PENALIZED METHODS for functional inference

- The idea of "penalty is ad-hoc
- It does not arise "naturally" in classical inference
- It appears very naturally in Bayesian inference
  - ➔ L_2 penalty: equivalent to Gaussian prior
  - ➔ L_1 penalty: equivalent to double exponential prior

The concept of penalized likelihood (example in the mixed linear model)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

$$\mathbf{y}|\boldsymbol{\beta}, \mathbf{u}, \mathbf{R} \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \mathbf{R})$$

$$\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$$

$$p(\mathbf{y}|\boldsymbol{\beta}, \mathbf{u}, \mathbf{R}) = \frac{1}{(2\pi)^{\frac{N}{2}}|\mathbf{R}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})\right]$$

$$p(\mathbf{u}|\mathbf{G}) = \frac{1}{(2\pi)^{\frac{q}{2}}|\mathbf{G}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\mathbf{u}'\mathbf{G}^{-1}\mathbf{u}\right]$$

Assuming known variance components, the log of the joint density of the data and random effects is termed "penalized likelihood

$$l(\boldsymbol{\beta},\mathbf{u}|\mathbf{y},\mathbf{R},\mathbf{G}) = K - \tfrac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u}) - \tfrac{1}{2}\mathbf{u}'\mathbf{G}^{-1}\mathbf{u}$$

$$-2l(\beta,u|y,R,G) = K + (y-X\beta-Zu)'(y-X\beta-Zu) + u'G^{-1}u$$

Penalized SS

$$\frac{\partial l(\boldsymbol{\beta},\mathbf{u}|\mathbf{y},\mathbf{R},\mathbf{G})}{\partial \boldsymbol{\beta}} = \mathbf{X}'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})$$

$$\frac{\partial l(\boldsymbol{\beta},\mathbf{u}|\mathbf{y},\mathbf{R},\mathbf{G})}{\partial \mathbf{u}} = \mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u}) - \mathbf{G}^{-1}\mathbf{u}$$

Setting the derivatives to 0 yields

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

- The solution to these equations produces the "maximum penalized likelihood" estimates of **β** and **u**
- These solutions are also the BLUE(**β**) and BLUP(**u**)

8. Reproducing Kernel Hilbert spaces mixed model

Function of molecular information **x (vector of SNP variables)**

$$SS[g(\mathbf{x}),\lambda] = \sum_{i=1}^{n}[y_i - \mathbf{w}_i'\boldsymbol{\beta} - \mathbf{z}_i'\mathbf{u} - g(x_i)]^2 + \lambda||g(\mathbf{x})||_H^2$$

Smoothing parameter (λ)

"Penalized sum of squares"

Some norm under Hilbert space (*H*) of functions

Variational problem: find ***g(x)*** over entire space of functions minimizing SS(.)

Solution to variational problem: linear function

No. individuals with molecular data

$$g(.) = \alpha_0 + \sum_{j=1}^{n} \alpha_j K(., \mathbf{x}_j)$$

reduction of dimension
p (# SNPs)➔ # indiv.

Regression coefficient

Reproducing kernel

Example of reproducing kernel:

$$K_h(\mathbf{x}, \mathbf{x}_j) = \exp\left[-\frac{(\mathbf{x}-\mathbf{x}_j)'(\mathbf{x}-\mathbf{x}_j)}{h} \right]$$

In an Euclidean space of dimension n, the dot product between vectors $\mathbf{v}$ and $\mathbf{w}$ is $\sum_{i=1}^{n} v_i w_i$,

and the norm is $\|\mathbf{v}\| = \sqrt{\sum_{i=1}^{n} v_i^2}$

Inner product generalizes dot product to vectors of infinite dimension. For instance, in a vector space of real functions with domain $[a, b]$, the inner product is

$$< g_1, g_2 >= \int_a^b g_1(x) g_2(x) dx,$$

$$\|g_1\| = \sqrt{\int_a^b g_1(x)^2 dx}$$

IF *x* is a random variable with pdf *p(x)*

$$< g_1, g_2 >= \int_a^b g_1(x) g_2(x) p(x) dx = E[g_1(x) g_2(x)]$$

➔Definition of positive-definite kernel (the theory deals with "reproducing kernels) function

$$\int k(\mathbf{x},\mathbf{t})g(\mathbf{x})g(\mathbf{t})p(\mathbf{x},\mathbf{t})d\mathbf{x}d\mathbf{t} > 0$$

➔Positive-definite kernel matrix; symmetric, with *k(i,j,h)=k(j,i,h)*

$$\mathbf{K}_h = \begin{bmatrix} k(1,1,\mathbf{h}) & k(1,2,\mathbf{h}) & \cdots & k(1,n,\mathbf{h}) \\ k(2,1,\mathbf{h}) & k(2,2,\mathbf{h}) & \cdots & k(2,n,\mathbf{h}) \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ k(n,1,\mathbf{h}) & k(n,2,\mathbf{h}) & \cdots & k(n,n,\mathbf{h}) \end{bmatrix}$$

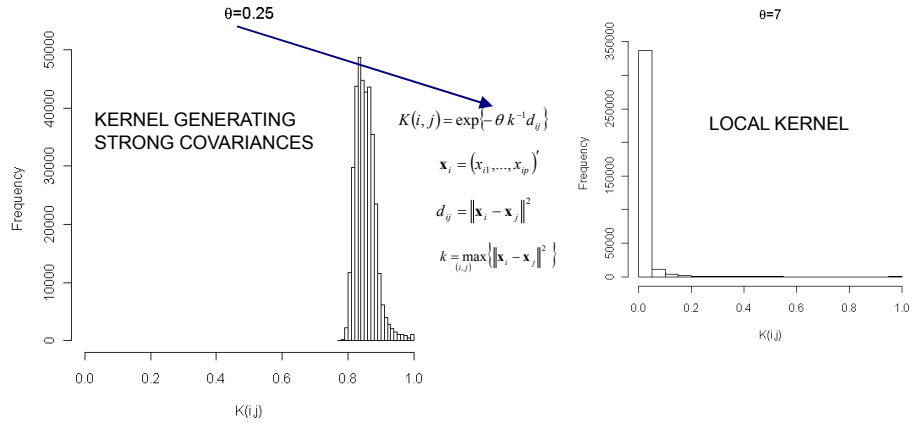
h= scalar or vector of bandwidth parameters

MEASURES OF DISTANCE THAT CAN BE USED IN KERNELS

Euclidean $$d(x,y) = \| x - y \| = \sqrt{\sum_{k=1}^{p} (x_k - y_k)^2}$$

Manhattan $$d(x,y) = \sum_{k=1}^{p} | x_k - y_k |,$$

Bray-Curtis $$d_{ij} = (\Sigma_k \mid x_{ik} - x_{jk} \mid)/(\Sigma_k x_{ik} + x_{jk})$$



Histogram of evaluations of Gaussian kernel by value of bandwidth parameter

Mixed model representation (enhancing pedigrees…)

$$y_i = \mathbf{w}_i'\boldsymbol{\beta} + \mathbf{z}_i'\mathbf{u} + \sum_{j=1}^{n} \exp\left[-\frac{(\mathbf{x}_i-\mathbf{x}_j)'(\mathbf{x}_i-\mathbf{x}_j)}{h}\right]\alpha_j + e_i$$

Define row vector

$$\mathbf{t}_i'(h) = \left\{\exp\left[-\frac{(\mathbf{x}_i-\mathbf{x}_j)'(\mathbf{x}_i-\mathbf{x}_j)}{h}\right]\right\}$$

$$\mathbf{T}(h) = \begin{bmatrix} \mathbf{t}_1'(h) \\ \mathbf{t}_2'(h) \\ . \\ \mathbf{t}_n'(h) \end{bmatrix}$$

t'$_i$(h)= k'$_i$(h)

T(h)=K(h)

Then:

Bandwidth parameter

$$\mathbf{y} = \mathbf{W}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{T}(h)\boldsymbol{\alpha} + \mathbf{e}$$

$$\sigma_\alpha^2 = \frac{1}{\lambda}$$

Do:

$$\boldsymbol{\alpha} \sim \mathbf{N}(\mathbf{0}, \mathbf{T}^{-1}(h)\sigma_\alpha^2)$$

