Two-variance models, re-expressed to give a simple RL

For the next bit we consider two-variance models, defined as:

- mixed linear models $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}, \text{cov}(\mathbf{u}) = \mathbf{G}, \text{cov}(\boldsymbol{\epsilon}) = \mathbf{R}$
- $\mathbf{R} = \sigma_e^2 \Sigma_e$, $\mathbf{G} = \sigma_s^2 \Sigma_s$
- σ_e^2 and σ_s^2 are unknown
- Σ_e and Σ_s are known and positive definite
  - WLOG set $\Sigma_e = \mathbf{I}_n$ and $\Sigma_s = \mathbf{I}_q$.

We proceed by

- re-parameterizing $(\boldsymbol{\beta}, \mathbf{u})$ to a canonical parameterization, which
- immediately gives the desired simple form for the RL.

Complication: $(\mathbf{X}|\mathbf{Z})$ is often not of full rank $\Rightarrow$ some messiness.

Overview of the following math-choked slides

Here's what the math does:

(1) Derive orthonormal bases for $R(\mathbf{X})$, $R(\mathbf{X}|\mathbf{Z})/R(\mathbf{X})$, and $R(\mathbf{X}|\mathbf{Z})^c$.

These three spaces partition real n-space.

Projecting $\mathbf{y}$ onto these spaces partitions it into info about, respectively,

- the fixed effects;
- the random effects and error mixed together; and
- error.

(2) Pick the basis for $R(\mathbf{X}|\mathbf{Z})/R(\mathbf{X})$ so that the re-parameterized RE has diagonal covariance.

(3) That makes the RL have a simple form, which opens a lot of doors.

Define

- $s_X = \text{rank}(\mathbf{X}) \in \{1, 2, \ldots, p\}$; $s_Z = \text{rank}(\mathbf{X}|\mathbf{Z}) - s_X \in \{1, 2, \ldots, q\}$
- $s_X + s_Z \leq p + q$; assume $s_X, s_Z > 0$.

Define

- $\mathbf{\Gamma}_X$ $n \times s_X$ with columns an orthonomal basis for the col($\mathbf{X}$).
- $\mathbf{\Gamma}_Z$ $n \times s_Z \ni \mathbf{\Gamma}_Z' \mathbf{\Gamma}_X = \mathbf{0}$ and the columns of $(\mathbf{\Gamma}_X | \mathbf{\Gamma}_Z)$ are an orthonormal basis for col($\mathbf{X}|\mathbf{Z}$).
- $\mathbf{\Gamma}_c$ $n \times (n - s_X - s_Z) \ni \mathbf{\Gamma}_c' \mathbf{\Gamma}_X = \mathbf{0}$, $\mathbf{\Gamma}_c' \mathbf{\Gamma}_Z = \mathbf{0}$, and $(\mathbf{\Gamma}_X | \mathbf{\Gamma}_Z | \mathbf{\Gamma}_c)$ is an orthonormal basis for real n-space.

Define $\mathbf{M}$ $(s_X + s_Z) \times (p + q) \ni (\mathbf{X}|\mathbf{Z}) = (\mathbf{\Gamma}_X | \mathbf{\Gamma}_Z)\mathbf{M}$, partitioned as

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{XX} & \mathbf{M}_{XZ} \\ \mathbf{0} & \mathbf{M}_{ZZ} \end{bmatrix} \qquad \begin{matrix} \mathbf{M}_{XX} \ s_X \times p & \quad \mathbf{M}_{XZ} \ s_X \times q \\ \mathbf{0} \ s_Z \times p & \quad \mathbf{M}_{ZZ} \ s_Z \times q \end{matrix}$$

so $\mathbf{X} = \mathbf{\Gamma}_X \mathbf{M}_{XX}$ and $\mathbf{Z} = \mathbf{\Gamma}_X \mathbf{M}_{XZ} + \mathbf{\Gamma}_Z \mathbf{M}_{ZZ}$.

