of the richness of the air and ethanol fuel mixture.

- Brinkman, N.D. (1981) Ethanol Fuel – A Single-Cylinder Engine Study of Efficiency and Exhaust Emissions. *SAE transactions*, 90, 1410–1424.

Ethanol data in R

```
library(SemiPar)
data(ethanol)
pairs(ethanol)
?ethanol
attach(ethanol)
```

Linear model in R

```
> summary(lm(NOx~E+C)) # linear in E and C

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.559101   0.662396   3.863 0.000218
E           -0.557137   0.601464  -0.926 0.356912
C           -0.007109   0.031135  -0.228 0.819941
---
Residual standard error: 1.14 on 85 degrees of freedom
Multiple R-squared: 0.01095,    Adjusted R-squared: -0.01232
F-statistic: 0.4707 on 2 and 85 DF,  p-value: 0.6262

> summary(lm(NOx~E+I(E^2)+C)) # linear C, quadratic E

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -21.2030     1.2398 -17.102  < 2e-16 ***
E            52.4110     2.7037  19.385  < 2e-16 ***
I(E^2)      -29.0899     1.4782 -19.679  < 2e-16 ***
C             0.0635     0.0137   4.635 1.3e-05 ***
---
Residual standard error: 0.484 on 84 degrees of freedom
Multiple R-squared:  0.8237,    Adjusted R-squared:  0.8174
F-statistic: 130.8 on 3 and 84 DF,  p-value: < 2.2e-16
```

Generalized linear models

- For non-normal responses: y_i is Bernoilli, Poisson, gamma, & other members of the class of exponential families.
- Good for analyzing count data & data with non-constant variance without transforming response.
- **Linear predictor** is $\eta_i = \beta_0 + \beta_1 x_{i1} + \cdots \beta_J x_{iJ} = \mathbf{x}_i' \boldsymbol{\beta}$.
- Specific models that people typically fit are
  - $y_i \sim N(\eta_i, \sigma^2)$
  - $y_i \sim Poisson(t_i \exp(\eta_i))$
  - $y_i \sim Bern\{\exp(\eta_i)/[1 + exp(\eta_i)]\}$
  - $y_i \sim \Gamma(\exp(\eta_i), \nu)$
  - $y_i \sim Mult(K, \{\Phi(\gamma_k + \eta_i) : k = 1, \ldots, K\})$
- Transformations of predictors more difficult...

Bernoulli data example

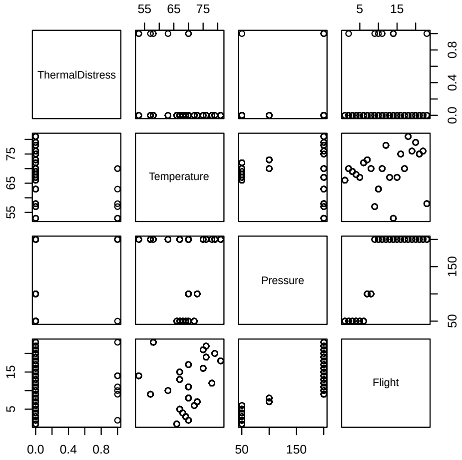
From `help(orings)`:

- Motivation: explosion of USA Space Shuttle Challenger on 28 January, 1986.
- Rogers commission concluded that the Challenger accident was caused by gas leak through the 6 o-ring joints of the shuttle.
- Dalal, Fowlkes & Hoadley (1989) looked at number distressed o-rings (among 6) versus launch temperature (Temperture) and pressure (Pressure) for 23 previous shuttle flights, launched at temperatures between 53^{o}F and 81^{o}F.
- Model: $y_i \sim \text{Bern}(\pi_i), \quad \text{logit}(\pi_i) = \beta_0 + \beta_1 T_i + \beta_2 P_i$.

O-ring data variables

- Data frame with 138 observations on the following 4 variables.
  - ThermalDistress: a numeric vector indicating wether the o-ring experienced thermal distress
  - Temperature: a numeric vector giving the launch temperature (degrees F)
  - Pressure: a numeric vector giving the leak-check pressure (psi)
  - Flight: a numeric vector giving the temporal order of flight
- Dalal, S.R., Fowlkes, E.B., and Hoadley, B. (1989). Risk analysis of space shuttle : Pre-Challenger prediction of failure. *Journal of the American Statistical Association*, 84: 945–957.