Smoothing parameter

$$\begin{bmatrix} \mathbf{W}'\mathbf{W} & \mathbf{W}'\mathbf{Z} & \mathbf{W}'\mathbf{T}(h) \\ \mathbf{Z}'\mathbf{W} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\frac{\sigma_e^2}{\sigma_u^2} & \mathbf{Z}'\mathbf{T}(h) \\ \mathbf{T}'(h)\mathbf{W} & \mathbf{T}'(h)\mathbf{Z} & \mathbf{T}'(h)\mathbf{T}(h) + \mathbf{T}(h)\frac{\sigma_e^2}{\sigma_\alpha^2} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \\ \hat{\boldsymbol{\alpha}} \end{bmatrix} = \begin{bmatrix} \mathbf{W}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{T}'(h)\mathbf{y} \end{bmatrix}$$

h assumed known here

Penalized estimation

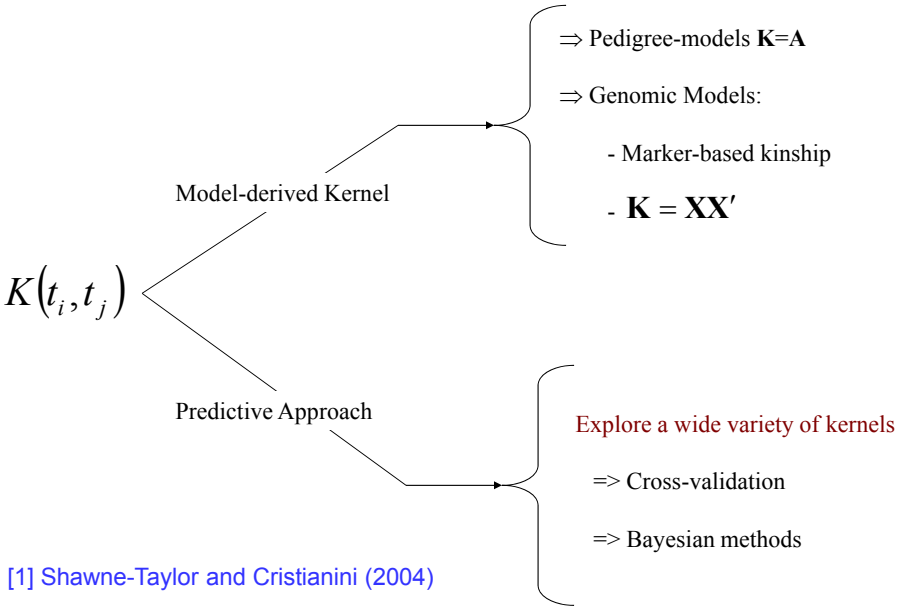
$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \left\{ (\mathbf{y} - \mathbf{K}\boldsymbol{\alpha})'(\mathbf{y} - \mathbf{K}\boldsymbol{\alpha}) + \lambda\boldsymbol{\alpha}'\mathbf{K}\boldsymbol{\alpha} \right\}$$

Bayesian View

$$\begin{cases} \mathbf{y} = \mathbf{K}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \\ p(\boldsymbol{\varepsilon}, \boldsymbol{\alpha}) = N(\boldsymbol{\varepsilon}|\mathbf{0}, \mathbf{I}\sigma_\varepsilon^2) N(\boldsymbol{\alpha}|\mathbf{0}, \mathbf{K}^{-1}\sigma_\alpha^2) \end{cases}$$

[1] Kimeldorf, G.S. & Wahba, G. (1970).

How to Choose the Reproducing Kernel? [1]



[1] Shawne-Taylor and Cristianini (2004)

THE "ANIMAL MODEL" IS A PARTICULAR CASE OF RKHS

$$
\begin{aligned}
y &= A\alpha + e \\
\alpha &\sim N(0, A^{-1}\sigma_a^2) \\
e &\sim N(0, I\sigma_e^2) \\
\Rightarrow u &= A\alpha \sim N(0, A\sigma_a^2)
\end{aligned}
$$

*Use **A** as kernel matrix*

$$
\begin{aligned}
\left(A'A + A\frac{\sigma_e^2}{\sigma_a^2}\right)\hat{\alpha} &= A'y \\
A\left(A + I\frac{\sigma_e^2}{\sigma_a^2}\right)\hat{\alpha} &= Ay \\
\hat{\alpha} &= \left(A + I\frac{\sigma_e^2}{\sigma_a^2}\right)^{-1} y \\
A\hat{\alpha} &= \left(I + A^{-1}\frac{\sigma_e^2}{\sigma_a^2}\right)^{-1} y = \text{BLUP(additive effects)}
\end{aligned}
$$

Predicted Genetic signal

GENOMIC BLUP IS A PARTICULAR CASE OF RKHS

$$y = XX'\alpha + e$$

$$\alpha \sim N\left(0, (XX')^{-1}\sigma^2_\beta\right)$$

$$e \sim N(0, I\sigma^2_e)$$

$$\Rightarrow u = XX'\alpha \sim N(0, XX'\sigma^2_\beta)$$

$$\left(XX'XX' + XX'\frac{\sigma^2_e}{\sigma^2_\beta}\right)\hat{\alpha} = XX'y$$

$$(XX')\left(XX' + I\frac{\sigma^2_e}{\sigma^2_\beta}\right)\hat{\alpha} = XX'y$$

$$\hat{\alpha} = \left(XX' + I\frac{\sigma^2_e}{\sigma^2_\beta}\right)^{-1} y$$

Predicted Genetic signal

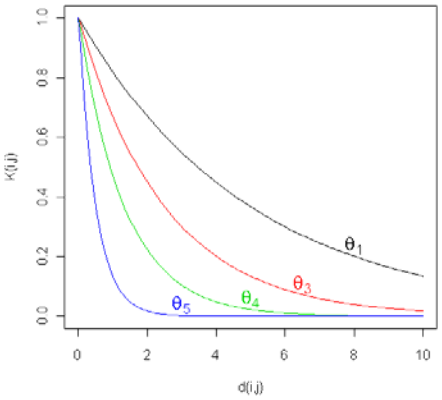
$$XX'\hat{\alpha} = XX'\left(XX' + I\frac{\sigma^2_e}{\sigma^2_\beta}\right)^{-1} y$$

$$\left(I + (XX')^{-1}\frac{\sigma^2_e}{\sigma^2_\beta}\right)^{-1} y = \text{"GENOMIC BLUP"}$$

Choosing the RK based on predictive ability

$d(\mathbf{x}_i, \mathbf{x}_j)$:
(genetic) distance between individuals

$$\Rightarrow \quad K(i,j|\theta) = Exp\left\{ -\theta \times d(\mathbf{x}_i, \mathbf{x}_j) \right\}$$



Strategies

- Grid of Values of θ + CV
- Fully Bayesian: assign a prior to θ (computationally demanding)
- Kernel Averaging [1]

$$K(i,j) = \alpha_1 K(i,j|\theta_1) + (1-\alpha_1)K(i,j|\theta_5)$$

[1] de los Campos et al. (2010) Genetics Research

Example 1 of RKHS

$$\begin{bmatrix} y_2 = 5 \\ y_3 = 3 \\ y_4 = 7 \\ y_5 = 8 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \left(\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} + \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \end{bmatrix} \right) + \begin{bmatrix} e_2 \\ e_3 \\ e_4 \\ e_5 \end{bmatrix}$$

$$= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}(\mathbf{a} + \mathbf{d}) + \mathbf{e}.$$

Additive Dominance

Henderson (1985) assumed $\sigma_a^2 = 5, \sigma_d^2 = 4$ and $\sigma_e^2 = 20$

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 1 & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 1 & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{4} & \frac{1}{4} & 1 \end{bmatrix} \text{ and } \mathbf{D} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{4} & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Application of BLUP paradigm leads to

$$\hat{\boldsymbol{\beta}}' = \begin{bmatrix} 5.145 & 0.241 \end{bmatrix},$$
$$\hat{\mathbf{a}}' = \begin{bmatrix} 0.045 & -0.192 & -0.343 & 0.096 & 0.242 \end{bmatrix},$$
$$\hat{\mathbf{d}}' = \begin{bmatrix} 0 & -0.073 & -0.365 & 0.162 & 0.234 \end{bmatrix}.$$

$$\hat{\mathbf{g}} = \hat{\mathbf{a}} + \hat{\mathbf{d}} = \begin{bmatrix} 0.045 & -0.265 & -0.708 & 0.259 & 0.477 \end{bmatrix}$$

Next, do RKHS with K=A+D as positive-definite kernel matrix

$$\mathbf{K} = \mathbf{A} + \mathbf{D} = \begin{bmatrix} 2 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & 2 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 2 & \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} & \frac{3}{4} & 2 & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{4} & \frac{1}{4} & 2 \end{bmatrix}$$

$$\begin{bmatrix} y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} 2 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 2 & \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{3}{4} & 2 & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{1}{4} & 2 \end{bmatrix} \begin{bmatrix} \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{bmatrix} + \begin{bmatrix} e_2 \\ e_3 \\ e_4 \\ e_5 \end{bmatrix}$$

$$= \mathbf{X}\boldsymbol{\beta} + \mathbf{K}\boldsymbol{\alpha} + \mathbf{e}.$$

$\sigma^2_{\alpha} = \sigma^2_a + \sigma^2_d = 9$ ➔ This is 1/λ

$$\begin{bmatrix} \hat{\beta}_0 = 5.289 & \hat{\beta}_1 = 0.200 & \hat{\alpha}_2 = -0.128 & \hat{\alpha}_3 = -0.781 & \hat{\alpha}_4 = 0.487 & \hat{\alpha}_5 = 0.422 \end{bmatrix}$$

$$\begin{bmatrix} \hat{g}_{K,1} \\ \hat{g}_{K,2} \\ \hat{g}_{K,3} \\ \hat{g}_{K,4} \\ \hat{g}_{K,5} \end{bmatrix} = \begin{bmatrix} 0.036 \\ -0.210 \\ -0.569 \\ 0.206 \\ 0.382 \end{bmatrix}$$

COMPARED WITH

$$\hat{\mathbf{g}} = \hat{\mathbf{a}} + \hat{\mathbf{d}} = \begin{bmatrix} 0.045 & -0.265 & -0.708 & 0.259 & 0.477 \end{bmatrix}$$

