Colombian Women's Life Patterns: A Multivariate Density Regression Approach

March 10, 2022

Kimi Kreilgaard and Linea Hedemark
M.Sc.Stud. in non-Nanoscience
Department of not-Chemistry

KØBENHAVNS UNIVERSITET

Introduction

- Goal is to make predictions for multivariate system in a Bayesian framework

- Framework is complicated as it has to reliably take multivariate input

- The case study used in the article as an example is survey data of Columbian women

The Data

Variety of discrete, categorical or binary data types

Predictors:

x1 Age at interview

x2 Area (urban/rural)

x3 Region (in Columbia)

x4 Family Violence

Responses:

z1 Age at Sexual Debut

z2 Age at Union

z3 Age at First Child

z4 Work Status

The Method

- Goal of the method is ambitious;

  - Be able to give correlated, multivariate input

  - Be adaptable to different data sizes and types

The Method

Combination of different methods:

- Adaptive Markov Chain Monte Carlo (aMCMC)
  - Gibbs Sampling
  - Metropolis Hastings
  - Fixed Truncation

- Sequential Monte Carlo (SMC)
  - Adaptive truncation

Transformation of Data to Suit the Method

Link functions

$$z_\ell = h_\ell(\mathbf{y}, \mathbf{x}) = c_\ell(\mathbf{y}, \mathbf{x}) \lfloor \exp(y_\ell) \rfloor, \quad \text{for } \ell = 1, 2,$$

$$z_3 = h_3(\mathbf{y}, \mathbf{x}) = \mathbb{1}_{[0,\infty)}(y_3),$$

Inverse Link functions

$$(l_\ell, u_\ell) = \begin{cases} (\log(x_1 + 1), \infty) & \text{for censored } z_\ell = 0 \\ (\log(z_\ell), \log(z_\ell + 1)) & \text{for uncensored } z_\ell \neq 0 \end{cases}, \quad \text{when } \ell = 1, 2,$$

$$(l_3, u_3) = \begin{cases} (-\infty, 0) & \text{for } z_3 = 0 \\ (0, \infty) & \text{for } z_3 = 1 \end{cases}.$$

Gibbs Sampling

- Useful for multivariate distributions

- The joint: P(x,y) is difficult to sample

- The conditionals: P(x|y) and P(y|x) are easier to sample. One variable is updated at a time, the other is fixed.

- When conditionals are unknown we add a Metropolis-Hastings step.
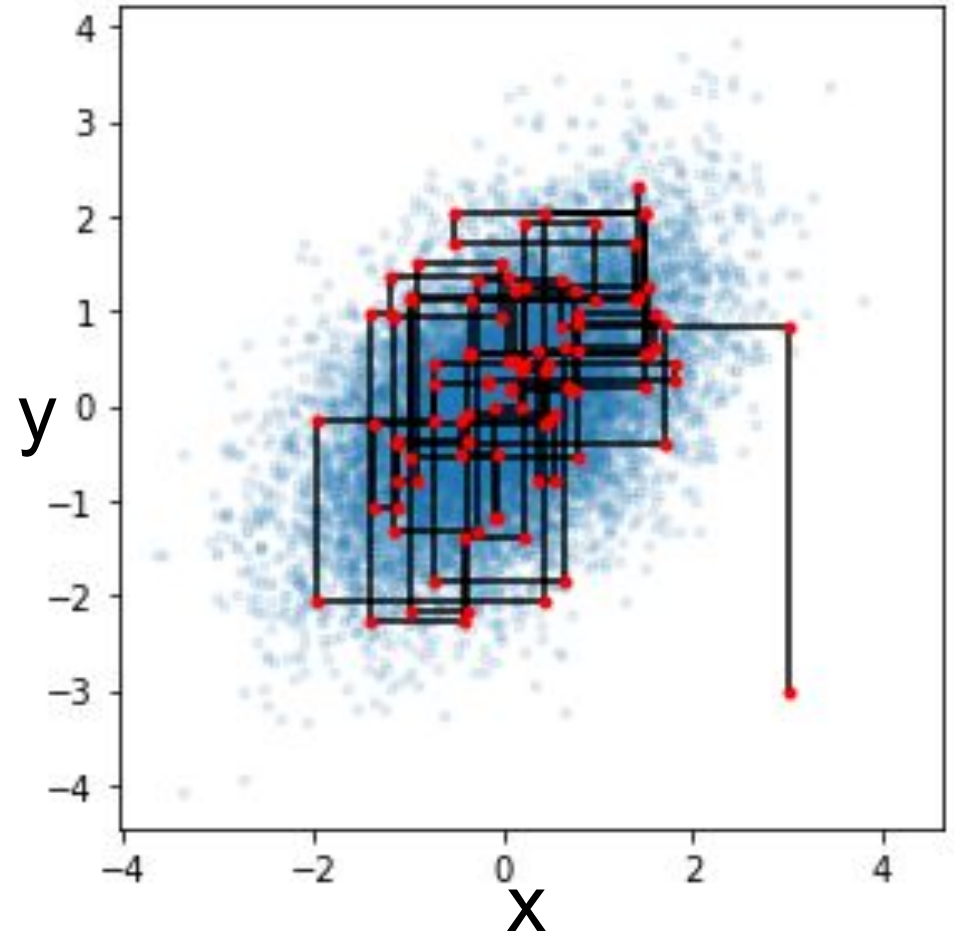
Truncation Level and SMC