$(\mathbf{X}|\mathbf{Z}) = (\mathbf{\Gamma}_X|\mathbf{\Gamma}_Z)\mathbf{M}$ for

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{XX} & \mathbf{M}_{XZ} \\ \mathbf{0} & \mathbf{M}_{ZZ} \end{bmatrix} \qquad \begin{matrix} \mathbf{M}_{XX}\ s_X \times p & \mathbf{M}_{XZ}\ s_X \times q \\ \mathbf{0}\ s_Z \times p & \mathbf{M}_{ZZ}\ s_Z \times q \end{matrix}$$

Let $\mathbf{M}_{ZZ}$ have SVD $\mathbf{M}_{ZZ} = \mathbf{P}\mathbf{A}^{0.5}\mathbf{L}'$, so $\mathbf{M}_{ZZ}\mathbf{M}'_{ZZ} = \mathbf{P}\mathbf{A}\mathbf{P}'$

- $\mathbf{P}$ $s_Z \times s_Z$ and $\perp$
- $\mathbf{A}$ $s_Z \times s_Z$ and diagonal
- $\mathbf{L}'$ is $s_Z \times q$ with orthonormal rows.

Now re-parameterize the mixed linear model as

$$\begin{aligned} \mathbf{y} = (\mathbf{X}|\mathbf{Z}) \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{u} \end{bmatrix} + \boldsymbol{\epsilon} &= (\mathbf{\Gamma}_X|\mathbf{\Gamma}_Z)\mathbf{M} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{u} \end{bmatrix} + \boldsymbol{\epsilon} \\ &= (\mathbf{\Gamma}_X|\mathbf{\Gamma}_Z\mathbf{P}) \begin{bmatrix} \boldsymbol{\beta}^* \\ \mathbf{v} \end{bmatrix} + \boldsymbol{\epsilon} \end{aligned}$$

- $\boldsymbol{\beta}^* = \mathbf{M}_{XX}\boldsymbol{\beta} + \mathbf{M}_{XZ}\mathbf{u}$ is a fixed effect: Precision$(\boldsymbol{\beta}^*) = \mathbf{0}$
- $\mathbf{v} = \mathbf{A}^{0.5}\mathbf{L}'\mathbf{u}$ is $s_Z \times 1$ with cov$(\mathbf{v}) = \sigma_s^2\mathbf{A}$, diagonal

Deriving the RL from the re-parameterized model

Having re-parameterized the mixed linear model as

$$\mathbf{y} = (\mathbf{\Gamma}_X | \mathbf{\Gamma}_Z \mathbf{P}) \left[\begin{array}{c} \boldsymbol{\beta}^* \\ \mathbf{v} \end{array} \right] + \boldsymbol{\epsilon}, \qquad \mathbf{v} \sim N_{s_Z}(\mathbf{0}, \sigma_s^2 \mathbf{A}), \quad \mathbf{A} \text{ diagonal}, \qquad (1)$$

define $\mathbf{K} = (\mathbf{\Gamma}_Z \mathbf{P} | \mathbf{\Gamma}_c), n \times (n - s_X)$; pre-multiply (1) by $\mathbf{K}'$ to give

$$\mathbf{K}'\mathbf{y} = \left[\begin{array}{c} \mathbf{v} \\ \mathbf{0}_{(n-s_X-s_Z)\times 1} \end{array} \right] + \boldsymbol{\xi}, \qquad \boldsymbol{\xi} \sim N(\mathbf{0}, \sigma_e^2 \mathbf{I}_{n-s_X})$$

So

- $\mathbf{P}'\mathbf{\Gamma}_Z' y = \mathbf{v} + \xi_1 \sim N(\mathbf{0}, \sigma_s^2 \mathbf{A} + \sigma_e^2 \mathbf{I}_{s_Z})$, independent of
- $\mathbf{\Gamma}_c' y = \xi_2 \sim N(\mathbf{0}, \sigma_e^2 \mathbf{I}_{n-s_X-s_Z})$