PREDICTING FUTURE RECORDS UNDER THE SAME ENVIRONMENTAL CONDITIONS; PARAMETRICALLY

$$\begin{bmatrix} y^f_1 \\ y^f_2 \\ y^f_3 \\ y^f_4 \\ y^f_5 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 3 \\ 1 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} + \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \end{bmatrix} + \begin{bmatrix} e^f_1 \\ e^f_2 \\ e^f_3 \\ e^f_4 \\ e^f_5 \end{bmatrix}$$

$$= \mathbf{M}_P\boldsymbol{\theta}_P + \mathbf{e}^f,$$

PREDICTION OF FUTURE RECORDS NON-PARAMETRICALLY

$$\begin{bmatrix} y_1^f \\ y_2^f \\ y_3^f \\ y_4^f \\ y_5^f \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 3 \\ 1 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 2 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 2 & \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{3}{4} & 2 & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{1}{4} & 2 \end{bmatrix} \begin{bmatrix} \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{bmatrix} + \begin{bmatrix} e_1^f \\ e_2^f \\ e_3^f \\ e_4^f \\ e_5^f \end{bmatrix}$$

$$= \mathbf{M}_K \boldsymbol{\theta}_K + \mathbf{e}^f.$$

FOR BOTH APPROACHES THE PREDICTIVE DISTRIBUTION IS

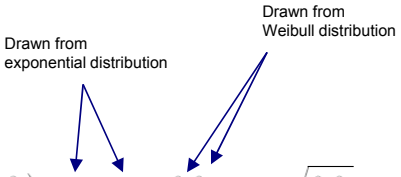
$$\begin{bmatrix} y_1^f \\ y_2^f \\ y_3^f \\ y_4^f \\ y_5^f \end{bmatrix} \Big| \begin{bmatrix} y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix}, \text{dispersion (smoothing) parameters}$$

$$\sim \left(\mathbf{M}_{\cdot} \hat{\boldsymbol{\theta}}_{\cdot}, (\mathbf{M}_{\cdot} \mathbf{C}_{\cdot}^{-1} \mathbf{M}_{\cdot}' + \mathbf{I}_f) \sigma_e^2 \right),$$

For the two procedures the mean and SD of the predictive distributions are:

$$P = \begin{bmatrix} 5.674 \pm 6.020 \\ 5.364 \pm 5.460 \\ 5.162 \pm 5.353 \\ 5.646 \pm 5.834 \\ 6.828 \pm 6.115 \end{bmatrix}; K = \begin{bmatrix} 5.754 \pm 5.576 \\ 5.286 \pm 5.659 \\ 4.735 \pm 5.561 \\ 5.919 \pm 5.940 \\ 7.061 \pm 6.157 \end{bmatrix}$$

Example 2 of RKHS



$$E\left(y|\alpha_i,\alpha_j,\beta_i,\beta_j\right)=\alpha_i+\alpha_j+\beta_i\beta_j+\alpha_i\alpha_j\sqrt{\beta_i\beta_j}, \tag{21}$$

where α_i (β_i) and α_j $\left(\beta_j\right)$ are effects of alleles i and j at the $\alpha\,(\beta)$ locus. The system is non-linear on allelic effects, as indicated by the first derivatives of the conditional expectation function with respect to the $\alpha's$ or $\beta's$. For instance

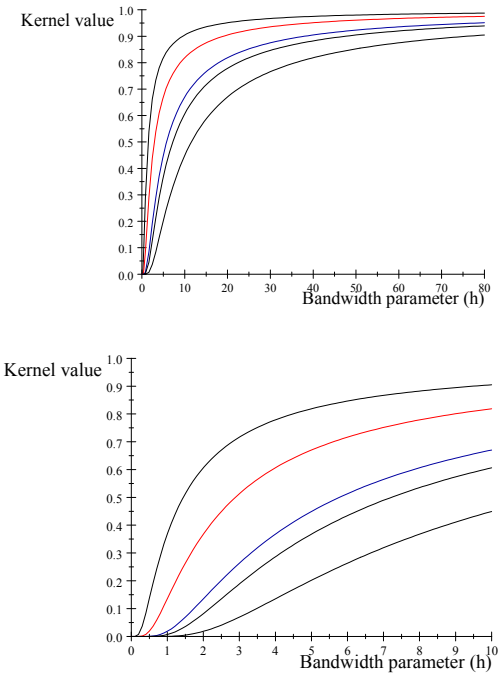
$$\frac{\partial E\,(.)}{\partial\alpha_j}=1+\alpha_i\sqrt{\beta_i\beta_j};\quad \frac{\partial E\,(.)}{\partial\beta_j}=\beta_i+\frac{1}{2}\alpha_i\alpha_j\sqrt{\frac{\beta_i}{\beta j}}.$$

Arbitrary Gaussian kernel adopted for the RKHS regression
using as covariate a 2×1 vector: number of alleles at each of the two loci,
e.g., $x_{AA}=2, x_{Aa}=1$ and $x_{aa}=0$. For example, the kernel entry $AABB$ and $AAbb$ is

$$k(\mathbf{x}_{AABB},\mathbf{x}_{AAbb},h)=\exp\left[-\frac{(2-2)^2+(2-0)^2}{h}\right]=\exp\left[-\frac{4}{h}\right],$$

$$\mathbf{K}_h=\begin{bmatrix}
 & AABB & AABb & AAbb & AaBB & AaBb & Aabb & aaBB & aaBb & aabb\\
AABB & 1 & e^{-\frac{1}{h}} & e^{-\frac{4}{h}} & e^{-\frac{1}{h}} & e^{-\frac{2}{h}} & e^{-\frac{5}{h}} & e^{-\frac{4}{h}} & e^{-\frac{5}{h}} & e^{-\frac{8}{h}}\\
AABb & e^{-\frac{1}{h}} & 1 & e^{-\frac{1}{h}} & e^{-\frac{2}{h}} & e^{-\frac{1}{h}} & e^{-\frac{2}{h}} & e^{-\frac{5}{h}} & e^{-\frac{4}{h}} & e^{-\frac{5}{h}}\\
AAbb & e^{-\frac{4}{h}} & e^{-\frac{1}{h}} & 1 & e^{-\frac{5}{h}} & e^{-\frac{2}{h}} & e^{-\frac{1}{h}} & e^{-\frac{8}{h}} & e^{-\frac{5}{h}} & e^{-\frac{4}{h}}\\
AaBB & e^{-\frac{1}{h}} & e^{-\frac{2}{h}} & e^{-\frac{5}{h}} & 1 & e^{-\frac{1}{h}} & e^{-\frac{4}{h}} & e^{-\frac{1}{h}} & e^{-\frac{2}{h}} & e^{-\frac{5}{h}}\\
AaBb & e^{-\frac{2}{h}} & e^{-\frac{1}{h}} & e^{-\frac{2}{h}} & e^{-\frac{1}{h}} & 1 & e^{-\frac{1}{h}} & e^{-\frac{2}{h}} & e^{-\frac{1}{h}} & e^{-\frac{2}{h}}\\
Aabb & e^{-\frac{5}{h}} & e^{-\frac{2}{h}} & e^{-\frac{1}{h}} & e^{-\frac{4}{h}} & e^{-\frac{1}{h}} & 1 & e^{-\frac{5}{h}} & e^{-\frac{2}{h}} & e^{-\frac{1}{h}}\\
aaBB & e^{-\frac{4}{h}} & e^{-\frac{5}{h}} & e^{-\frac{8}{h}} & e^{-\frac{1}{h}} & e^{-\frac{2}{h}} & e^{-\frac{5}{h}} & 1 & e^{-\frac{1}{h}} & e^{-\frac{4}{h}}\\
aaBb & e^{-\frac{5}{h}} & e^{-\frac{4}{h}} & e^{-\frac{5}{h}} & e^{-\frac{2}{h}} & e^{-\frac{1}{h}} & e^{-\frac{2}{h}} & e^{-\frac{1}{h}} & 1 & e^{-\frac{1}{h}}\\
aabb & e^{-\frac{8}{h}} & e^{-\frac{5}{h}} & e^{-\frac{4}{h}} & e^{-\frac{5}{h}} & e^{-\frac{2}{h}} & e^{-\frac{1}{h}} & e^{-\frac{4}{h}} & e^{-\frac{1}{h}} & 1
\end{bmatrix}$$

Kernel value $k(.,.;h) = \exp\left(-\frac{S}{h}\right)$ against bandwidth parameter h. Curves, from upper to lower, correspond to $S = 1, 2, 4, 5, 8$



$h = 1.75$ as bandwidth parameter
6 unique entries in the **K** matrix:
1.0 (diagonal elements, the two individuals have identical genotypes)
0.565 (3 alleles in common in a pair of individuals)
0.319 (2 alleles in common, 1 per locus)
0.102 (2 alleles in common at only one locus)
0.06 (1 allele in common)
0.01 (no alleles shared).

Training set

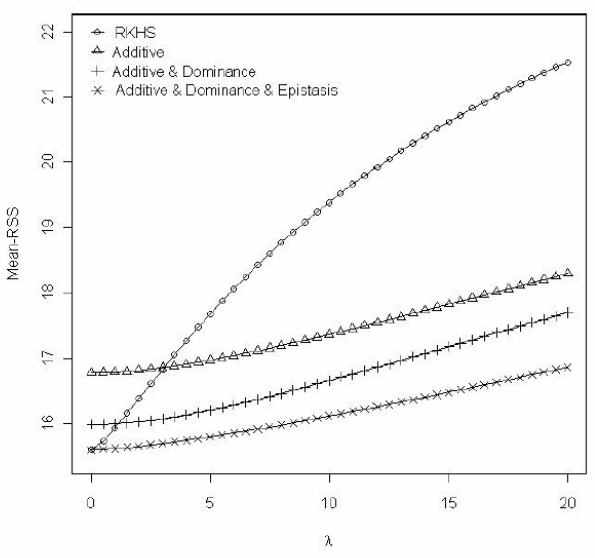
Residuals were drawn from the normal distribution $N(0, 20)$, and added to (21) to form phenotypes. The resulting phenotypic distribution is unknown, because y is a non-linear function of exponential and Weibull variates, plus of an additive normally distributed residual. There were 5 individuals with records for each of the $AABB$,$AABb$,$AAbb$ genotypes; 20 for each of $AaBB$, $AaBb$ and $Aabb$, and 5 of each of $aaBB$, $aaBb$ and $aabb$. Thus, there were 90 individuals with phenotypic records, in total.