- The truncation level is fixed for the aMCMC:

    the predictive model is a mixture of a fixed number of component models

- The truncation level is adaptive SMC:

    SMC samples component models from aMCMC output

    the number of components is chosen by balancing computational burden and quality of approximation to an infinite-dimensional model

Results

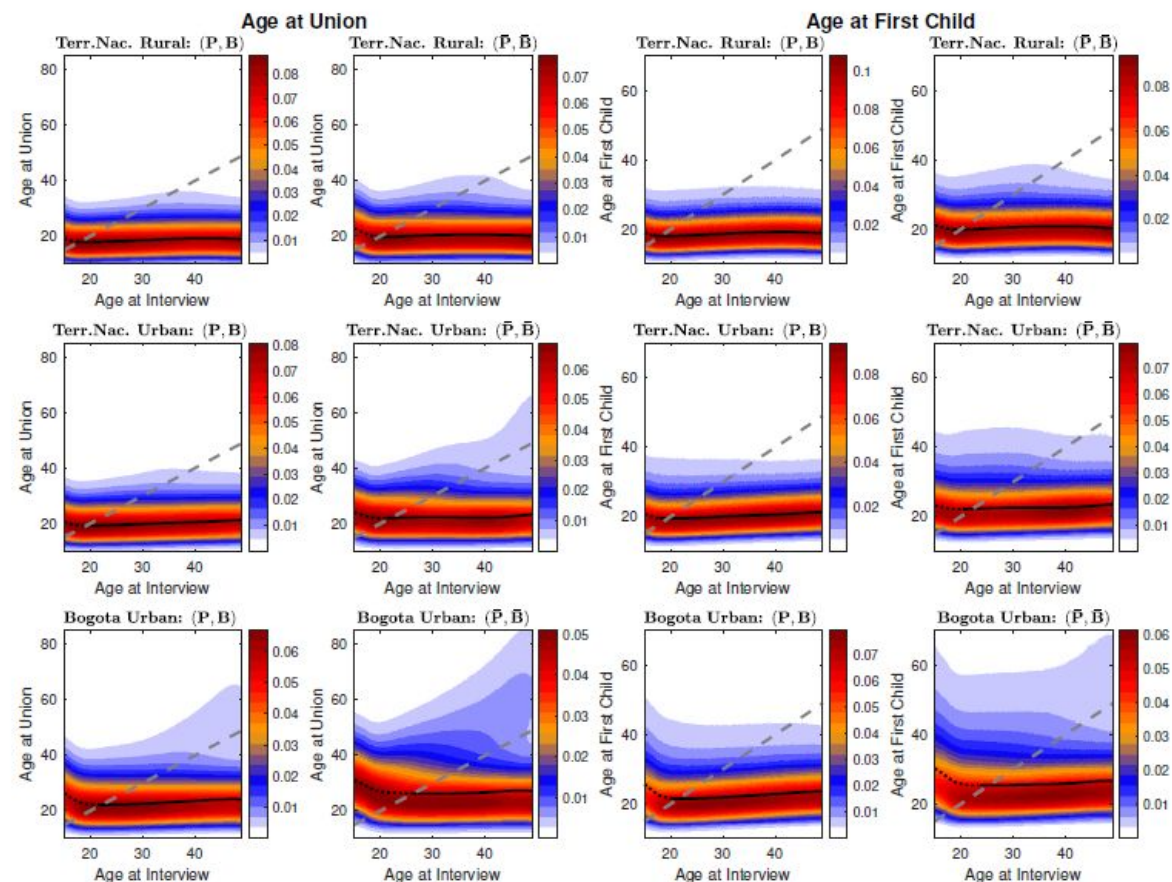An example of the type of results as given by the model

Figure 6: Predictive densities of the ages at union and first child as functions of *Age* for women who grew up in violent $(\mathbf{P}, \mathbf{B})$ and non-violent $(\bar{\mathbf{P}}, \bar{\mathbf{B}})$ families. Results are reported for urban and rural areas of the least developed region (Territorios nacionales) and for the capital (Bogota). The region above the dashed line indicates when age at event exceeds *Age*. The black line is the posterior median function.

KØBENHAVNS UNIVERSITET

Results

An example of the type of results as
given by the model