Raw data scatterplot



Logistic regression in DPpackage

```
> # linear temperature and pressure effects
> mcmc  = list(nburn=2000,nsave=2000,nskip=5,ndisplay=10,tune=1.1)
> prior = list(beta0=rep(0,3), Sbeta0=diag(10000,3))
> fit3  = Pbinary(ThermalDistress~Temperature+Pressure,link="logit",prior=prior,
              mcmc=mcmc,state=state,status=TRUE)
> summary(fit3)
Bayesian parametric binary regression model

Call:
Pbinary.default(formula = ThermalDistress ~ Temperature + Pressure,
    link = "logit", prior = prior, mcmc = mcmc, state = state,
    status = TRUE)

Posterior Predictive Distributions (log):
     Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
-6.916000  -0.040890  -0.022730  -0.197400  -0.011160  -0.006015

Regression coefficients:
             Mean        Median      Std. Dev.   Naive Std.Error   95%HPD-Low   95%HPD-Upp
(Intercept)   8.4128601   8.2787009   5.4440859   0.1217335        -1.3493922   19.5698479
Temperature  -0.1896291  -0.1853123   0.0715294   0.0015994        -0.3334473   -0.0566012
Pressure      0.0044974   0.0033455   0.0108284   0.0002421        -0.0150397    0.0262874

Acceptance Rate for Metropolis Step =  0.4344286

Number of Observations: 138
```

Additive models for normal data

- An additive model considers J simultaneous transformations of each predictor

$$y_i = \beta_0 + f_1(x_{i1}) + f_2(x_{i2}) + \cdots + f_J(x_{iJ}) + e_i.$$

- One approach to modeling the $f_1(x_1), \ldots, f_J(x_J)$ is via B-splines.
- Penalized least-squares criterion

$$\underbrace{\sum_{i=1}^{n}\left[y_i - \sum_{j=1}^{J} f_j(x_{ij})\right]^2}_{\text{makes } \sum_{j=1}^{J} f_j(x_{ij}) \text{ close to } y_i} + \underbrace{\sum_{j=1}^{J}\lambda_j \int_{a_j}^{b_j} [f_j''(x)]^2 dx}_{\text{bigger } \lambda \Rightarrow \text{ less wiggly } f_j(x)}.$$

B-splines widely used and wildly useful

USC statistics professors in our department that use B-splines in their research:

Edsel Peña, Karl Gregory, Shan Huang, David Hitchcock, Dewei Wang, John Grego, Lianming Wang, and Tim Hanson.

Tim has used B-splines (incuding Bernstein polynomials) in 10 papers over the last four years.

B-splines

B-splines, or "basis splines" are a type of spline written

$$f(x) = \sum_{k=1}^{K} \xi_k B_k(x),$$

where $B_k(x)$ is the kth B-spline basis function of degree d over the domain $[a, b]$. A simple nonparametric regression model is

$$y_i = f(x_i) + \epsilon_i, \;\; \epsilon_i \overset{iid}{\sim} N(0, \sigma^2).$$

Note that this is written as a multiple regression model

$$\mathbf{y} = \mathbf{B}\xi + \epsilon,$$

where the ith row of $\mathbf{B}$ is $(B_1(x_i), \ldots, B_K(x_i))$.

B-splines

A B-spline includes all polynomials of the same degree or less over $[a, b]$. Thus B-splines *generalize polynomial regression*. For example, a B-spline of order $d = 2$ includes all constant ($d = 0$), linear ($d = 1$), and quadratic ($d = 2$) functions over $[a, b]$ as special cases.

Having K too large leads to overfitting unless we shrink adjacent elements of $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_K)'$ to be close together. Doing so leads to a penalized B-spline.

Cardinal B-splines have equidistant knots. An alternative is to take knots to coincide with quantiles of your predictors. A very common method for nonparametric modeling of smooth trends is to use penalized B-splines with equidistant knots.

A few references

The literature of B-splines is vast. some key references related to generalized additive (mixed) models are:

- de Boor, C. (1978). *A practical Guide to Splines*. Springer, Berlin.
- Hastie, T. & Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, 1, 297–318.
- Gray, R.J. (1992). Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of the American Statistical Association*, 87, 942–951.
- Eilers, P.H.C. & Marx, B.D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11, 89–121.
- Lang, S. & Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics*, 13, 183–212.

References, more recent...