The RL is the likelihood for (σ_s^2, σ_e^2) arising from the transformed data

- $\mathbf{P}'\mathbf{\Gamma}_Z' y = \mathbf{v} + \xi_1 \sim N_{s_Z}(\mathbf{0}, \sigma_s^2 \mathbf{A} + \sigma_e^2 \mathbf{I}_{s_Z})$, independent of
- $\mathbf{\Gamma}_c' y = \xi_2 \sim N(\mathbf{0}, \sigma_e^2 \mathbf{I}_{n-s_X-s_Z})$

Specifically,

$$\begin{aligned}
\log RL(\sigma_s^2, \sigma_e^2 | \mathbf{y}) &= B - \frac{n - s_X - s_Z}{2} \log(\sigma_e^2) - \frac{1}{2\sigma_e^2} \mathbf{y}' \mathbf{\Gamma}_c \mathbf{\Gamma}_c' \mathbf{y} \\
&\quad - \frac{1}{2} \sum_{j=1}^{s_Z} \left[\log(\sigma_s^2 a_j + \sigma_e^2) + \frac{\hat{v}_j^2}{\sigma_s^2 a_j + \sigma_e^2} \right],
\end{aligned}$$

for $\hat{\mathbf{v}} = (\hat{v}_1, \ldots, \hat{v}_{s_Z})' = \mathbf{P}'\mathbf{\Gamma}_Z' \mathbf{y}$, a known function of $\mathbf{y}$.

Examining the restricted likelihood

$$\log RL(\sigma_s^2, \sigma_e^2|\mathbf{y}) = B - \frac{n - s_X - s_Z}{2}\log(\sigma_e^2) - \frac{1}{2\sigma_e^2}\mathbf{y}'\mathbf{\Gamma}_c\mathbf{\Gamma}_c'\mathbf{y} \quad (2)$$

$$-\frac{1}{2}\sum_{j=1}^{s_Z}\left[\log(\sigma_s^2 a_j + \sigma_e^2) + \frac{\hat{v}_j^2}{\sigma_s^2 a_j + \sigma_e^2}\right], \quad (3)$$

for $\hat{\mathbf{v}} = (\hat{v}_1, \ldots, \hat{v}_{s_Z})' = \mathbf{P}'\mathbf{\Gamma}_Z'\mathbf{y}$, a known function of $\mathbf{y}$.

Eq'n (2) are the *free terms* for σ_e^2; they

- are a function of σ_e^2 but not σ_s^2
- use $\mathbf{y}$ only through $\mathbf{y}'\mathbf{\Gamma}_c\mathbf{\Gamma}_c'\mathbf{y}$, the residual sum of squares from an unshrunk fit of $\mathbf{y}$ on $(\mathbf{X}|\mathbf{Z})$

Eq'n (3) are the *mixed terms* for σ_s^2; they

- are a function of both σ_e^2 and σ_s^2
- use $\mathbf{y}$ through $\hat{\mathbf{v}}$, the estimate of $\mathbf{v}$ from the unshrunk fit.

The RL is the likelihood from a particular GLM

Specifically, a GLM with gamma errors, identity link, and:

GLM notation	j^{th} mixed term $j = 1, \ldots, s_Z$	Free terms $j = s_Z + 1$
Data y_i	$\hat{v}_j^2$	$\hat{v}_{s_Z+1}^2 = \mathbf{y}'\mathbf{\Gamma}_c\mathbf{\Gamma}_c'\mathbf{y}/(n - s_X - s_Z)$
Canonical parameter θ_i	$-1/(\sigma_s^2 a_j + \sigma_e^2)$	$-1/\sigma_e^2$
Shape parameter ν_i	$1/2$	$(n - s_X - s_Z)/2$
$E(y_i) = -1/\theta_i$	$\sigma_s^2 a_j + \sigma_e^2$	σ_e^2
$\text{Var}(y_i) = [E(y_i)]^2/\nu_i$	$2(\sigma_s^2 a_j + \sigma_e^2)^2$	$2(\sigma_e^2)^2/(n - s_X - s_Z)$

GLMs provide a lot of tools (residuals, case influence, etc.).

