Bayesian Non-parametrics for Marathon Modeling

Melanie F. Pradier Francisco J.R. Ruiz Fernando Perez-Cruz
Email: {melanie,franrruiz,fernando}@tsc.uc3m.es, University Carlos III of Madrid

Problem Formulation

Aim: comparative density estimation of group data + temporal analysis

Some motivating applications

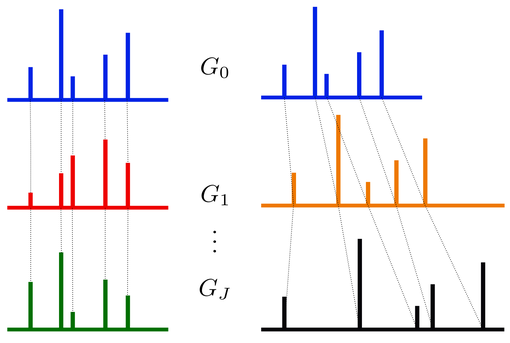
- Pediatrics: children weight and height evolution with age.
- Social Sciences: gender impact on salary income.
- Pharmaceutics: drug responses according to patient's characteristics.

A concrete application:

- Study of age/gender/environment impact on marathon performance.
- Analysis of temporal running patterns.

single-p Dependent Dirichlet Process

- The DDP places a prior over a collection $G_1 \dots G_J$ of random distributions
- Based on Dirichlet Process $G \sim \mathrm{DP}(\mathrm{H}, \alpha)$
  - α: concentration parameter
  - H: base measure
- $G_j = \sum_{k=1}^{\infty} \pi_{jk} \delta_{\phi_{jk}}$
- single-p DDP: Weights π_{jk} shared across groups



Hierarchical DP

$G_0 \sim \mathrm{DP}(\alpha, \mathrm{H})$

$G_j \sim \mathrm{DP}(\gamma, \mathrm{G_0})$

Atom-Dependent DP

$G_0 \sim \mathrm{DP}(\alpha, \mathrm{H})$

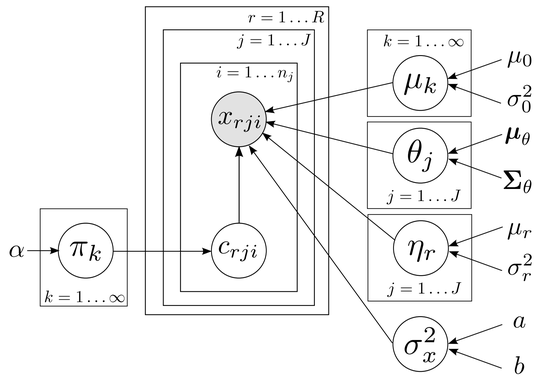
$G_j = \mathrm{T}_j[G_0]$

Modeling of the finishing time

Estimation with Infinite Gaussian Mixtures

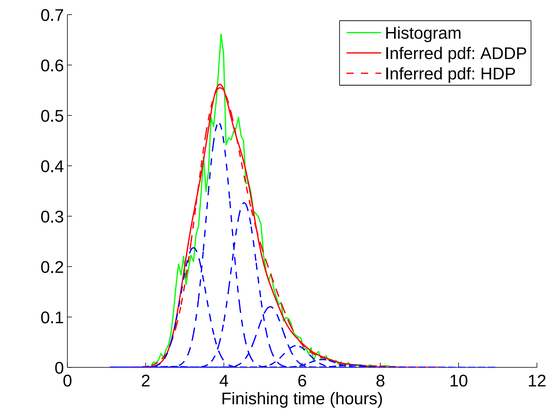
Notation

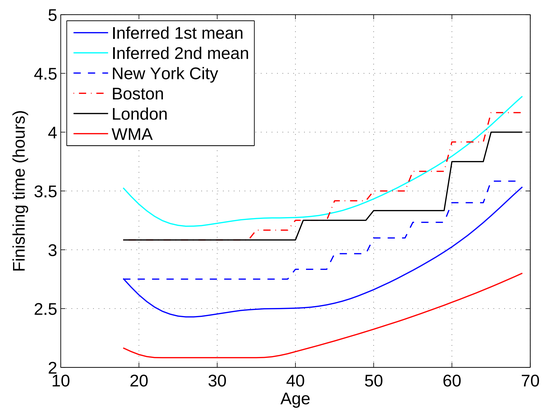
- R number of races
- J: number of runner populations
- n_j number of runners in group j

