MARIE CURIE ACTIONS

Initial Training Network "Machine Learning for Personalized Medicine"

Bayesian Non-Parametrics for Personalized Medicine

Melanie F. Pradier, UC3M

Stochastic modeling and graphical models for the analysis and prediction of phenotype interactions

MID-TERM REVIEW, September 15, 2014

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

Who I am

Academical Background

- Telecommunication Engineer, Technical University of Madrid (4 years)
- MSc. in Information Technology, University of Stuttgart (2.5 years)

Professional Experience

- Research Internship at Sony EuTEC, Stuttgart (9 months)
- Research Internship at Sony Corporation, Tokyo (1 year)

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

Previous Research Projects

- 2009 – Bachelor Th *"Statistical Bounds for DOA estimation in Antenna Arrays"*
- 2010 – Sony EuTec: *"Personalization and Recommendation Systems"*
- 2011 – MSc. Th: *"Emotion Recognition from Speech signals and Perception of Music"*
- 2012 – Sony Corporation: *"Adaptive Learning Technologies for Education"*
- 2013 – *"Scalar Quantization for Lossy Source Coding of Continuous Sources"*

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

Research Overview

Aim: Model complex relationships between genetics, epigenetics and environment to infer latent information.

1. Modeling
2. Inference

A1 Biomarker discovery

A2 Data Integration

A3 Causal Mechanism of Disease

A4 Gene-Environment Interactions

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

1. Modeling

Goal: comparative density estimation of group data

Some motivating applications

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

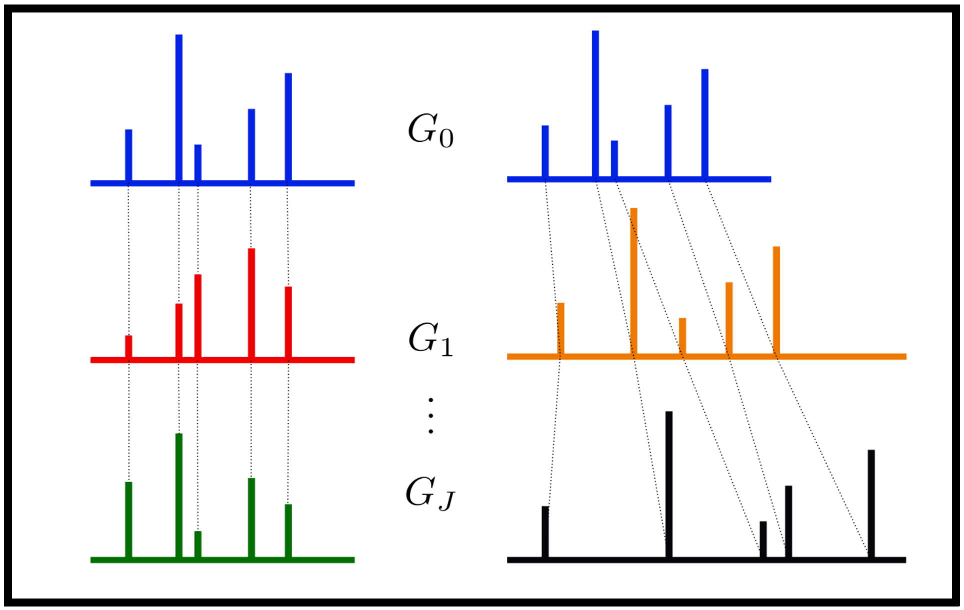
This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 316861