- Brezger, A., Kneib, T., and Lang, S. (2005). `BayesX`: Analyzing Bayesian structured additive regression models. *Journal of Statistical Software*, 14, 1–22.
- Hennerfeind, A., Brezger, A., & Fahrmeir, L. (2006). Geoadditive survival models. *Journal of the American Statistical Association*, 101, 1065–1075.
- Kneib, T. (2006). Mixed Model Based Inference in Structured Additive Regression. Ph.D. Thesis, Munich University.
- Krivobokova, T. (2007). Theoretical and Practical Aspects of Penalized Spline Smoothing. Ph.D. Thesis, der Universität Bielefeld
- Krivobokova, T., Kneib, T., & Claeskens, G. (2010). Simultaneous confidence bands for penalized spline estimators. *Journal of the American Statistical Association*, 105, 852–863.
- Kneib, T., Konrath, S. & Fahrmeir, L. (2011). High-dimensional structured additive regression models: Bayesian regularisation, smoothing and predictive performance, *Applied Statistics*, 60, 51-70.

Basis “mother”

A B-spline is a linear combination of *basis* functions (the “B” in B-spline). Quadratic B-spline basis function on $[0,3]$:

$$\phi(x) = \left\{ \begin{array}{ll} 0.5x^2 & 0 \leq x \leq 1 \\ 0.75 - (x - 1.5)^2 & 1 \leq x \leq 2 \\ 0.5(3 - x)^2 & 2 \leq x \leq 3 \\ 0 & \text{otherwise} \end{array} \right\}.$$

The basis functions are just shifted, shrunk/stretched versions of these.

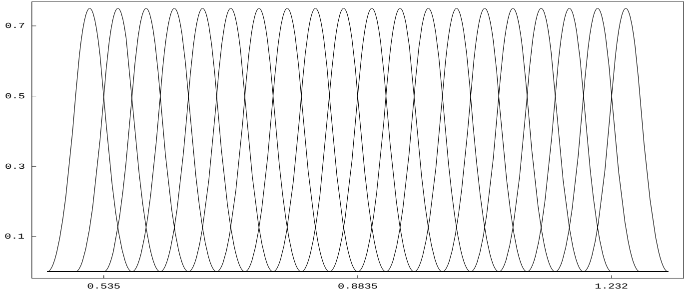
K basis functions

- Want K basis functions, typically $K = 20$.
- Without detail, kth basis function for predictor j is

$$B_{jk}(x) = \phi\left(\frac{x - a_j}{\Delta_j} + 3 - k\right), \quad \Delta_j = \frac{b_j - a_j}{K - 2}.$$

- Here $a_j = \max\{x_{1j}, x_{2j}, \ldots, x_{nj}\}$ and $b_j = \min\{x_{1j}, x_{2j}, \ldots, x_{nj}\}$.
- So (a_j, b_j) is the range of the j*th* predictor *in the data*.
- For ethanol data, $x_{i1} \in (0.535, 1.232)$.
- Next slide is $\{B_{11}(x), \ldots, B_{1,20}(x)\}$ for ethanol $x_{i1} \in (0.535, 1.232)$.

$K = 20$ quadratic basis functions over $x_{i1} \in (a_1, b_1)$



Example: B-spline basis by hand and using `splines` package.

Enforcing the same level of smoothness over the curve

- The jth predictor $f_j(x) = \sum_{k=1}^{K} \xi_{jk} B_{jk}(x)$.
- **Main idea**: Use lots of basis functions (e.g. $K = 20$ or more), but penalize $f_j(x)$ for being too "wiggly."
- This puts constraints on the $\xi_1, \ldots, \xi_J$.
- Common approach: penalize second derivative (how much slope can change) over range of the predictor.

Second order random walk prior

- For equispaced, quadratic (& cubic) B-splines,

$$\int_{a_j}^{b_j} |f_j''(x)|^2 dx = ||\mathbf{D}_2 \xi_j \Delta_j||^2.$$

- Here,

$$\mathbf{D}_2 = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix} \in \mathbb{R}^{(K-2)\times K},$$

  is second order difference penalty matrix.
- Let $\boldsymbol{\xi}_j = (\xi_{j1}, \ldots, \xi_{jK})'$ B-spline coefficients for predictor j. **Prior** is $\mathbf{D}_2 \xi_j \sim N_{K-2}(\mathbf{0}, \frac{1}{\lambda_j}\mathbf{I}_{K-2})$.
- Gives "2nd order random walk prior." As λ_j becomes large, $f_j''(x)$ is forced toward zero, and $f_j(x)$ becomes linear.