Testing set

100

A more important issue, at least from the perspective taken in this paper, is "out of sample" predictive ability. To examine this, 3 new (independent) samples of phenotypes were generated, assuming the residual distribution $N(0, 20)$, as before, and with 5 individuals per genotype, i.e., there were 45 subjects in each sample. The predictive

IMPORTANT ISSUE TO DISCUSS HERE



TRAINING SET

Figure 4. Average (over 90 data points) squared residual for four models fitted to the training sample (RKHS= reproducing kernel Hilbert spaces regression with Gaussian kernel and bandwidth= 1.75) for each value of the smoothing parameter λ.

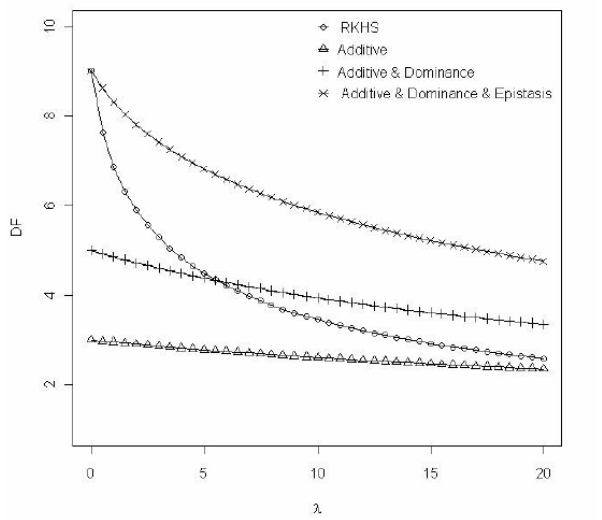
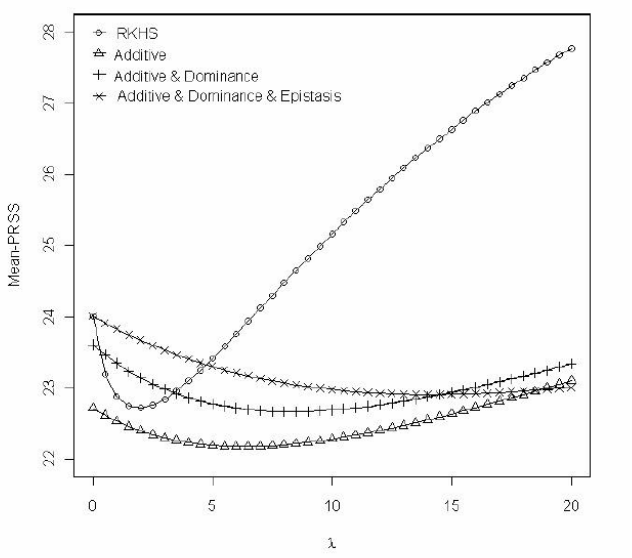


Figure 5. Effective degrees of freedom for four models fitted to the training sample (RKHS= reproducing kernel Hilbert spaces regression with Gaussian kernel and bandwidth= 1.75) at each value of the smoothing parameter λ.



TESTING SET

Figure 6. Average (over 100 samples with 45 realized observations in each) squared prediction error for four models fitted to the predictive sample (RKHS= reproducing kernel Hilbert spaces regression with Gaussian kernel and bandwidth= 1.75) for each value of the smoothing parameter λ.

Explanation of results

How does one explain the paradox that a simple additive model had better predictive performance when gene action was non-linear, as simulated here? In order to address this question, consider the "true" mean value of the 9 genotypes simulated:

	BB	Bb	bb
AA	11.933	8.000	6.417
Aa	3.626	2.919	2.757
aa	0.916	0.304	0.185

The "corrected" sum of squares among these means is 125.23. A fixed effects analysis of variance of these "true" values (assuming genotypes were equally frequent) gives the following partition of sequential sum of squares, apart from rounding errors: 1) additive effect of locus A : 82.8%; 2) additive effect of locus B after accounting for A : 7.06%; 3) dominance effects of loci A and B : 4.2%, and 3) epistasis: 6.2%. Thus, even though the genetic system was non-linear, most of the variation among genotypic means can be accounted for with a linear model on additive effects. The additive model had the worst fit to the data (even worse than the models that assume dominance and epistasis) and, yet, it had the best predictive ability, followed by RKHS for (roughly) $0.5 < \lambda < 3$. !!

Example
Of RKHS 2

		CC	Cc	cc
AA	BB	3	0	3
AA	Bb	0	6	0
AA	bb	3	0	3
Aa	BB	1	2	3
Aa	Bb	3	2	1
Aa	bb	2	2	2
aa	BB	2	2	2
aa	Bb	2	2	2
aa	bb	2	2	2

$E(AA) = (3+3+6+3+3)/9 = 2$
$E(Aa) = (1+2+3+3+2+1+2+2+2)/9 = 2$
$E(aa) = 2 \times 9/9 = 2$
$E(BB) = (3+0+3+1+2+3+2+2+2)/9 = 2$
$E(Bb) = (0+6+0+3+2+1+2+2+2)/9 = 2$
$E(bb) = (3+0+3+2+2+2+2+2+2)/9 = 2$
$E(CC) = (3+0+3+1+3+2+2+2+2)/9 = 2$
$E(Cc) = (0+6+0+2+2+2+2+2+2)/9 = 2$
$E(cc) = (3+0+3+3+1+2+2+2+2)/9 = 2$

- There is no additive variability at any of the three loci, since adding or removing a "large" allele does not affect mean values
- There is no dominance at any of the three loci, as indicated by a zero difference betwen heterozygotes and the average of the homozygotes
- There is considerable interaction. If genotypes are AA, there is pure dominance at each of the B and C loci. In $AaBB$ individuals, removing the C allele increases the mean, with the opposite being true in $AaBb$. In $Aabb$ individuals the $C-locus$ genotype is inmaterial. In aa genotypes, nothing happens.

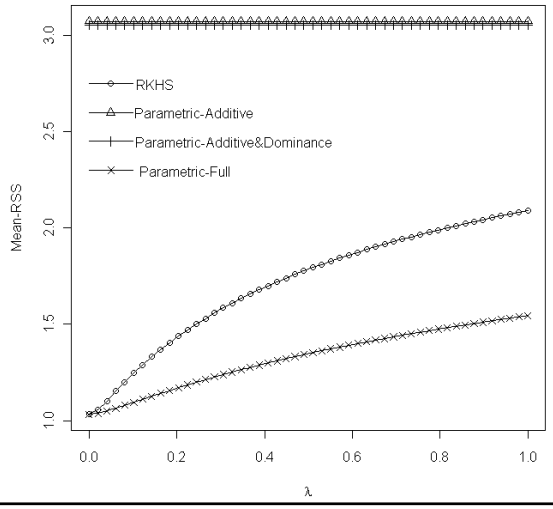
Source	DF	Anova SS	Mean Square	F Value	Pr > F
a	2	0.00000000	0.00000000	0.00	1.0000
b	2	0.00000000	0.00000000	0.00	1.0000
c	2	0.00000000	0.00000000	0.00	1.0000
a*b	4	0.00000000	0.00000000	0.00	1.0000
a*c	4	0.00000000	0.00000000	0.00	1.0000
b*c	4	13.33333333	3.33333333	1.00	0.4609
Error (a*b*c)	8	26.66666667	3.33333333		

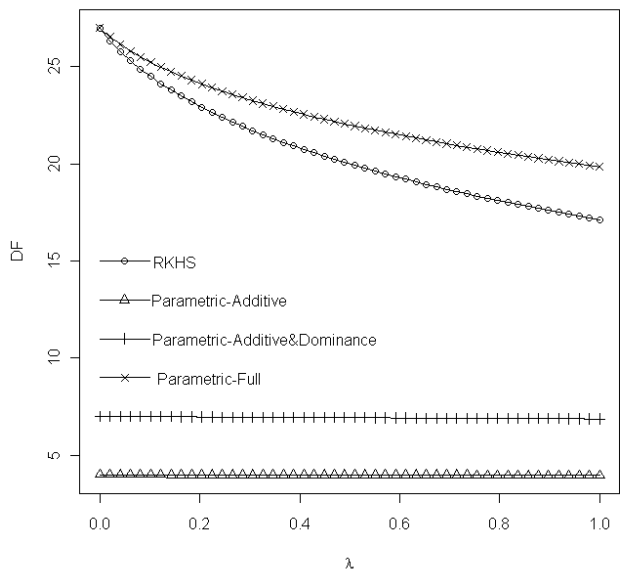
Variation between genotypic values is pure interaction

Training set:
- 27 genotypes,
- 5 replicates per genotype,
- residual variance 1.5

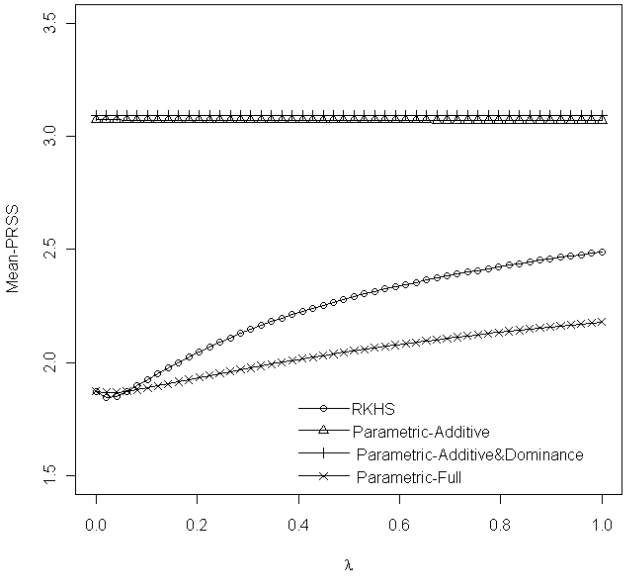
Testing set: 50 MC replicates, each as the training set.