Dependent Dirichlet Process

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

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

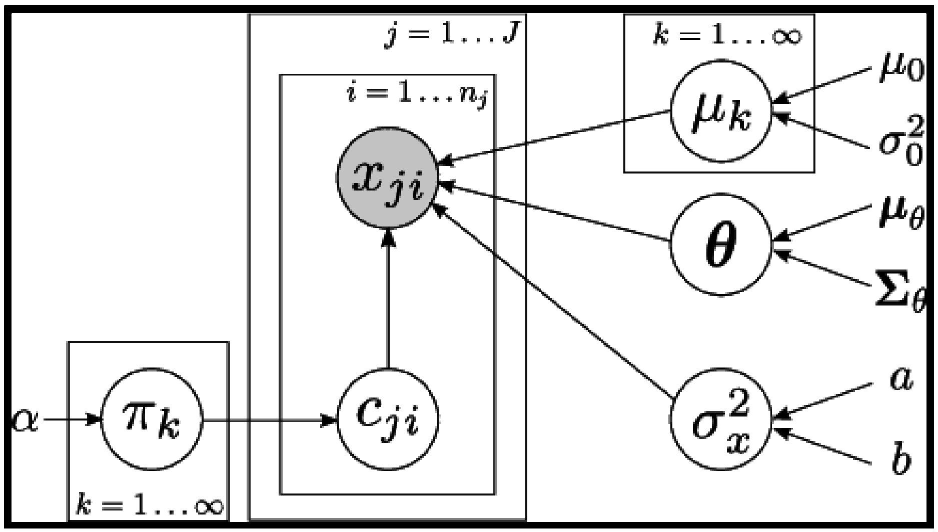
Hierarchical DP vs Single-p DDP



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

Application to Marathon dB

Study of Age/gender impact on marathon performance



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

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

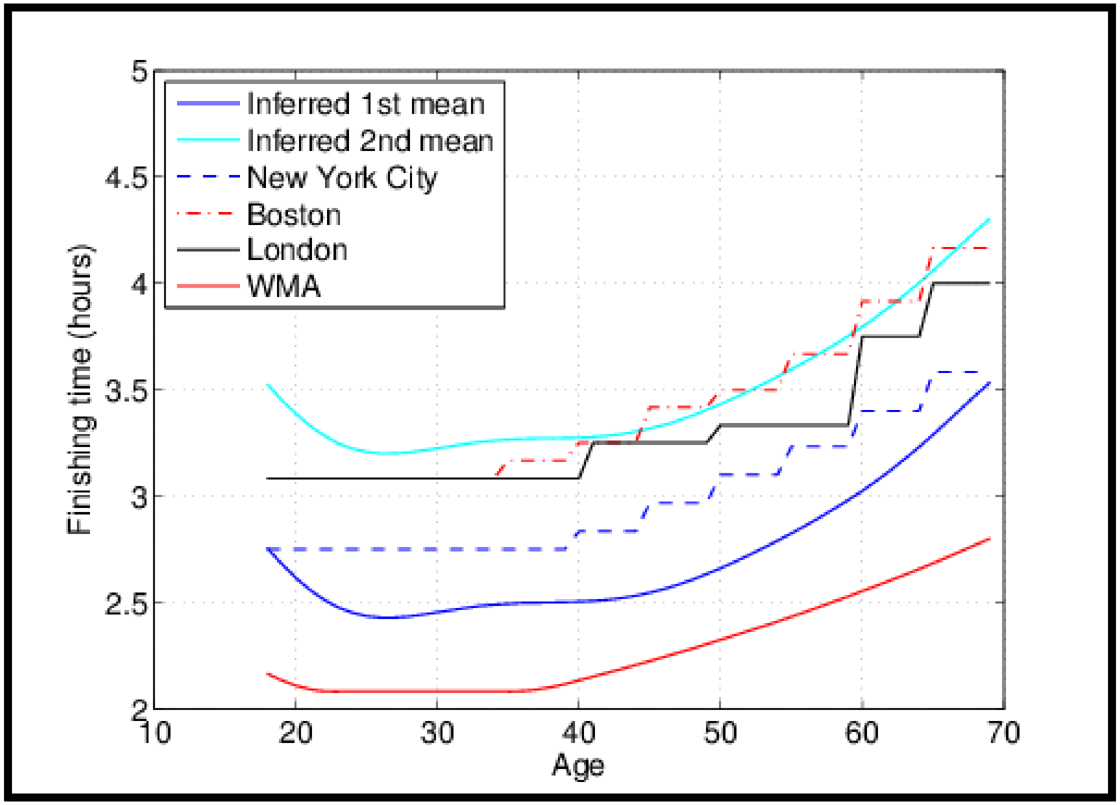
European Commission

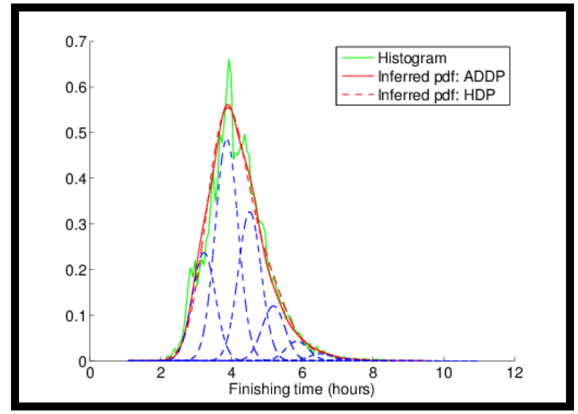
MARIE CURIE ACTIONS

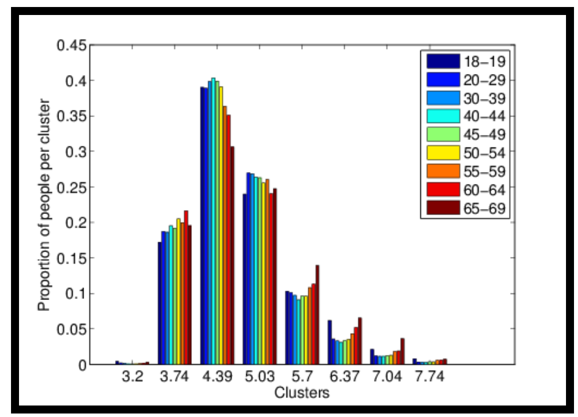
Initial Training Network "Machine Learning for Personalized Medicine"

