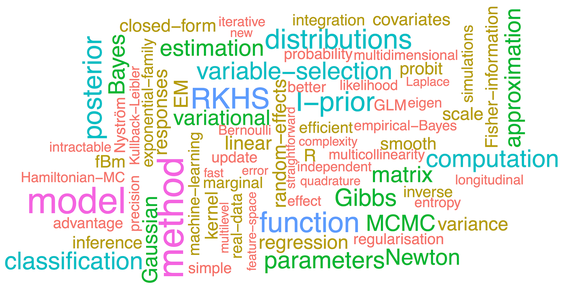
Haziq J & W Bergsma
Department of Statistics
London School of Economics and
Political Science
https://phd.haziqj.ml

Binary and Multinomial Regression using Fisher Information Covariance Kernels (I-priors)

[Updated: 22/12/2019]

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE



Introduction

Consider the following regression model for $i = 1, \dots, n$:

$$\begin{aligned} y_i &= \alpha + f(x_i) + \epsilon_i \\ (\epsilon_1, \dots, \epsilon_n)^\top &\sim \mathrm{N}_n(0, \Psi^{-1}) \end{aligned} \tag{1}$$

where $y_i \in \mathbb{R}$, $x \in \mathcal{X}$, and $f \in \mathcal{F}$. Let $\mathcal{F}$ be a reproducing kernel Hilbert space (RKHS) with kernel $h_\lambda : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. The Fisher information for f evaluated at x and x' is

$$\mathcal{I}\big(f(x), f(x')\big) = \sum_{k=1}^{n} \sum_{l=1}^{n} \Psi_{k,l} h_\lambda(x, x_k) h_\lambda(x', x_l). \tag{2}$$

The I-prior

The entropy maximising prior distribution for f, subject to identifying constraints, is

$$\mathbf{f} = \big(f(x_1), \dots, f(x_n)\big)^\top \sim \mathrm{N}_n\left(\mathbf{f}_0, \mathcal{I}[f]\right).$$

Equivalently, $f(x) = f_0(x) + \sum_{i=1}^{n} h_\lambda(x, x_i) w_i$, with

$$(w_1, \dots, w_n)^\top \sim \mathrm{N}_n(0, \Psi).$$

Of interest are

- the posterior distribution for the regression function

$$p(\mathbf{f}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{f})p(\mathbf{f})}{\int p(\mathbf{y}|\mathbf{f})p(\mathbf{f})\,\mathrm{d}\mathbf{y}}; \text{ and}$$

- the posterior predictive distribution given new data

$$p(y_{\text{new}}|\mathbf{y}) = \int p(y_{\text{new}}|f_{\text{new}}, \mathbf{y}) p(f_{\text{new}}|\mathbf{y})\,\mathrm{d}f_{\text{new}}.$$

Model parameters (error precision Ψ, RKHS scale parameters λ, and any others) may need to be estimated.

A Unified Regression Framework

- Multiple linear regression (linear RKHS)
- Smoothing models (fBm RKHS)
- Multilevel regression (ANOVA RKHS: linear & Pearson)

$$f(x_i^{(j)}) = f_1(j) + f_2(x_i^{(j)}) + f_{1:2}(x_i^{(j)}, j)$$

- Longitudinal modelling (ANOVA RKHS: fBm & Pearson)

$$f(x_i, t_i) = f_1(t_i) + f_2(x_i) + f_{1:2}(x_i, t_i)$$

- Functional covariates ($\mathcal{X}$ is a Hilbert-Sobolev space)