Results in training set





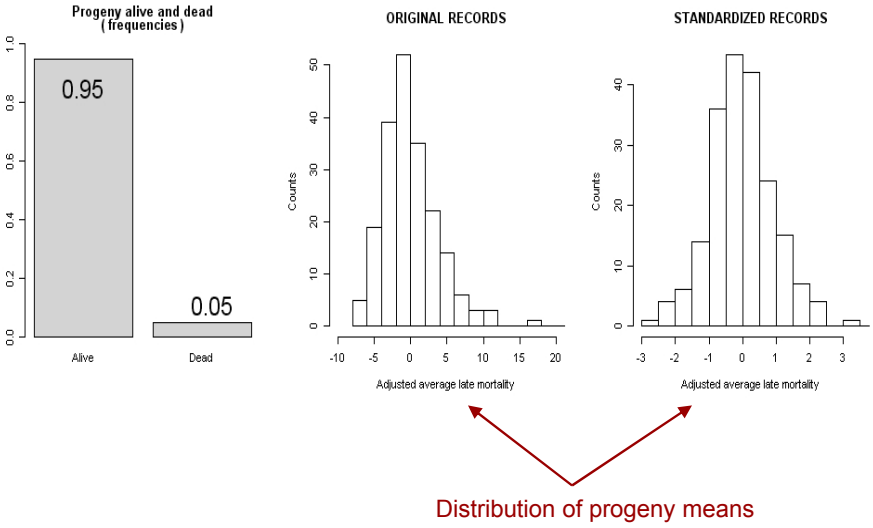
Results testing set.



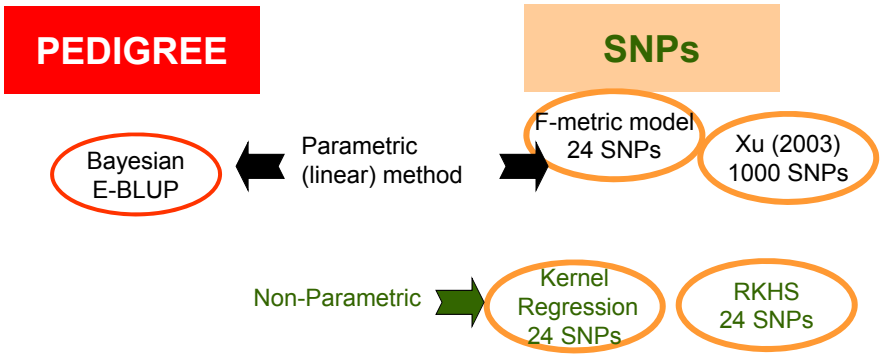
EXAMPLE 3: CHICKENDATA

- Average progeny “late mortality” (lm) in low hygiene environment for 200 sires of line29 (12,167 progenies).
  - Pre-corrected for hatch, age of dam and dam,
  - Standardized log-transformed means
- SNPs: filter and wrapper strategy (Long et al., 2007)
  - 24 SNPs selected out of over 5000 genotyped on sires

DATA



MODELS



E-BLUP

$$\mathbf{y} = \mathbf{Zu} + \mathbf{e}$$

$$\mathbf{u} \sim N(\mathbf{0}, \mathbf{A}\sigma_u^2) \qquad \sigma_u^2 \sim \upsilon_u s_u^2 \chi_{\upsilon_u}^{-1}$$

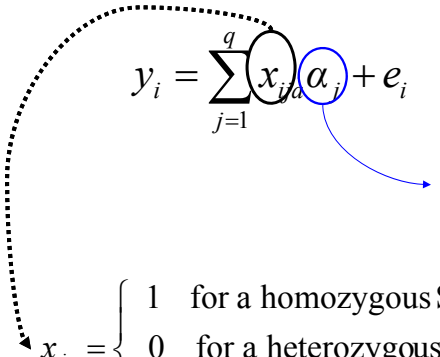
$$e \sim N(\mathbf{0}, \mathbf{R} = \mathbf{N}^{-1}\sigma_e^2) \qquad \sigma_e^2 \sim \upsilon_e s_e^2 \chi_{\upsilon_e}^{-1}$$

Number of progeny of sire *i*.
Weighted residuals. (Varona and Sorensen, 2007)

GIBBS SAMPLING
200,000 samples
50,000 burn-in
10 thinning period

F-metric model (Least-squares Regression)

Van der Veen (1959); Zeng et al. (2005)



$$y_i = \sum_{j=1}^{q} x_{ija}\alpha_j + e_i$$

q= 24 markers

$$\boldsymbol{\alpha} = \{\alpha_{ja}\}$$

$$x_{ja} = \begin{cases} 1 & \text{for a homozygous SNP (say AA)} \\ 0 & \text{for a heterozygous SNP (say Aa)} \\ -1 & \text{for a homozygous SNP (say aa)} \end{cases}$$

F-metric model (Linear Regression)

$$y_i = \sum_{j=1}^{q} x_{ija}\alpha_j + e_i$$

$$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_{24})'$$

Coefficients: Bounded uniform priors (-99999, 99999)

$$e \sim N(\mathbf{0}, \mathbf{R} = \mathbf{N}^{-1}\sigma_e^2)$$

Residual variance: Inverse chi-squared

$$\sigma_e^2 \sim \upsilon_e s_e^2 \chi_{\upsilon_e}^{-1}$$

GIBBS SAMPLING
200,000 samples
50,000 burn-in
10 thin period

Bayesian Regression (Xu, 2003)

1000 SNPs chosen randomly along the genome

$$y_i = \sum_{j=1}^{1000} x_{ija} b_j + e_i$$

$$x_{ia} = \begin{cases} 1 & \text{for a homozygous SNP (say AA)} \\ 0 & \text{for a heterozygous SNP (say Aa)} \\ -1 & \text{for a homozygous SNP (say aa)} \end{cases}$$

Bayesian Regression (Xu, 2003)

(similar to Bayes A of Meuwissen et al. 2001)

1000 SNPs chosen randomly along the genome

$$y_i = \sum_{j=1}^{1000} x_{ija} b_j + e_i$$

b_i Regression coefficient for SNP *i*, assumed distributed as $b_i \sim N(0, \sigma_i^2)$

Where σ_i^2 is the variance associated to each SNP

$$\sigma_i^2 \sim \upsilon\, s^2 \chi_\upsilon^{-1}$$

$$e \sim N(\mathbf{0}, \mathbf{R} = \mathbf{N}^{-1}\sigma_e^2)$$

Residual variance: Inverse chi-squared

$$\sigma_e^2 \sim \upsilon_e s_e^2 \chi_{\upsilon_e}^{-1}$$

GIBBS SAMPLING
200,000 samples
50,000 burn-in
10 thin period

The Gibbs sampler: not much new here…

➢The conditional posterior of location effects is MULVN with mean vector

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\boldsymbol{\beta}}_{-0} \end{bmatrix} = \begin{bmatrix} \frac{\mathbf{1}'\mathbf{1}}{\sigma_e^2} & \frac{\mathbf{1}'\mathbf{X}}{\sigma_e^2} \\ \frac{\mathbf{X}'\mathbf{1}}{\sigma_e^2} & \frac{\mathbf{X}'\mathbf{X}}{\sigma_e^2} + Diag\left(\frac{1}{\sigma_j^2}\right) \end{bmatrix}^{-1} \begin{bmatrix} \frac{\mathbf{1}'\mathbf{y}}{\sigma_e^2} \\ \frac{\mathbf{X}'\mathbf{y}}{\sigma_e^2} \end{bmatrix}$$

$$\hat{\boldsymbol{\beta}} = \mathbf{C}^{-1}\mathbf{r}$$

$$\boldsymbol{\beta}|ELSE \sim N(\mathbf{C}^{-1}\mathbf{r}, \mathbf{C}^{-1})$$

➢The conditional posterior distributions of the variances of marker effects and of residual variance are

$$p\left(\sigma_j^2|ELSE\right) = \left(b_j^2 + \nu S^2\right)\chi_{\nu+1}^{-2} \qquad \text{j=1,2,…,1000}$$

$$\sigma_e^2|ELSE \sim (\mathbf{y} - \mathbf{Xb})'(\mathbf{y} - \mathbf{Xb}) + \upsilon_e s_e^2$$

KERNEL REGRESSION

Gianola et al. (2006)

- Non-parametric regression

$$y_i = g(\mathbf{x}_i) + e_i$$

$\mathbf{x}_i$ is a (*q*x1) vector representing the genotype of sire *i*

$g(\mathbf{x}_i)$ some unknown function of the whole SNP genotype for sire *i*, representing the expected phenotypic value of sires possessing the *q*-dimensional SNPs combination

$\mathbf{e} = \{e_i\}$ assumed distributed independently of X, and around zero

KERNEL REGRESSION

- Non-parametric regression

$$y_i = g(\mathbf{x}_i) + e_i$$

g(x) =conditional expectation function.

