Application for Postdoctoral and Research Associate positions at MRC Biostatistics Unit, University of Cambridge

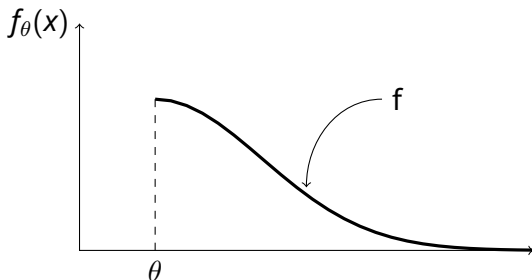Johan Van Der Molen Moris
PhD in Statistics
12th of November, 2020

Past Work

$$X_1, \ldots, X_n \sim f_\theta(x) = f(x - \theta)$$

Results:

- Bayesian joint estimation of θ and f.
- Bernstein-von Mises for marginal posterior of θ.
- Consistent and rate optimal estimation of f in L_1.
- Consistent estimation of f around 0 in sup norm.
- Automatic results for Scale models $Y_i \sim f_\theta(y) = \theta f(\theta y)$ via log transformation.

Motivation: Procurement Auctions

Estimation of the location of density jumps has many applications mainly in econometrics such as procurement auctions or equilibrium job-search models.

Example
Bids observed in Californian state procurement auctions of highway and street maintenance projects carried out by the California Department of Transportation (Caltrans) between January 2002 and December 2005, studied by [Krasnokutskaya and Seim, 2011].

Prior Models for f

Log Spline:

$$f(x; \eta) = \exp \left\{ \sum_{j=1}^{J_n} \eta_j B_{j,q}(x) - c(\eta) \right\}$$

$$c(\eta) = \log \int_0^\infty e^{\eta^t B_q(x)} dx$$

Mixture[1]:

$$f_{P,z}(x) = \int_x^\infty \frac{1}{\theta} \int_0^\infty \Gamma(\theta \mid z, z/\epsilon) dP(\epsilon) d\theta$$
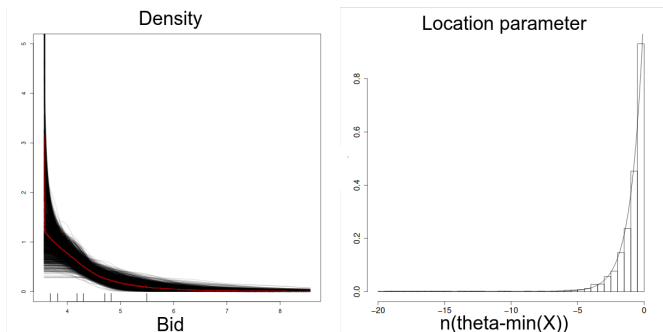
where:

$$dP(\epsilon) = \frac{p\epsilon^2 dQ^{(0)}(\epsilon) + (1-p)dQ^{(1)}(\epsilon)}{p \int_0^\delta \epsilon^2 dQ^{(0)}(\epsilon) + 1 - p}$$

[1]Joint work with Prof. Judith Rousseau and Dr. J.B. Salomond

3

Results

$$\left\| \pi(n(X_{(1)} - \theta) \mid X^n) - Exp(f_0(0)) \right\|_{TV} \to 0, \quad n \to \infty.$$

Figure 1: Left: Draws from the posterior distribution of f, red line is posterior mean; lower support point is the posterior mean of θ. Right: Posterior distribution of θ (recentred by the smallest bid and rescaled by sample size)

4

Credible Intervals:

	Scale model	Location model
95% CI for θ_0	$[2.809, 3.697]$	$[3.275, 3.697]$
95% CI for $f_{0,\theta_0}(\theta_0)$	$[0.262, 5.658]$	$[0.495, 6.596]$

PhD Thesis available here:
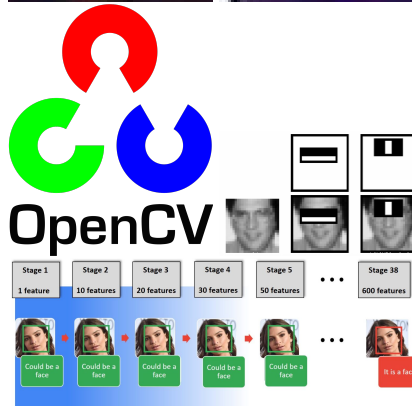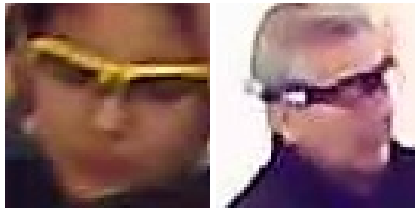https://jvandermolen.com/docs/thesis.pdf

Past Work

At Centre for Advanced Research in Education, University of Chile:

Prediction of elementary school students performance using

- Decision Trees
- SVM
- Matrix and Tensor Factorisation

$$\hat{p}_{s,i,m} = \mu + \mu_s + \mu_i + \sum_{r=1}^{R} a_{s,r} b_{i,r} c_{m,r}$$

Current Work

- Working on some extensions and writing papers from PhD work on mixtures.
    - relaxing smoothness condition.
    - generalising prior process $dP(\epsilon)$.
    - extending to regression on θ.
- Working remotely as a teaching assistant and co-supervising undergraduate students.
- Paper review.

My interest in this position

- Working with 'omics data presents many challenges
  - High-dimensional.
  - Nested data and complex correlation structures.
  - Many sources of uncertainty and levels of noise.
  - Longitudinal and spatial data.
  - Missing data.
- Possibility of working on machine learning and Bayesian methods.
- Collaboration between different fields
- Training

- Expand my area of research.
- Develop new methodology.
- Applications that have high impact.
- Theoretical results motivated by these applications.

What I can offer

- Hard work.
- Knowledge and experience from past work.
- Programming skills: R, LaTeX, Python.
- Capacity to collaborate.
- Transferable skills (eg organise reading group).
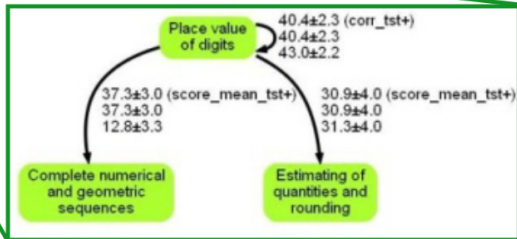- Teaching.

Thank you!

📄 Krasnokutskaya, E. and Seim, K. (2011).
Bid Preference Programs and Participation in Highway Procurement Auctions.
The American Economic Review, 101(6):2653–2686.

Current Work

Co-supervising 3 projects:

- Early detection of gender differences in reading and writing (Multilevel regression, linear mixed effects model).
- Classification of students answers to open-ended questions (Natural Language Processing).
- Analysis of the movement of the teacher in the classroom (Pose detection).