Alternative derivation: The RL as a marginal posterior

Begin with the re-parameterized mixed linear model

$$\mathbf{y} = (\boldsymbol{\Gamma}_X | \boldsymbol{\Gamma}_Z \mathbf{P}) \begin{bmatrix} \boldsymbol{\beta}^* \\ \mathbf{v} \end{bmatrix} + \boldsymbol{\epsilon}, \qquad \mathbf{v} \sim N_{s_Z}(\mathbf{0}, \sigma_s^2 \mathbf{A}), \quad \mathbf{A} \text{ diagonal.}$$

Pre-multiply both sides by the $\perp$ matrix $(\boldsymbol{\Gamma}_X | \boldsymbol{\Gamma}_Z \mathbf{P} | \boldsymbol{\Gamma}_c)'$ to give

$$\begin{bmatrix} \boldsymbol{\Gamma}_X' \\ \mathbf{P}'\boldsymbol{\Gamma}_Z' \\ \boldsymbol{\Gamma}_c' \end{bmatrix} \mathbf{y} = \begin{bmatrix} \boldsymbol{\beta}^* \\ \mathbf{v} \\ \mathbf{0}_{(n-s_X-s_Z)\times 1} \end{bmatrix} + \boldsymbol{\epsilon}, \tag{4}$$

The distribution of $\boldsymbol{\epsilon}$ is unchanged

Let $\pi(\sigma_e^2, \sigma_s^2)$ be the prior distribution for (σ_e^2, σ_s^2)

The joint posterior distribution of $(\boldsymbol{\beta}^*, \mathbf{v}, \sigma_e^2, \sigma_s^2)$ is easily shown to be ...

$$\pi(\boldsymbol{\beta}^*, \mathbf{v}, \sigma_e^2, \sigma_s^2 | \mathbf{y}) \propto \pi(\sigma_e^2, \sigma_s^2)$$

$$(\sigma_e^2)^{-s_X/2} \exp\left(-(\boldsymbol{\beta}^* - \boldsymbol{\Gamma}_X'\mathbf{y})'(\boldsymbol{\beta}^* - \boldsymbol{\Gamma}_X'\mathbf{y})/2\sigma_e^2\right) \tag{5}$$

$$\prod_{j=1}^{s_Z} \left(\sigma_e^2 \frac{a_j}{a_j + r}\right)^{-0.5} \exp\left(-\sum_{j=1}^{s_Z} \left(2\sigma_e^2 \frac{a_j}{a_j + r}\right)^{-1} (v_j - \tilde{v}_j)^2\right) \tag{6}$$

$$(\sigma_e^2)^{-(n-s_X-s_Z)/2} \exp\left(-\mathbf{y}'\boldsymbol{\Gamma}_c\boldsymbol{\Gamma}_c'\mathbf{y}/2\sigma_e^2\right) \tag{7}$$

$$\prod_{j=1}^{s_Z} \left(\sigma_s^2 a_j + \sigma_e^2\right)^{-0.5} \exp\left(-\sum_{j=1}^{s_Z} \hat{v}_j^2/2(\sigma_s^2 a_j + \sigma_e^2)\right), \tag{8}$$

where $\tilde{v}_j = \frac{a_j}{a_j+r}\hat{v}_j$ and $r = \sigma_e^2/\sigma_s^2$.

Eq'n (5) is $\pi(\boldsymbol{\beta}^*|\mathbf{y}, \sigma_e^2, \sigma_s^2)$; Eq'n (6) is $\pi(\mathbf{v}|\mathbf{y}, \sigma_e^2, \sigma_s^2)$; (7) + (8) is the RL.

(6): $v_j|\sigma_e^2, \sigma_s^2 \sim$ indep't $N(\tilde{v}_j, \sigma_e^2 a_j/(a_j + r))$, with DF $a_j/(a_j + r)$.

Thus given $r = \sigma_e^2/\sigma_s^2$, v_j is shrunk more for j with smaller a_j.