– How do we estimate g(x) ?

Nadaraya-Watson estimator
(Nadaraya, 1964; Watson, 1964)
Based on definition of
conditional mean

$$g(\mathbf{x}) = \frac{\int yp(\mathbf{x}, y)dy}{p(\mathbf{x})}$$

KERNEL REGRESSION

- Non-parametric regression

$$y_i = g(\mathbf{x}_i) + e_i$$

$$g(\mathbf{x}) = \frac{\int yp(\mathbf{x}, y)dy \approx \frac{1}{nh^q}\sum_{i=1}^{n} y_i K_h(X - x_i)}{p(\mathbf{x}) \approx \frac{1}{nh^q}\sum_{i=1}^{n} K_h(X - x_i)}$$

h: smoothing parameter

Trinomial Kernel

Pure non-parametric regression.

Trinomial KERNEL

K(X-x) = Some function measuring distances between focal points or objects (genotypes).

$$K_{h_1,h_2}(\mathbf{x}-\mathbf{x}_i) = h_1^{d_{i1}} h_2^{d_{i2}} (1-h_1-h_2)^{2q-d_{i1}-d_{i2}}$$

Focal genotype	Observed genotype		
	AA	Aa	aa
AA	0 0	1 0	0 1
Aa	1 0	0 0	1 1
aa	0 1	1 1	0 0

49

REPRODUCING KERNEL HILBERT SPACES REGRESSION

- Penalized sum of squares has the form:

$$J[g(\mathbf{x}) \mid \lambda] = \tfrac{1}{2}[\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-g(\mathbf{x})]'\mathbf{R}^{-1}[\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-g(\mathbf{x})] + \tfrac{\lambda}{2}\| g(\mathbf{x})\|_H^2$$

$$g(\mathbf{X}\mid h) = \begin{bmatrix} \mathbf{k}_1'(h) \\ \vdots \\ \mathbf{k}_j'(h) \\ \vdots \\ \mathbf{k}_q'(h) \end{bmatrix} \boldsymbol{\alpha} = \mathbf{K}_h\boldsymbol{\alpha}$$

$$\boldsymbol{\alpha} = [\alpha_0, \alpha_1, \ldots, \alpha_n]'$$

$$\mathbf{K}_h = \begin{bmatrix} K_h(x_1,x_1) & K_h(x_1,x_j) & K_h(x_1,x_n) \\ \ldots & K_h(x_i,x_j) & \ldots \\ K_h(x_n,x_1) & K_h(x_n,x_j) & K_h(x_n,x_n) \end{bmatrix}$$

$$K_h(\mathbf{x}-\mathbf{x}_i) = \exp\left[-\frac{(\mathbf{x}-\mathbf{x}_i)'(\mathbf{x}-\mathbf{x}_i)}{h}\right]$$

REPRODUCING KERNEL HILBERT SPACES

- Embedding all these expression in the penalized sum of squares:

$$\begin{bmatrix} \mathbf{1}'\mathbf{R}^{-1}\mathbf{1} & \mathbf{1}'\mathbf{R}^{-1}\mathbf{K}_h \\ \mathbf{K}_h'\mathbf{R}^{-1}\mathbf{1} & \mathbf{K}_h'\mathbf{R}^{-1}\mathbf{K}_h + \frac{1}{\lambda^{-1}}\mathbf{K}_h \end{bmatrix} \begin{bmatrix} \hat{\mu}_{\lambda,h} \\ \hat{\boldsymbol{\alpha}}_{\lambda,h} \end{bmatrix} = \begin{bmatrix} \mathbf{1}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{K}_h'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

$$\boldsymbol{\alpha} \mid \lambda, h \sim N(0, \mathbf{K}_h^{-1}\lambda^{-1}) \qquad \sigma_\alpha^2 \sim \upsilon_\alpha s_\alpha^2 \chi_{\upsilon_\alpha}^{-1}$$

$$\mathbf{e} \sim N(0, \mathbf{R}) \qquad \sigma_e^2 \sim \upsilon_e s_e^2 \chi_{\upsilon_e}^{-1}$$

GIBBS SAMPLING
200,000 samples
50,000 burn-in
10 thin period

Sequence alignment Kernel

Sequence alignment KERNEL

Dynamic programming algorithms

Similarity between two DNA sequences

Adapted to SNP sequences

$$K_h(\mathbf{x} - \mathbf{x}_i) = \exp\left[-Score(\mathbf{x} - \mathbf{x}_i)\right]$$

No need to tune *h*

(Delcher et al., 1999, 2002)

52

Variance component & parameter estimates

Parameter	Posterior features	E-BLUP	F-metric	RKHS	BR (Xu's)
σ_e^2	μ (s.d)	24.38 (3.88)	29.72 (3.56)	17.07 (3.02)	20.75 (2.91)
	HPD (95%)	16.88-32.04	23.60-37.51	11.78-23.64	15.62-27.09
σ_u^2	μ (s.d)	0.10 (0.06)	…	…	1.03 (0.71)
	HPD (95%)	0.03-0.24	…	…	0.67-1.95
σ_α^2	μ (s.d)	…	…	0.40 (0.07)	
	HPD (95%)	…	…	0.28-0.55	
h^2	μ (s.d)	0.02 (0.01)	…	…	…
	HPD (95%)	0.004-0.050	…	…	…

Sum of posterior means of variances of the 1000 markers

- Spearman (above diagonal) and Pearson correlations (below diagonal) between posterior means of sire effects

	E-BLUP	F-metric	Kernel	RKHS	BR
E-BLUP	…	0.52	0.77	0.84	**0.91**
F-metric	**0.56**		**0.48**	**0.51**	0.53
Kernel	0.66	0.38	…	0.93	0.76
RKHS	0.84	**0.50**	0.79	…	0.84
BR	**0.92**	**0.57**	0.58	0.80	…

- E-BLUP & Xu (2003) very similar.
- LR most different ranking.

MODEL FIT

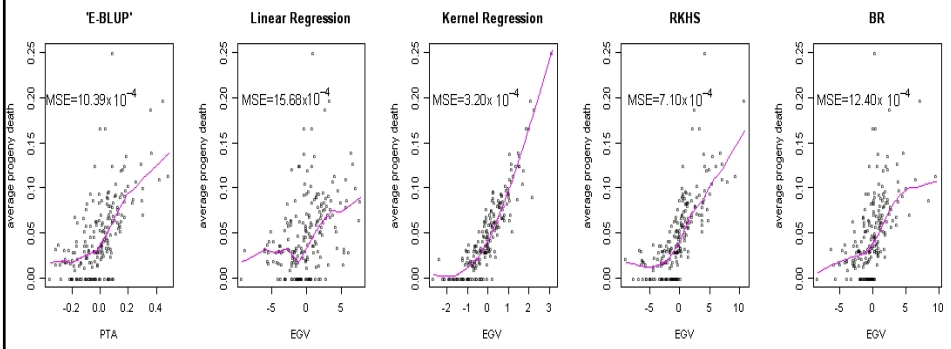
-Compute deviance measurement based on mean squared errors:

- A) Regression of adjusted average progeny on sire's PTA or EGV
- B) Regression of raw average progeny on sire's PTA or EGV

-Lowess regression
(Non-parametric locally weighted regression)

MODEL FIT

•Regression of adjusted raw progeny LM on sire's PTA or EGV



MODEL FIT

- Less dispersion in non-parametric models
- Lower MSE for kernel regression
- Worst for Linear regression (F-metric model)

Still….which model predicts the data best ?

Predictive ability

- Cross validation
  1. 5 subsets, letting 20% sire means missing each time at random
  2. Estimate PTA or EGV of sires with missing values from the augmented posterior distributions
  3. Calculate correlations between actual and inferred average progeny, for each method within subset.

Predictive ability

Subset	E-BLUP	F-metric	Kernel	RKHS	BR
1st	0.03	**0.27**	0.05	**0.27**	0.13
2nd	0.18	0.19	0.28	**0.37**	0.12
3rd	**0.18**	0.08	0.06	-0.01	0.17
4th	-0.04	0.07	0.13	**0.28**	0.15
5th	0.17	-0.12	0.23	0.15	**0.25**
GLOBAL	0.10	0.06	0.14	**0.20**	0.16

- RKHS showed better predictive ability
  - –25% higher reliability than Xu's method
  - –100% higher reliability than E-BLUP
  - –233% higher reliability than F-metric (linear regression on markers)
- RKHS better than fixed or random regression on markers and E-BLUP.