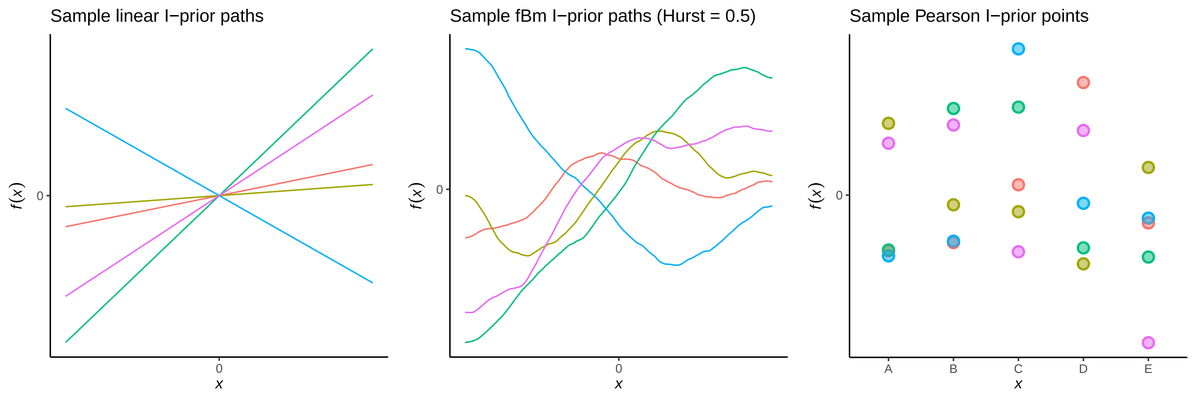


Figure 1: (L-R) Sample paths from the linear, fractional Brownian motion (fBm), and Pearson RKHS. The (reproducing) kernels corresponding to each RKHS are: $h_\lambda(x, x') = \lambda\langle x, x'\rangle_{\mathcal{X}}$ (linear), $h_\lambda(x, x') = -\frac{\lambda}{2}\big(\|x - x'\|_{\mathcal{X}}^{2\gamma} - \|x\|_{\mathcal{X}}^{2\gamma} - \|x'\|_{\mathcal{X}}^{2\gamma}\big)$ (fBm), and $h_\lambda(x, x') = \lambda\big(\delta_{xx'}/\mathrm{P}[X = x] - 1\big)$ (Pearson).

Categorical Responses

When each $y_i \in \{1, \dots, m\}$, normality assumptions are violated. Model instead $y_i = \arg\max_k y_{ik}^*$, where

$$\begin{aligned} y_{ij}^* &= \alpha_j + f_j(x_i) + \epsilon_{ij} \\ (\epsilon_{i1}, \dots, \epsilon_{im})^\top &\sim \mathrm{N}_m(0, \Sigma) \end{aligned} \tag{3}$$

with $\mathrm{Cov}(\epsilon_{ij}, \epsilon_{kj}) = 0$, for all $i \neq k$, $j = 1, \dots, m$, and iid I-priors on $\mathbf{f}_i = \big(f_1(x_i), \dots, f_m(x_i)\big)^\top$. Class probabilities p_{ij} are obtained using a *truncated* m*-variate normal* density

$$p_{ij} = \int\limits_{\{y_{ij}^* > y_{ik}^* \,|\, k \neq j\}} \mathrm{N}_m(\mathbf{y}_i^* \,|\, \mathbf{f}_i, \Sigma)\,\mathrm{d}\mathbf{y}_i^* =: g_j^{-1}(\mathbf{f}_i).$$

Now, the marginal, on which the posterior depends,

$$p(\mathbf{y}) = \int \prod_{i=1}^{n} \prod_{j=1}^{m} \left[\left\{ g_j^{-1}(\mathbf{f}_i) \right\}^{[y_i = j]} \cdot \mathrm{N}_m(\mathbf{f}_j \,|\, \mathbf{f}_0, \mathcal{I}[f])\,\mathrm{d}\mathbf{f}_j \right],$$

cannot be found in closed form. By working in a fully Bayesian setting, we append model parameters and employ a *variational approximation*.

Spatio-Temporal Modelling of BTB

Determine the existence of spatial segregation of multiple types of bovine tuberculosis (BTB) in Cornwall, and whether the spatial distribution had changed over time.

- Constant model (constant RKHS)

$$p_{ij} = g_j^{-1}(\alpha_k)_{k=1}^m$$

- Spatial segregation (fBm RKHS)

$$p_{ij} = g_j^{-1}\big(\alpha_k + f_{1k}(x_i)\big)_{k=1}^m$$

- Spatio-temporal segregation (ANOVA RKHS)

$$p_{ij} = g_j^{-1}\big(\alpha_k + f_{1k}(x_i) + f_{2k}(t_i) + f_{12k}(x_i, t_i)\big)_{k=1}^m$$

Evidence Lower Bound (ELBO) values for the three models are -1197.4, -665.3, and -656.2 respectively.