Why this re-expression is cool

The $\hat{v}_j$ are known linear functions of $\mathbf{y}$; if $\boldsymbol{\beta}$ includes an intercept, they are data contrasts.

$\hat{v}_j|\sigma_e^2, \sigma_s^2 \sim$ indep't $N(0, \sigma_s^2 a_j + \sigma_e^2)$; that's why the RL decomposes.

The a_j determine how $\hat{v}_j$ and $\mathbf{y}'\boldsymbol{\Gamma}_c\boldsymbol{\Gamma}_c'\mathbf{y}$ inform about σ_e^2 and σ_s^2.

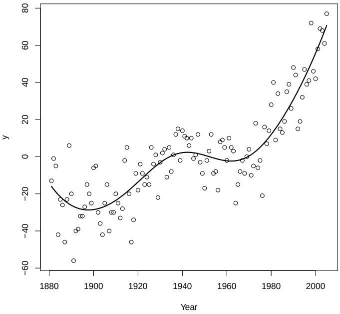
- The a_j are (obscure) functions of $\mathbf{X}$ and $\mathbf{Z}$, explored below.

$v_j|\mathbf{y}, \sigma_e^2, \sigma_s^2 \sim$ indep't normal with mean 0, simple variance and DF.

By assumption $s_Z > 0$; thus σ_e^2 and σ_s^2 are identified $\Leftrightarrow$ either
(a) $\exists$ free terms (i.e., $n - s_X - s_Z > 0$) or (b) $\exists \geq 2$ distinct a_j.

Recall: Penalized spline fit to the GMST data

$n = 125$, truncated quadratic basis, 30 knots at years $1880 + \{4, \ldots, 120\}$



$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^p \\ & \vdots & \\ 1 & x_n & x_n^2 \end{bmatrix} \qquad \mathbf{Z} = \begin{bmatrix} (x_1 - \kappa_1)_+^2 & \ldots & (x_1 - \kappa_K)_+^2 \\ & \vdots & \\ (x_n - \kappa_1)_+^2 & \ldots & (x_n - \kappa_K)_+^2 \end{bmatrix}$$

$s_X = 3$, $s_Z = 30$, though just barely

Recall: We re-parameterized the mixed linear model as

$$\mathbf{y} = (\boldsymbol{\Gamma}_X | \boldsymbol{\Gamma}_Z \mathbf{P}) \begin{bmatrix} \boldsymbol{\beta}^* \\ \mathbf{v} \end{bmatrix} + \epsilon, \quad \mathbf{v} \sim N_{s_Z}(\mathbf{0}, \sigma_s^2 \mathbf{A}), \quad \mathbf{A} = \text{diag}\{a_j\},$$

Here are the a_j, in decreasing order:

36.0	3.15	0.562	0.147	0.0493	0.0195
8.76e-3	4.32e-3	2.30e-3	1.30e-3	7.68e-4	4.75e-4
3.05e-4	2.01e-4	1.37e-4	9.51e-5	6.76e-5	4.89e-5
3.61e-5	2.71e-5	2.06e-5	1.60e-5	1.26e-5	1.01e-5
8.32e-6	7.00e-6	6.06e-6	5.42e-6	5.06e-6	3.75e-6

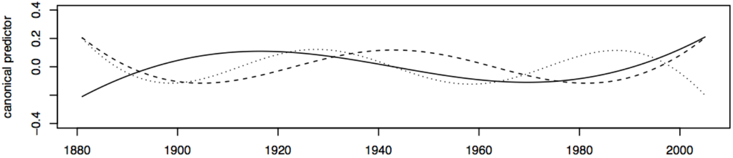
The a_j decline quickly: $a_1/a_6 = 1841$, the last 18 $a_1/a_j < 10^5$

Later we'll see this implies

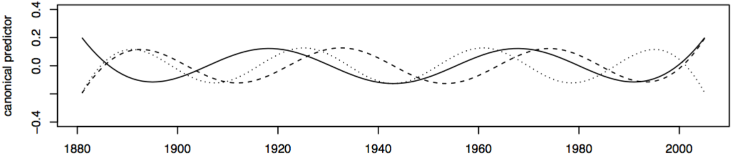
- the first few $\hat{v}_j$ are almost all of the data's info about σ_s^2.
- the remaining $\hat{v}_j$ are almost exclusively about σ_e^2.