EXAMPLE 4: CHICKEN DATA

Genomic-assisted prediction of a quantitative trait in parents and progeny: application to food conversion rate in chickens

FCR measured on progeny of 333 sires with 3481 SNPs
FCR measured on progeny of 61 birds (sons of the above sires)

➔2- generation data set

BAYES A --all markers
RKHS --all markers
RKHS --400 markers filtered using different INFOGAINS
BLUP (Bayes) –pedigree information

Training set: 333 sires of sons

Predictive set: 61 sons of sires

Table 1: Means, standard deviation (s.d.) and 95% confidence intervals (CI) of the Bootstrap distribution of Spearman correlations between predicted and observed phenotypes in the testing set (E-BLUP: Bayesian linear model; Bayes A: Bayesian regression on SNP; RKHS: reproducing kernel Hilbert spaces regression).

Whole genome methods			
method	mean	s.d	CI (95%)
E-BLUP	0.11	0.13	(-0.13, 0.35)
Bayes A	0.27	0.12	(0.04, 0.49)
RKHS	0.27	0.12	(0.03, 0.50)
Information gain using 2 classes (400 pre-selected SNPs) + RKHS			
percentile	mean	s.d	CI(95%)
0.15	0.33	0.12	(0.09, 0.56)
0.20	0.32	0.11	(0.10, 0.53)
0.25	0.36	0.11	(0.13, 0.57)
0.30	0.19	0.12	(-0.05, 0.42)
0.35	0.35	0.11	(0.12,0.55)
0.40	0.33	0.11	(0.10-0.53)
Information gain using 3 classes (400 pre-selected SNPs) + RKHS			
percentile	mean	s.d	CI(95%)
0.15	0.32	0.11	(0.10, 0.54)
0.20	0.24	0.13	(-0.01, 0.48)
0.25	0.39	0.11	(0.16, 0.59)
0.30	0.19	0.12	(-0.05, 0.42)
0.35	0.20	0.12	(-0.04, 0.43)
0.40	0.16	0.12	(-0.08, 0.40)

Note that the confidence bands of the predictive correlations are wide

Figure 1. Heat map of linkage disequilibrium (r^2) between SNPs preselected using two different criteria for classifing sires: 2 classes (high and low) with percentile =0.30 and 3 classes (high, medium and low) with percentile=0.25.

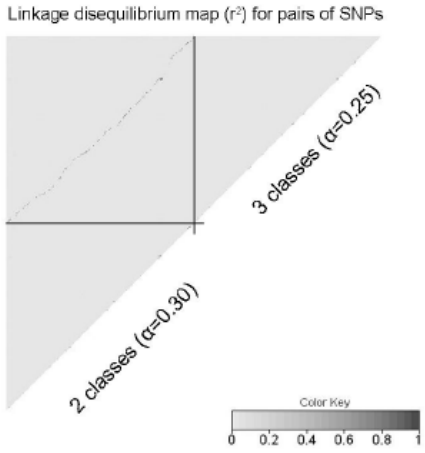
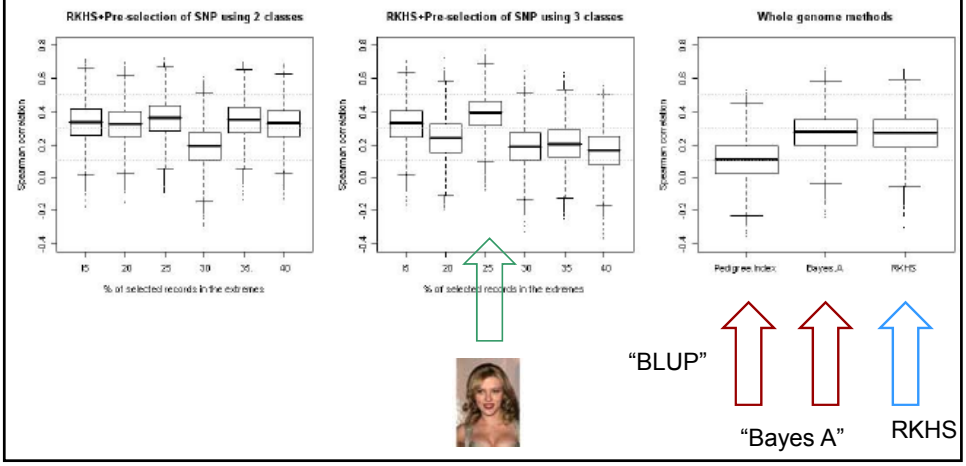


Figure 2. Box plots for the bootstrap distribution of Spearman correlations between predicted and observed phenotype in the testing set (progeny) obtained with: RKHS on 400 pre-selected SNPs using 2 or 3 classes to classify sires with different percentiles (left and middle panels, respectively) and methods using pedigree or all available SNPs (right pannel).



"BLUP"

"Bayes A"

RKHS

EXAMPLE 5: Application to US Jersey data

⇒ **US Jersey**

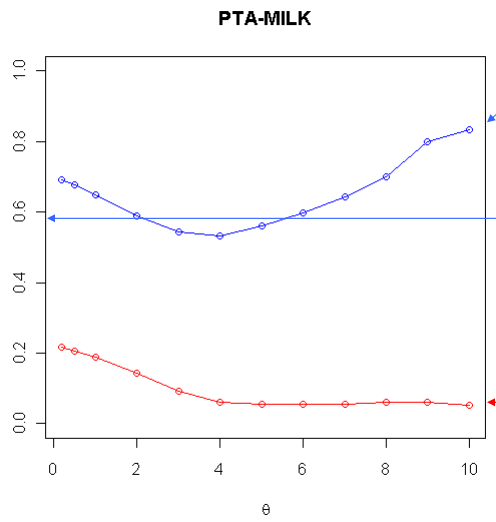
- **N=** 1,762 sires (n=1446, training n=1130 ; testing, n=316).
- **Markers:** BovineSNP50 BeadChip (50k).
- **Traits:** PTAs for Milk, Protein Content and Daughter Pregnancy Rate

⇒ **Models:**

- Linear model $\mathbf{K} = \mathbf{XX}'$
- Genomic-based kinship $\mathbf{K} = \mathbf{G}$ [1]
- Gaussian Kernel $K(i, j|\theta) = Exp\{ -\theta \times d(\mathbf{x}_i, \mathbf{x}_j) \}$
  - Fixed over a grid of values
  - Kernel averaging:

[1] Hayes and Goddard (2008) Journal of Animal Science.

Application to US Jersey data



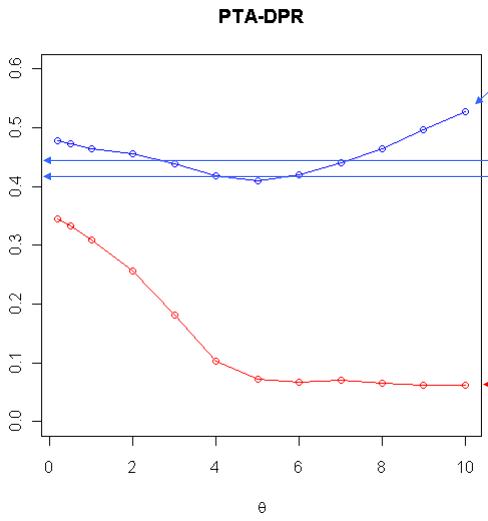
PMSE

(MSE of predictive residuals)

	PMSE
Linear Model	0.588
Marker-based Kinship	0.590
Kernel average	0.588

Residual Variance

Application to US Jersey data



PMSE

(MSE of predictive residuals)

	PMSE
Linear Model	0.459
Marker-based Kinship	0.460
Kernel Average	0.421

Residual Variance

Empirical Application

⇒ Kernel Averaging seems be an effective strategy kernel selection

⇒ In this example (PTAs):

Linear Model, Kinship and Kernel Averaging performed similarly

⇒ Not necessarily so for other traits and other populations

Predictive ability of models for genomic selection in Wheat [1]

	Predictive Correlation		Difference
Environment	BL	RKHS	(%)
E1	0.518	0.601	+16%
E2	0.493	0.494	0%
E3	0.403	0.445	+10%
E4	0.457	0.524	+15%

N= 599;

Trait: Grain Yield (4 environments);

Models: RKHS and Bayesian LASSO (BL)

[1] Crossa *et al.* (2010) Genetics.

Radial Basis Functions: another form of non-parametric regression

$$y_i = \mathbf{w}_i'\beta + \sum_{j=1}^{n} k_{\boldsymbol{\theta}}\left(\mathbf{x}_i, \mathbf{x}_j\right)\alpha_j + e_i$$

$\beta =$	nuisance location vector
$\mathbf{x} =$	$p \times 1$ vector of SNP genotypes
$\boldsymbol{\alpha} = \{\alpha_i\}$	$n \times 1$ vector of regressions
$k_{\boldsymbol{\theta}}\left(\mathbf{x}_i, \mathbf{x}_j\right)$	basis function, a transformation of $\mathbf{x}_i, \mathbf{x}_j$
$\boldsymbol{\theta}$	parameter vector, possibly of order $p \times 1$
$\mathbf{e} = \{e_i\}$	$\mathbf{e} \sim N\left(\mathbf{0}, \mathbf{I}\sigma_e^2\right)$

Note that the basis functions are adaptive: depend on parameters (θ)