Penalized likelihood, one predictor

- Byaesian analysis via random walk prior *entirely equivalent* to penalized likelihood, except that λ is estimated from the data. Otherwise λ can be chosen via simple rules-of-thumb, arguments involving effective *df*, or cross-validation.
- For one predictor minimize

$$||\mathbf{y} - \mathbf{B}\xi||^2 + \lambda||\mathbf{D}_2\xi_j\Delta_j||^2.$$

- Estimating τ is separate.
- Recall $\lambda \to \infty \Rightarrow f(x) = \hat{\beta}_0 + \hat{\beta}_1 x$.
- **Example:** NOx vs. E, penalized and unpenalized for different λ.

First order random walk prior

- A first order random walk prior is given by $\mathbf{D}_1\xi_j \sim N_{K-1}(\mathbf{0}, \frac{1}{\lambda_j}\mathbf{I}_{K-1})$, where
- $$\mathbf{D}_1 = \begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1 \end{bmatrix} \in \mathbb{R}^{(K-1)\times K},$$
- Encourages all pairs of adjacent basis functions to have the same degree of "nearness" to each other.
- When λ_j is large, adjacent basis functions are forced closer and $f_j'(x)$ is forced toward zero, yielding $f_j(x)$ is constant.

Reduced-rank normal prior on $\{\xi_j\}$

- Either prior implies the improper prior (Speckman and Sun, 2003; Kneib, 2006)

$$p(\xi_j|\lambda_j) \propto \lambda_j^{(K-o)/2} \exp(-0.5\lambda_j \xi_j' [\mathbf{D}_o' \mathbf{D}_o] \xi_j),$$

  where $o = 1$ for 1st order and $o = 2$ for 2nd order random walk.
- Prior is informative in some directions, but not others: not informative in the space spanned by the null vectors of penalty matrix $\mathbf{D}_o'\mathbf{D}_o$.
- What are these vectors for $\mathbf{D}_1$ and $\mathbf{D}_2$? What does this imply?

Additive normal-errors model with J predictors

- The jth additive function is $f_j(x) = \sum_{k=1}^{K} \xi_{jk} B_{jk}(x)$.
- The jth matrix of B-spline basis evaluations at the observed predictors is

$$\mathbf{B}_j = \begin{bmatrix} B_{j1}(x_{1j}) & B_{j2}(x_{1j}) & \cdots & B_{jK}(x_{1j}) \\ B_{j1}(x_{2j}) & B_{j2}(x_{2j}) & \cdots & B_{jK}(x_{2j}) \\ \vdots & \vdots & \ddots & \vdots \\ B_{j1}(x_{nj}) & B_{j2}(x_{nj}) & \cdots & B_{jK}(x_{nj}) \end{bmatrix} \in \mathbb{R}^{n \times K}.$$

- The full conditional distributions for β_0, $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_J$ and τ are closed form! Gibbs sampling easy. Alternatively, the model can be written slightly differently and fit in JAGS or any mixed-model software...

Mixed model representation for 2nd order prior

- Kneib (2006) gives mixed model representation of the prior that explicitly makes use of "noninformative" and "informative" directions. Also see Eilers and Marx (2010).
- Accumulate J constant terms into one overall intercept β_0.
- Let $\mathbf{e} = (1, 2, \ldots, K)'$. Write coefficients ξ_j as mixed model with variance component λ_j:

$$\xi_j = \beta_0 + \beta_j(\mathbf{e} - \tfrac{K}{2})\mathbf{Z}\mathbf{b}_j, \quad \mathbf{Z} = \mathbf{D}_2'(\mathbf{D}_2\mathbf{D}_2')^{-1},$$

- where $p(\beta_j) \propto 1$ independent of $\mathbf{b}_j \sim N_{K-2}(\mathbf{0}, \frac{1}{\lambda_j}\mathbf{I}_{K-2})$.
- Easy to code in JAGS!
- Implies $\mathbf{D}_2\xi_j \sim N_{K-2}(\mathbf{0}, \frac{1}{\lambda_j}\mathbf{I}_{K-2})$ but separates out linear & non-linear portions.