Detecting Cardiac Arrhythmia

Predict whether patients suffers from a cardiac disease based on features such as age, height, weight and a myriad of electrocardiogram (ECG) data ($p = 271$).

Table 1: Mean out-of-sample misclassification rates and standard errors for 100 runs of various training (n) and test ($451 - n$) sizes for the cardiac arrhythmia binary classification task.

	Misclassification rate (%)		
Method	$n = 50$	$n = 100$	$n = 200$
I-probit (linear)	34.5 (0.4)	31.4 (0.4)	29.7 (0.4)
I-probit (fBm)	34.7 (0.6)	27.3 (0.3)	24.5 (0.3)
GP (Gaussian)	37.3 (0.4)	33.8 (0.4)	29.3 (0.4)
L-1 logistic	34.9 (0.4)	30.5 (0.3)	26.1 (0.3)
SVM (linear)	36.2 (0.5)	35.6 (0.4)	35.2 (0.4)
SVM (Gaussian)	48.4 (0.5)	47.2 (0.5)	46.9 (0.4)
RF	31.7 (0.4)	26.7 (0.3)	22.4 (0.3)
k-NN	40.6 (0.3)	38.9 (0.3)	35.8 (0.4)

Conclusions

- Simple estimation of various categorical models:
  - Choice models (with or without IIA);
  - Random-effects models;
  - Binary and multiclass classification.
- Inference is straightforward (e.g. model comparison or transformed parameter significance).
- Often gives better predictions.

References

[1] Wicher Bergsma.
Regression with I-priors.
Econometrics and Statistics, to appear, 2019.

[2] Mark Girolami and Simon Rogers.
Variational Bayesian multinomial probit regression with Gaussian process priors.
Neural Computation, 18(8), 2006.

[3] Robert E McCulloch, Nicholas G Polson, and Peter E Rossi.
A Bayesian analysis of the multinomial probit model with fully identified parameters.
Journal of Econometrics, 99(1):173–193, 2000.

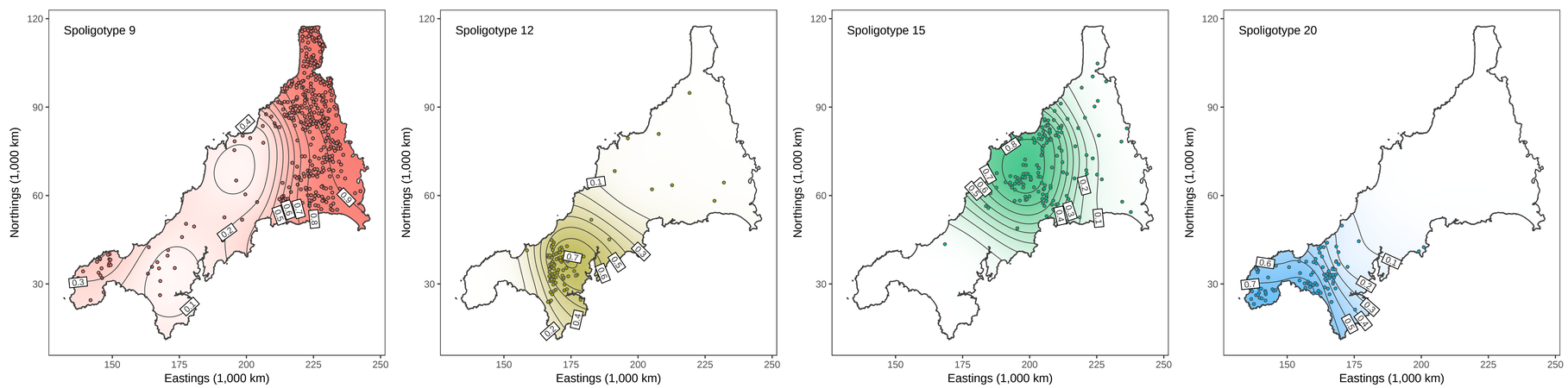


Figure 2: Predicted probability surfaces for BTB contraction in Cornwall for the four largest spoligotypes of the bacterium *Mycobacterium bovis* over the entire time period 1989–2002.