Machine Learning for Personalized Medicine

Results: Bayesian Non-Parametrics for Marathon Analysis







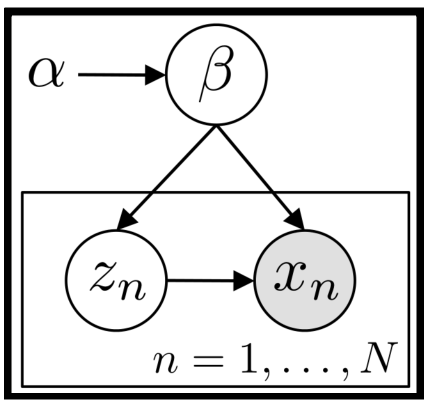
MID-TERM REVIEW, September 15, 2014

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

2. Inference

Scala pBNP toolbox for parallel MCMC Inference in BNP models

$$p(\beta, \mathbf{x}, \mathbf{z}|\alpha) = p(\beta|\alpha) \prod_{n=1}^{N} p(x_n, z_n|\beta)$$

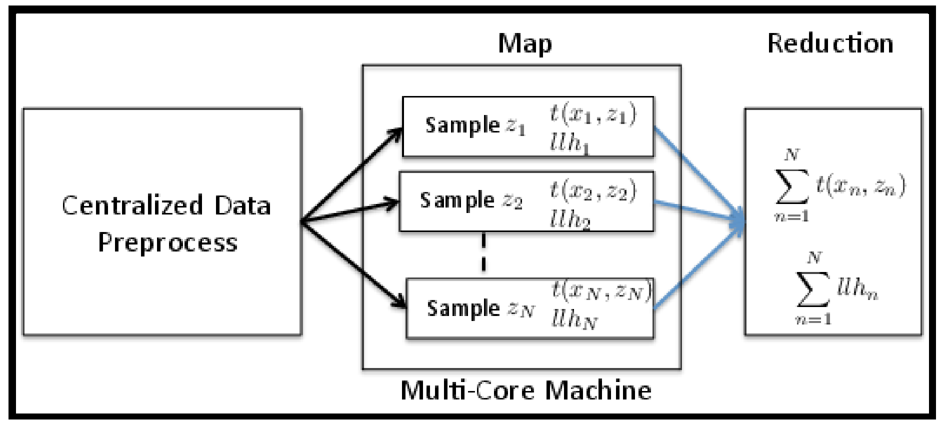


1. Sample β from $p(\beta|\mathbf{z}, \mathbf{x}, \alpha)$
2. Sample z_n from $p(z_n|x_n, \beta)$ for $n = 1, \ldots, N$

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

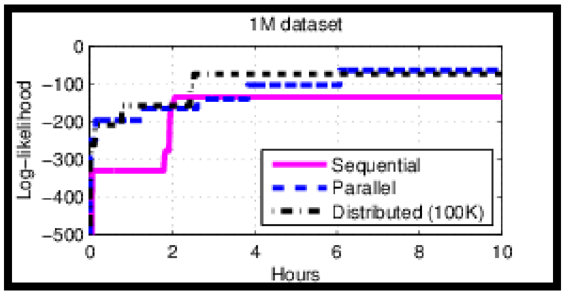
Scala pBNP toolbox

1. Functional programming
2. Parallel/distributed
3. DP/IBP models
4. Easy extension to any lik model
5. Flexible Framework

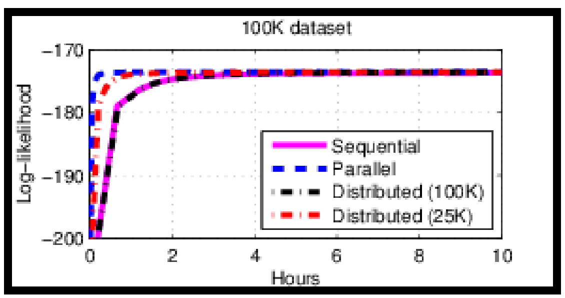


This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 316861

Results: Scala pBNP toolbox



Algorithm \ N	100K	1M	5M	50M
Sequential	0.1349	1.3963	-	-
Parallel	0.0123	0.1397	0.8736	-
Distributed (100K)	0.1795	0.1512	0.2143	1.3429



Algorithm \ N	100K	1M
Sequential	9.9169	-
Parallel	0.6791	26.4195
Distributed (100K)	10.7375	32.8588
Distributed (25K)	3.1215	9.6388

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

Impact of Marie Curie-ITN

1. Quality Training
2. Multidisciplinary work
3. Link to Society

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