Thus the model is

$$\begin{aligned}\mathbf{Y} &= \mathbf{X}\beta + (\mathbf{B}_1\mathbf{Z}\mathbf{b}_1 + \cdots + \mathbf{B}_J\mathbf{Z}\mathbf{b}_J) + \epsilon \\ &= \mathbf{X}\beta + [\mathbf{B}_1 \cdots \mathbf{B}_J][\mathbf{I}_J \otimes \mathbf{Z}]\mathbf{b} + \epsilon,\end{aligned}$$

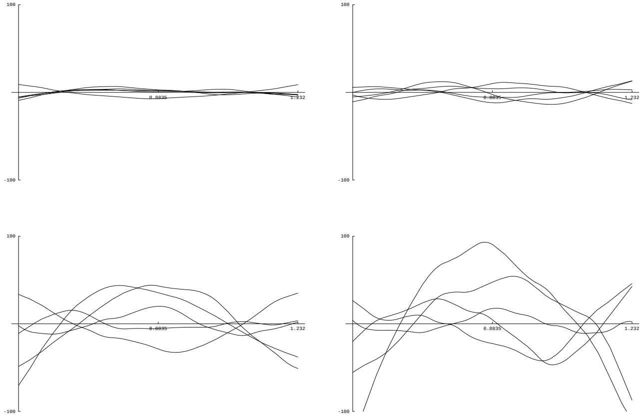
where

$$\mathbf{b}_j|\boldsymbol{\lambda} \stackrel{\perp}{\sim} N_{K-2}(\mathbf{0}, \tfrac{1}{\lambda_j}\mathbf{I}_K).$$

Example: is coming up for Ache hunting data; first need to discuss GAMs.

$\lambda = 1, 0.25, 0.04, 0.01$: prior draws of $f_1(x_1)$

This is the “wiggly part” about the linear trend $\beta_1 x_1$.



Additive model: ethanol data

Want to fit additive, normal errors model:

$$y_i = \beta_0 + f_1(E_i) + f_2(C_i) + e_i,$$

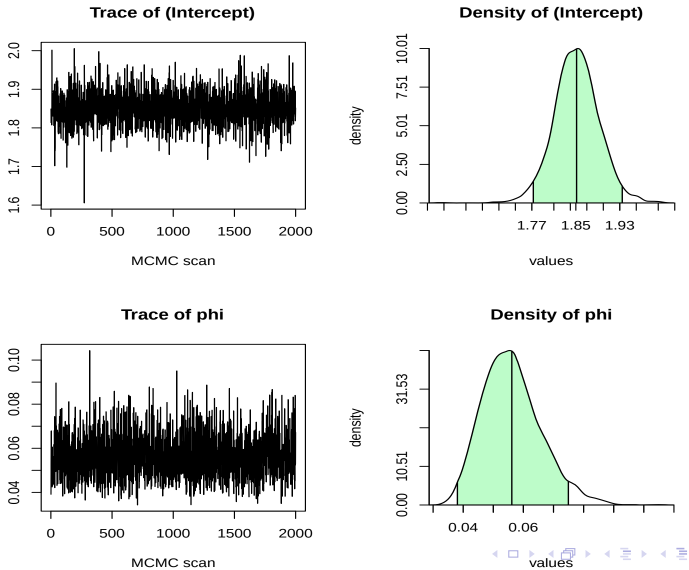
where

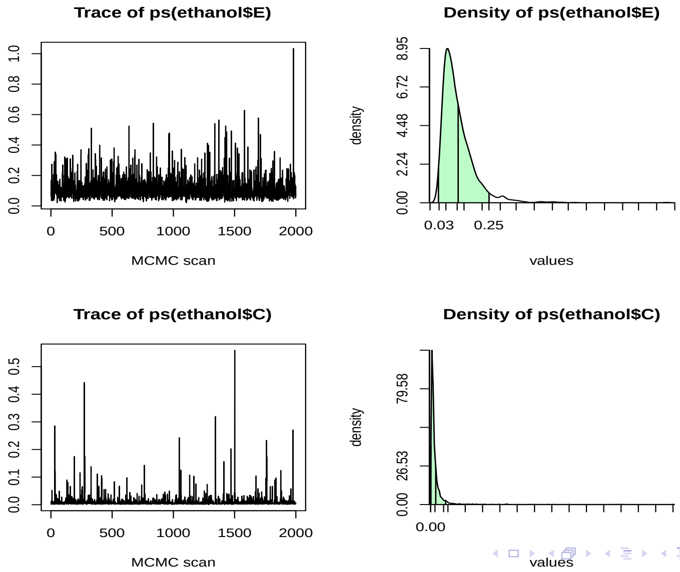
- y_i is NOx (nitrogen oxides).
- E_i is equivalence ratio.
- C_i is compression ratio of the engine.