Matrix form

$$y = \mathbf{W}\beta + \mathbf{K}_{\boldsymbol{\theta}}\boldsymbol{\alpha} + \mathbf{e}$$

$$\mathbf{K}_{\boldsymbol{\theta}} = \begin{bmatrix} \exp\left[-\sum_{k=1}^{p}\theta_k\left(x_1^{[k]} - x_1^{[k]}\right)^2\right] & \cdots & \exp\left[-\sum_{k=1}^{p}\theta_k\left(x_1^{[k]} - x_n^{[k]}\right)^2\right] \\ \exp\left[-\sum_{k=1}^{p}\theta_k\left(x_2^{[k]} - x_1^{[k]}\right)^2\right] & \cdots & \exp\left[-\sum_{k=1}^{p}\theta_k\left(x_2^{[k]} - x_n^{[k]}\right)^2\right] \\ \vdots & \cdots & \vdots \\ \exp\left[-\sum_{k=1}^{p}\theta_k\left(x_n^{[k]} - x_1^{[k]}\right)^2\right] & \cdots & \exp\left[-\sum_{k=1}^{p}\theta_k\left(x_n^{[k]} - x_n^{[k]}\right)^2\right] \end{bmatrix}$$

The kernel matrix $\boldsymbol{K}_\theta$ need not be positive-definite here (contrary to RKHS regression)

Genotypes can be coded, for example as

	$SNP1$	$SNP2$	$SNP3$		$SNP1$	$SNP2$	$SNP3$
Ind. 1	AA	Gg	TT	$\Rightarrow$	2	1	2
Ind. 2	AA	gg	tt	$\Rightarrow$	2	0	0

$$
\begin{aligned}
k_{\boldsymbol{\theta}}(\mathbf{x}_1, \mathbf{x}_2) &= \exp\left[-\sum_{k=1}^{3} \theta_k \left(x_1^{[k]} - x_2^{[k]} \right)^2 \right] \\
&= \exp\left[-\theta_1 (2-2)^2 - \theta_2 (1-0)^2 - \theta_3 (2-0)^2 \right] \\
&= \exp[-\theta_2 - 4\theta_3] \\
k_{\boldsymbol{\theta}}(\mathbf{x}_2, \mathbf{x}_1) &= \exp\left[-\sum_{k=1}^{3} \theta_k \left(x_2^{[k]} - x_1^{[k]} \right)^2 \right] \\
&= \exp\left[-\theta_1 (2-2)^2 - \theta_2 (0-1)^2 - \theta_3 (0-2)^2 \right] \\
&= \exp[-\theta_2 - 4\theta_3]
\end{aligned}
$$

Bayesian structure: priors

Notations: $\boldsymbol{\tau} = (\tau_1^2, \ldots, \tau_j^2, \ldots, \tau_n^2)$; $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k, \ldots, \theta_p)$

$$
\begin{aligned}
p(\boldsymbol{\beta}) &\propto \text{constant} && (1) \\
p(\boldsymbol{\alpha}|\boldsymbol{\tau}, \sigma_e^2) &= N(\mathbf{0}, \sigma_e^2 \boldsymbol{D}_{\boldsymbol{\tau}}), \quad \boldsymbol{D}_{\boldsymbol{\tau}} = Diag(\tau_1^2, \ldots, \tau_j^2, \ldots, \tau_n^2) && (2) \\
p(\boldsymbol{\tau}|\lambda^2) &= \prod_{j=1}^{n} \text{Expon}\left(\frac{\lambda^2}{2}\right) = \prod_{j=1}^{n} \frac{\lambda^2}{2} \exp\left(-\frac{\lambda^2 \tau_j^2}{2}\right) && (3) \\
p(\sigma_e^2|a, \nu) &= \text{IG}(\text{shape} = a, \text{scale} = \nu) \propto (\sigma_e^2)^{-a-1} \exp\left(-\frac{\nu}{\sigma_e^2}\right) && (4) \\
p(\lambda^2|\gamma_1, \delta_1) &= \text{Gamma}(\text{shape} = \gamma_1, \text{rate} = \delta_1) \propto \left(\lambda^2\right)^{\gamma_1 - 1} \exp\left(-\delta_1 \lambda^2\right) && (5) \\
p(\boldsymbol{\theta}|\rho) &= \prod_{k=1}^{p} \text{Expon}(\rho) = \prod_{k=1}^{p} \rho \exp(-\rho \theta_k) && (6) \\
p(\rho|\gamma_2, \delta_2) &= \text{Gamma}(\text{shape} = \gamma_2, \text{rate} = \delta_2) \propto \rho^{\gamma_2 - 1} \exp(-\delta_2 \rho) && (7)
\end{aligned}
$$

Hyper-parameters are: a, ν, γ_1, γ_2, δ_1 , δ_2

4.1 Joint posterior

$$
\begin{aligned}
p(\text{Param}|\text{Data}) &\propto p(\mathbf{y}|\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\theta},\sigma_e^2)p(\boldsymbol{\beta})p(\boldsymbol{\alpha}|\boldsymbol{\tau},\sigma_e^2)p(\sigma_e^2|a,\nu)p(\boldsymbol{\tau}|\lambda^2)p(\lambda^2|\gamma_1,\delta_1)p(\boldsymbol{\theta}|\rho)p(\rho|\gamma_2,\delta_2) \\
&= \left(\sigma_e^2\right)^{-n/2}\exp\left[-\frac{1}{2\sigma_e^2}(\mathbf{y}-\boldsymbol{W\beta}-\boldsymbol{K_\theta\alpha})'(\mathbf{y}-\boldsymbol{W\beta}-\boldsymbol{K_\theta\alpha})\right] \\
&\times\prod_{j=1}^{n}\left[\left(2\pi\sigma_e^2\tau_j^2\right)^{-\frac{1}{2}}\exp\left(-\frac{\alpha_j^2}{2\sigma_e^2\tau_j^2}\right)\right]\times\left(\sigma_e^2\right)^{-a-1}\exp\left(-\frac{\nu}{\sigma_e^2}\right)\times\prod_{j=1}^{n}\frac{\lambda^2}{2}\exp\left(-\frac{\lambda^2\tau_j^2}{2}\right) \\
&\times\left(\lambda^2\right)^{\gamma_1-1}\exp\left(-\delta_1\lambda^2\right)\times\prod_{k=1}^{p}\rho\exp(-\rho\theta_k)\times\rho^{\gamma_2-1}\exp(-\delta_2\rho)
\end{aligned}
\tag{8}
$$

4.2 Fully conditionals

$$
\begin{aligned}
p(\beta|\text{else}) &= N\left[(\boldsymbol{W'W})^{-1}\boldsymbol{W'}(\mathbf{y}-\boldsymbol{K_\theta\alpha}),(\boldsymbol{W'W})^{-1}\sigma_e^2\right] && (9)\\
p(\alpha|\text{else}) &\propto p(\mathbf{y}|\boldsymbol{\beta},\boldsymbol{\alpha},\theta,\sigma_e^2)p(\boldsymbol{\alpha}|\tau,\sigma_e^2) \\
&= N\left[\left(\boldsymbol{K_\theta'K_\theta}+\boldsymbol{D_\tau^{-1}}\right)^{-1}\boldsymbol{K_\theta'}(\mathbf{y}-\boldsymbol{W\beta}),\sigma_e^2\left(\boldsymbol{K_\theta'K_\theta}+\boldsymbol{D_\tau^{-1}}\right)^{-1}\right] && (10)\\
p(\sigma_e^2|\text{else}) &\propto p(\mathbf{y}|\boldsymbol{\beta},\boldsymbol{\alpha},\theta,\sigma_e^2)p(\sigma_e^2|a,\nu)p(\boldsymbol{\alpha}|\tau,\sigma_e^2) \\
&\propto \left(\sigma_e^2\right)^{-n-a-1}\exp\left\{-\tfrac{1}{2\sigma_e^2}\left[(\mathbf{y}-\boldsymbol{W\beta}-\boldsymbol{K_\theta\alpha})'(\mathbf{y}-\boldsymbol{W\beta}-\boldsymbol{K_\theta\alpha})+\boldsymbol{\alpha'D_\tau^{-1}\alpha}+2\nu\right]\right\} \\
&= \text{IG}\left(\text{shape}=n+a+1,\text{scale}=\tfrac{1}{2}\cdots\right) && (11)\\
p(\tau|\text{else}) &: p(1/\tau_j^2)=\text{Inverse Gaussian}\left(\text{mean}=\sqrt{\frac{\lambda^2\sigma_e^2}{\alpha_j^2}},\text{scale}=\lambda^2\right) && (12)\\
p(\lambda^2|\text{else}) &= \left(\lambda^2\right)^{\gamma_1-1}\exp\left(-\delta_1\lambda^2\right)\prod_{j=1}^{n}\frac{\lambda^2}{2}\exp\left(-\frac{\lambda^2\tau_j^2}{2}\right) \\
&= \text{Gamma}\left(\text{shape}=n+\gamma_1,\text{rate}=\delta_1+\frac{1}{2}\sum_{j=1}^{n}\tau_j^2\right) && (13)\\
p(\rho|\text{else}) &= \rho^{\gamma_2-1}\exp(-\delta_2\rho)\prod_{k=1}^{p}\rho\exp(-\rho\theta_k) \\
&= \text{Gamma}\left(\text{shape}=p+\gamma_2,\text{rate}=\delta_2+\sum_{k=1}^{p}\theta_k\right) && (14)\\
p(\theta|\text{else}) &\propto \exp\left[-\frac{1}{2\sigma_e^2}(\mathbf{y}-\boldsymbol{W\beta}-\boldsymbol{K_\theta\alpha})'(\mathbf{y}-\boldsymbol{W\beta}-\boldsymbol{K_\theta\alpha})-\rho\sum_{k=1}^{p}\theta_k\right] && (15)
\end{aligned}
$$

Metropolis-Hastings