Here are the columns of $\mathbf{\Gamma}_Z\mathbf{P}$ that go with selected a_j:

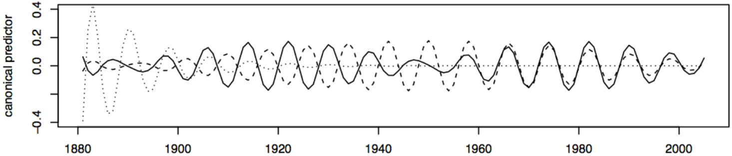
For $a_1 = 36.0$ (solid), $a_2 = 3.15$ (dashed), $a_3 = 0.562$ (dotted)



For $a_4 = 0.147$ (solid), $a_5 = 0.049$ (dashed), $a_6 = 0.0195$ (dotted)



For $a_{28} = 5.42\text{e-}6$ (solid), $a_{29} = 5.06\text{e-}6$ (dashed), $a_{30} = 3.75\text{e-}6$ (dotted)



This penalized spline can be understood as

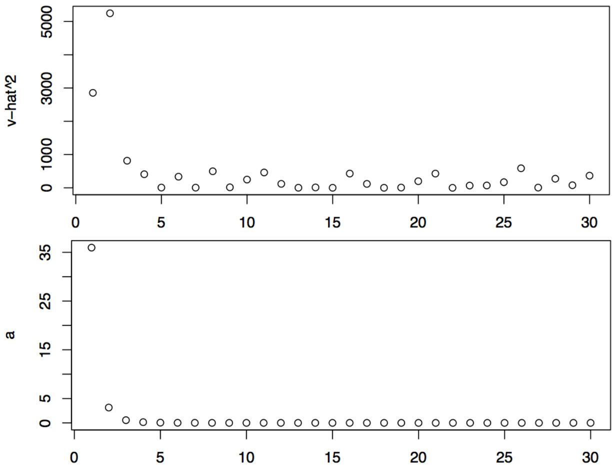
- a quadratic regression with unshrunk coefficients PLUS
- a regression on higher-order polynomials with shrunken coefficients
- WHERE the extent of shrinkage increases with the polynomial order.

What controls shrinkage:

- σ_s^2 controls shrinkage of all coefficients
- the a_j control the *relative* degrees of shrinkage of different v_j
- v_j with smaller a_j are shrunk more; broadly, variation in those directions is mostly treated as error.

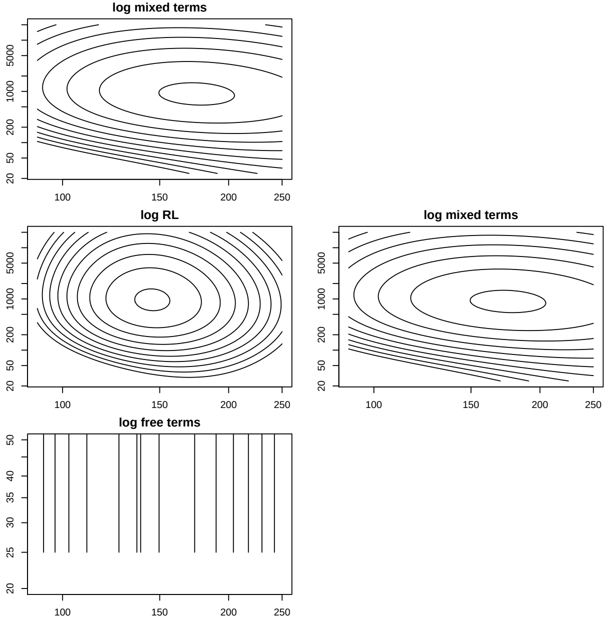
This appears to generalize for splines with truncated polynomial bases.