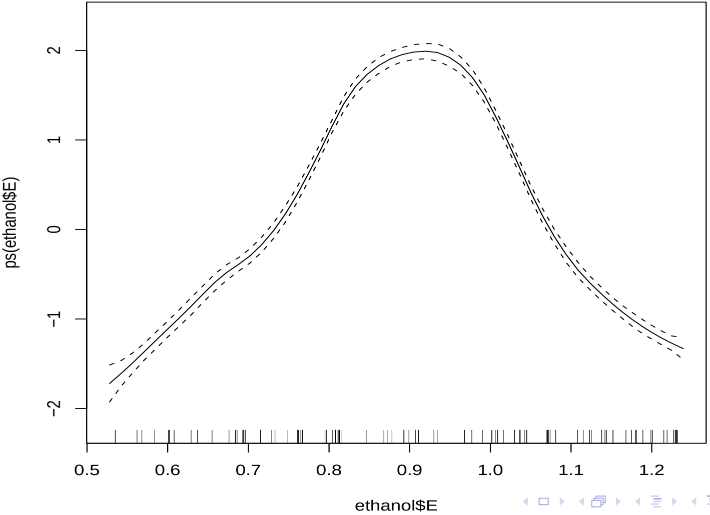
DPpackage `PSgam` function for R

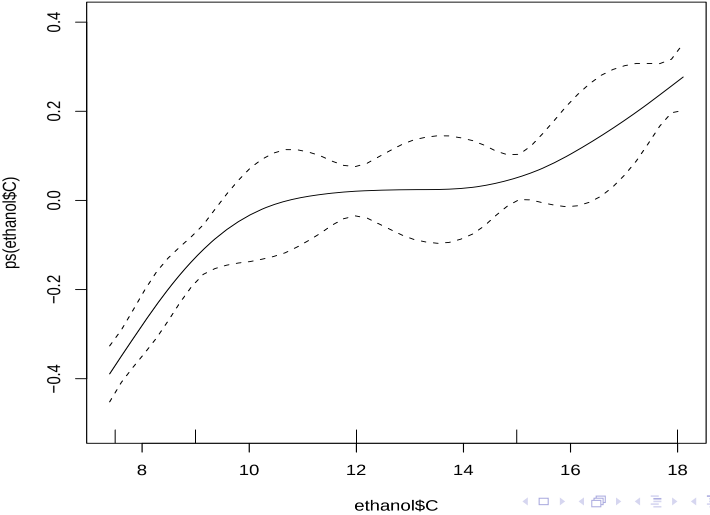
```
library(DPpackage); library(lattice)
data(ethanol); attach(ethanol); plot(ethanol)
# Additive model with additive E and C functions
prior=list(taub1=0.01,taub2=0.01,beta0=rep(0,1),Sbeta0=diag(100,1),tau1=0.001,tau2=0.001)
mcmc =list(nburn=2000,nsave=2000,nskip=199,ndisplay=500)
fit2 =PSgam(formula=ethanol$NOx~ps(E,C,k=18,degree=2,pord=2),
          family=gaussian(identity),prior=prior,mcmc=mcmc,ngrid=50,state=NULL,status=TRUE)
plot(fit2)
```

We will run this one.


Trace of (Intercept)
2.0
1.9
1.8
1.7
1.6
0
500
1000
1500
2000
MCMC scan
Density of (Intercept)
density
10.01
7.51
5.01
2.50
0.00
1.77
1.85
1.93
values
Trace of phi
0.10
0.08
0.06
0.04
0
500
1000
1500
2000
MCMC scan
Density of phi
density
31.53
10.51
0.00
0.04
0.06
values


Trace of ps(ethanol$E)
1.0
0.8
0.6
0.4
0.2
0.0
0
500
1000
1500
2000
MCMC scan
Density of ps(ethanol$E)
density
8.95
6.72
4.48
2.24
0.00
0.03
0.25
values
Trace of ps(ethanol$C)
0.5
0.4
0.3
0.2
0.1
0.0
0
500
1000
1500
2000
MCMC scan
Density of ps(ethanol$C)
density
79.58
26.53
0.00
0.00
values


ps(ethanol$E)
2
1
0
−1
−2
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
ethanol$E


ps(ethanol$C)
0.4
0.2
0.0
−0.2
−0.4
8
10
12
14
16
18
ethanol$C

Model	LPML
$\eta_i = \beta_0 + f_1(E_i)$	-28.5
$\eta_i = \beta_0 + f_1(E_i) + f_2(C_i)$	-5.8
$\eta_i = \beta_0 + f_1(E_i) + \beta_2 C_i$	-7.2
$\eta_i = \beta_0 + f_1(E_i) + \beta_{C_i}$	-6.0

Models with transformed E_i and some version of C_i predict best.