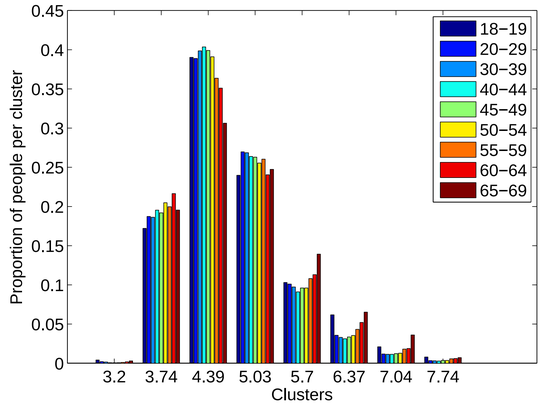


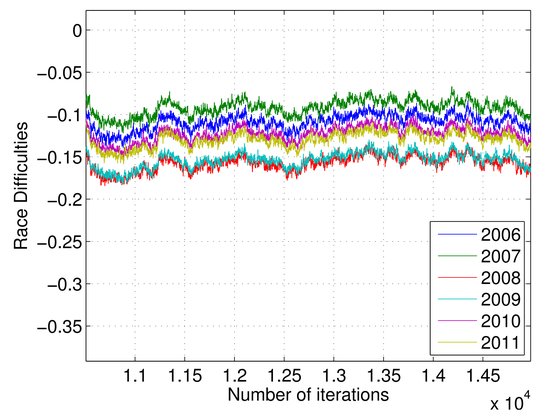
- $G_j = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_{jk}}$
- $x_{rji} | c_{rji} = k, \mu_k, \theta_j, \sigma_x^2, \eta_r \sim \mathcal{N}\left(x_{rji} | \eta_r (\mu_k + \theta_j), \sigma_x^2\right)$
- $\theta \sim \mathcal{N}(0, \Sigma_\theta)$
- $(\Sigma_\theta)_{ij} = \sigma_\theta^2 \cdot \exp\left(-\frac{(i-j)^2}{2\nu^2}\right) + \kappa\delta(i-j)$
- $\eta_r \sim \mathcal{N}(\mu_r, \sigma_r^2)$

Results: Inference of Age-grading curve and race difficulties

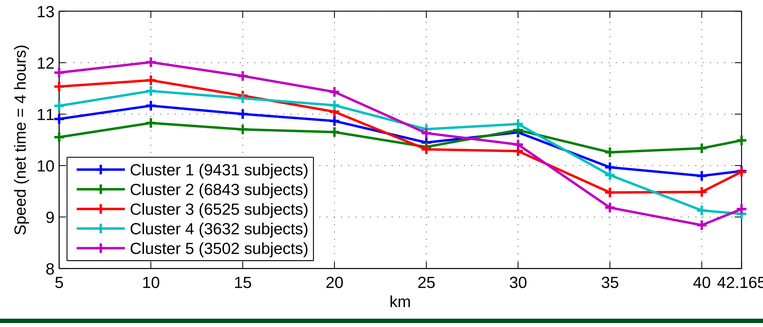


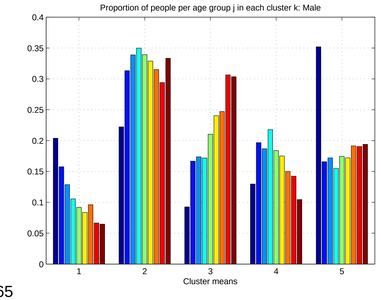






Modeling of temporal running patterns





Conclusions

- Non-parametric model to compare different group distributions
- Inference of robust age grading curves and race difficulty factors
- Inference of hidden temporal running patterns

Funding

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 316861