The RL is a gamma regression of $\hat{v}_j^2$ on a_j with slope σ_s^2 and intercept σ_e^2. Here are plots vs. j of $\hat{v}_j^2$ (top) and a_j (bottom).



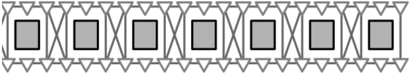
For large j, the $\hat{v}_j^2$ are telling you about the intercept σ_e^2.

Contributions to the log RL of free and mixed terms (1 log contours)



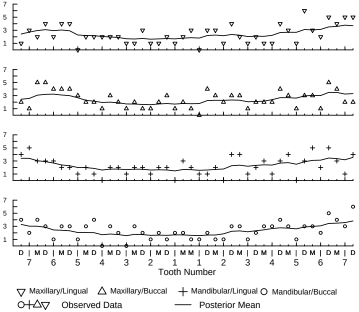
Example: Simple ICAR model for periodontal data

ICAR with these neighbor pairs for $n = 168$ sites.



Priors: Flat on the two island (arch) means; σ_e^2 and $\sigma_s^2 \sim$ IG(0.01,0.01).

Posterior medians: σ_e^2 1.25, σ_s^2 0.25, σ_e^2/σ_s^2 4.0 – very smooth fit.



Re-expressing this ICAR model

Recall that the ICAR's precision matrix is $\mathbf{Q}/\sigma_s^2$, where

- Q_{ii} = number of region i's neighbors
- Q_{ij} = -1 if $i \sim j$ and 0 otherwise.

Spectral decomposition: $\mathbf{Q} = \mathbf{V}$ diag$(d_1, \ldots, d_{166}, 0, 0)$ $\mathbf{V}'$

- $d_1 \geq \cdots \geq d_{166} > 0$, $\mathbf{V} = (\mathbf{V}_1 | \mathbf{V}_2)$ where
- $\mathbf{V}_2$ has $s_X = 2$ columns, one for each arch (island)
- $\mathbf{V}_1$ has $s_Z = 166$ columns.

In the re-expression,

- $\mathbf{\Gamma}_X = \mathbf{V}_2$, the two arch (island) means are the fixed effects;
- $\mathbf{P} = \mathbf{I}_{166}$, $\mathbf{\Gamma}_Z = \mathbf{V}_1[, 166 : 1]$;
- $a_j = 1/d_{167-j}$ so $a_1 \geq \cdots \geq a_{166} > 0$.

Here are the a_j (multiplicities are even because the 2 arches are identical):

a_j	multiplicity	# distinct a_j
149.0	2	1
37.33	2	1
16.64	2	1
9.40 to 1	2	11
0.843	4	1
0.695	24	1
0.672 to 0.288	2 or 4	15
0.200	24	1
0.1996 to 0.1800	2 or 4	14
0.1798	24	1

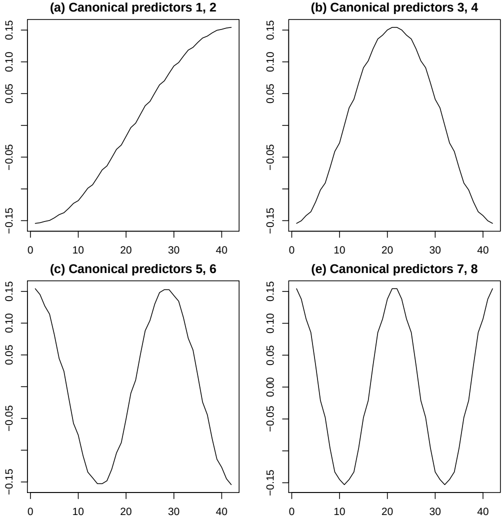
a_1/a_6: p-spline $1,841$; ICAR 9 ($a_1/a_{12} = 35$).

a_1/a_{15}: p-spline $263,083$; ICAR, 61 ($a_1/a_{30} = 177$).