Output of summary

```
> summary(fit2)

Bayesian semiparametric generalized additive model using P-Splines

Call:
PSgam.default(formula = ethanol$NOx ~ ps(ethanol$E, ethanol$C,
    k = 18, degree = 2, pord = 2), family = gaussian(identity),
    prior = prior, mcmc = mcmc, state = NULL, status = TRUE,
    ngrid = 50)

Posterior Predictive Distributions (log):
    Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-3.21000  -0.13540   0.25140  -0.06642   0.36990   0.45030

Model's performance:
   Dbar      Dhat      pD        DIC       LPML
    -5.079   -20.087   15.008     9.929   -5.845

Parametric component:
             Mean       Median     Std. Dev.  Naive Std.Error  95%HPD-Low  95%HPD-Upp
(Intercept)  1.8513673  1.8511907  0.0410706  0.0009184        1.7726287   1.9345199
phi          0.0561522  0.0554392  0.0096374  0.0002155        0.0380783   0.0749396

Penalty parameters:
               Mean       Median     Std. Dev.  Naive Std.Error  95%HPD-Low  95%HPD-Upp
ps(ethanol$E)  0.1142265  0.0948515  0.0749338  0.0016756        0.0271463   0.2504989
ps(ethanol$C)  0.0118754  0.0056149  0.0270512  0.0006049        0.0006712   0.0348543

Number of Observations: 88
```

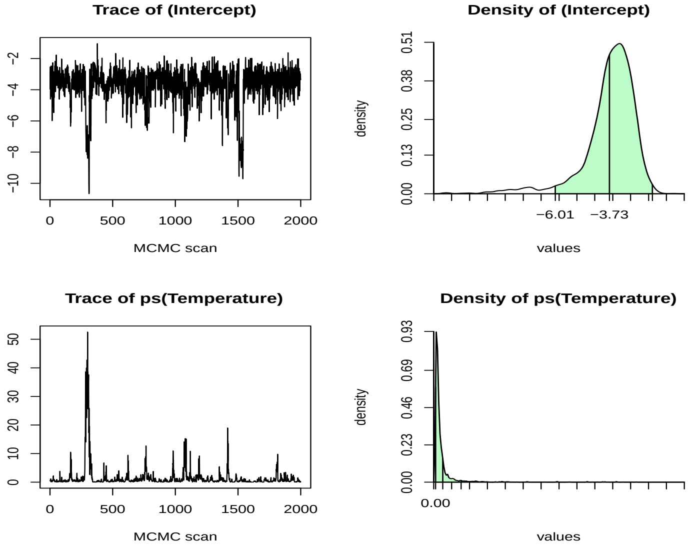
Generalized additive models

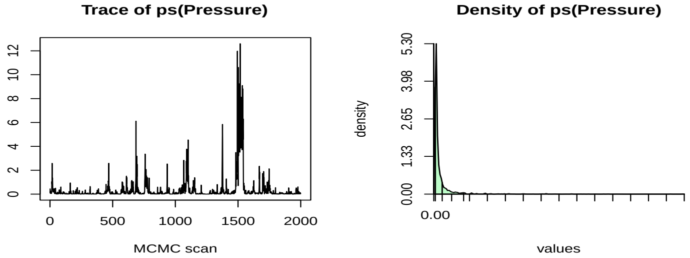
Means are parameterized:

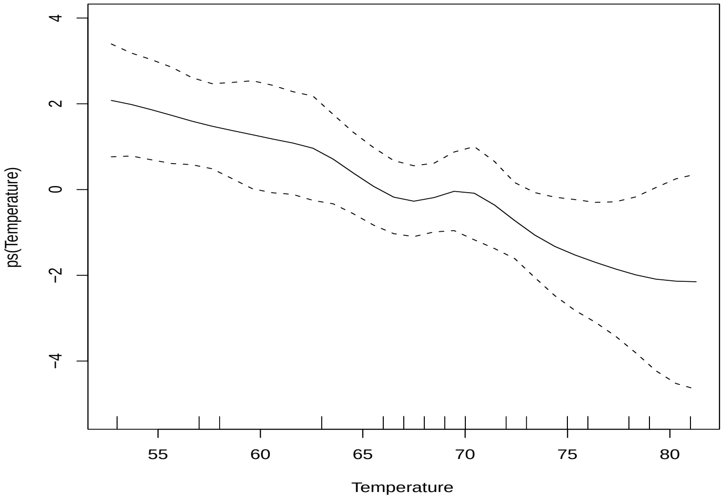
- Generalized linear model: $\eta_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_J x_{iJ}$.
- Generalized additive model: $\eta_i = \beta_0 + f_1(x_{i1}) + \cdots + f_J(x_{iJ})$.
- As before, model $f_1(x_1), \ldots, f_J(x_J)$ via penalized B-splines.
- Fitting proceeds via Gamerman's (1997) approach for GLMM.
- Can be fit in DPpackage `PSgam` or in BayesX.

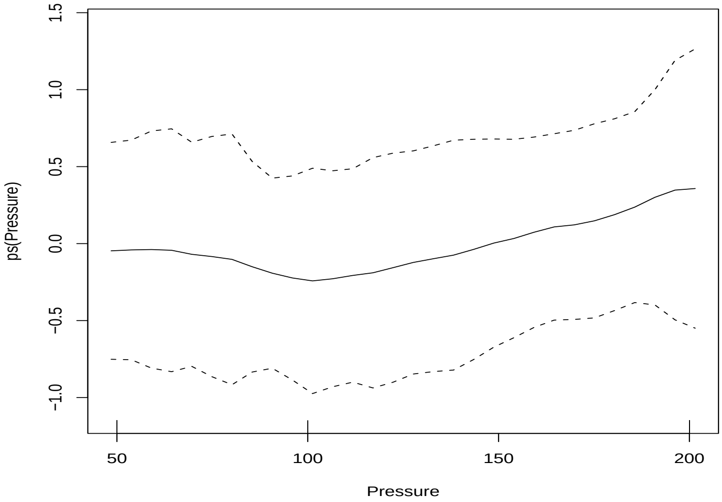
O-ring data: additive model

```
library(DPpackage)
# help(PSgam) gives function description
data(orings); attach(orings)
# additive effects in both temperature and pressure
# number of basis functions is 20, simple pairwise difference prior, 2nd order gives error
prior<-list(taub1=0.01,taub2=0.01,beta0=rep(0,1),Sbeta0=diag(100,1))
mcmc <- list(nburn=2000,nsave=2000,nskip=5,ndisplay=10)
fit1<-PSgam(formula=ThermalDistress~ps(Temperature,Pressure,k=18,degree=2,pord=1),
                family=binomial(logit),prior=prior,
                mcmc=mcmc,ngrid=30,
                state=NULL,status=TRUE)
```


Trace of (Intercept)
−2
−4
−6
−8
−10
0
500
1000
1500
2000
MCMC scan
Density of (Intercept)
0.51
0.38
0.25
0.13
0.00
density
−6.01
−3.73
values
Trace of ps(Temperature)
50
40
30
20
10
0
0
500
1000
1500
2000
MCMC scan
Density of ps(Temperature)
0.93
0.69
0.46
0.23
0.00
density
0.00
values


Trace of ps(Pressure)
0
2
4
6
8
10
12
0
500
1000
1500
2000
MCMC scan
Density of ps(Pressure)
density
0.00
1.33
2.65
3.98
5.30
0.00
values


ps(Temperature)
4
2
0
-2
-4
55
60
65
70
75
80
Temperature


ps(Pressure)
1.5
1.0
0.5
0.0
−0.5
−1.0
50
100
150
200
Pressure

Add random effects to model...

$$\eta_i = \sum_{j=1}^{J} f_j(x_{ij}) + \mathbf{z}_i' \mathbf{u}_{g_i}, \quad \mathbf{u}_1, \ldots, \mathbf{u}_G \overset{\perp}{\sim} N_d(\mathbf{0}, \mathbf{\Sigma}).$$

- $g_i \in \{1, \ldots, G\}$ is group indicator.
- As before, model $f_1(x_1), \ldots, f_J(x_J)$ via penalized B-splines.
- Results in a *generalized linear additive model* (GAMM).
- Can fit in BayesX. Spatial structure can be incorporated.
- **Example**: Ache capuchin monkey hunting in JAGS & BayesX.