$a_1/a_{smallest}$: p-spline $9,593,165$; ICAR 829.

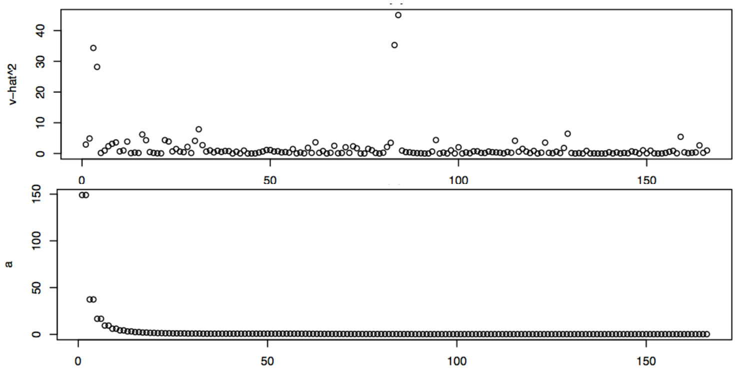
$\Rightarrow$ shrinkage of the v_j is much less differentiated for the ICAR.

Columns of canonical predictors $\mathbf{\Gamma}_Z \mathbf{P} = \mathbf{\Gamma}_Z$ (for one side of one arch)



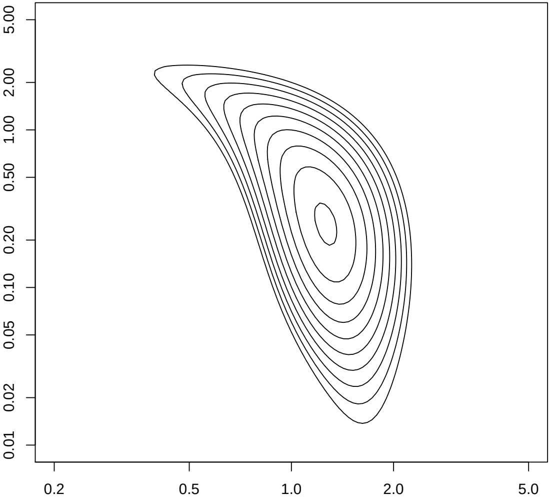
These are similar to the p-spline's canonical regressors.

The RL is a gamma regression of $\hat{v}_j^2$ on a_j with slope σ_s^2 and intercept σ_e^2.
Here are plots vs. j of $\hat{v}_j^2$ (top) and a_j (bottom).



If this model fits, the $\hat{v}_j^2$ should generally decline as j increases.
The outliers $\hat{v}_{83}, \hat{v}_{84}$ contrast direct vs interprox sites (one per arch).

Contour plot of log RL (1 log contours) – mixed terms only, but not bad.



σ_s^2 on vertical, σ_e^2 on horizontal.

Other examples in the book

Spatial confounding with the ICAR model

- This machinery gives some insight into
  - how the data determine σ_s^2, σ_e^2, and thus
  - which true underlying models can produce spatial confounding.

Dynamic linear model with one quasi-cyclic component

- The a_j decline like the ICAR's, not like the p-spline's.
- Canonical predictors look like superpositions of pairs of sine curves.
- Broadly, the canonical predictors' frequency increases as a_j decreases.
- I use fake data to illustrate how the $\hat{v}_j^2$ can show lack of fit.

A tentative collection of tools

Tools from generalized linear models.

- I use residuals, measures of leverage, case influence.
- I haven't used deviance, though it might be useful.

The canonical observations $\hat{v}_j$: mean 0, variance $\sigma_s^2 a_j + \sigma_e^2$.

A modified restricted likelihood that omits $j > m$;

- This helps show which mixed terms inform about which variance.

DF in the fit for v_j: $a_j/(a_j + r)$ for $r = \sigma_e^2/\sigma_s^2$.

- This is the contribution $(\mathbf{\Gamma}_Z \mathbf{P})_j$ makes to the fit.
- Larger DF indicate a greater contribution.