Augmented Ensemble MCMC sampling in Factorial Hidden Markov Models

Kaspar Märtens[1], Michalis K. Titsias[2], Christopher Yau[3,4]

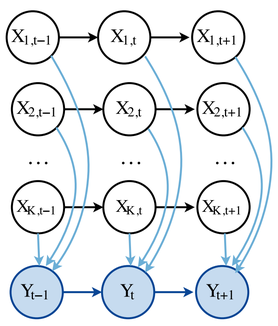
[1] Department of Statistics, University of Oxford
[2] Athens University of Economics and Business
[3] Centre for Computational Biology, University of Birmingham
[4] Alan Turing Institute, London, United Kingdom

Motivation

Bayesian inference for Factorial Hidden Markov Models models is challenging due to the exponentially sized latent variable space. Standard Monte Carlo samplers can have difficulties effectively exploring the posterior landscape and are often restricted to exploration around localised regions that depend on initialisation.

Background

Factorial HMM is an extended version of the standard HMM with multiple latent chains.



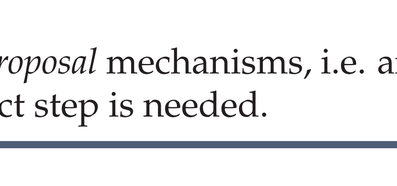
Ensemble MCMC methods such as parallel tempering can alleviate this problem. Suppose our goal is to sample from a target density π. Instead of sampling $\mathbf{x} \sim \pi(\cdot)$, ensemble MCMC introduces an extended product space $(\mathbf{x}_1, \ldots, \mathbf{x}_K)$ with a new target density π^* defined as follows

$$\pi^*(\mathbf{x}_1, \ldots, \mathbf{x}_K) = \prod_{k=1}^{K} \pi_k(\mathbf{x}_k),$$

Specifically, **parallel tempering** introduces a temperature ladder $1.0 = T_1 < \ldots < T_K$ and associates a temperature with each chain. Denoting the inverse temperature $\beta_k := 1/T_k$,

$$\pi^*(\mathbf{x}_1, \ldots, \mathbf{x}_K) = \prod_{k=1}^{K} \pi(\mathbf{x}_k)^{\beta_k}$$

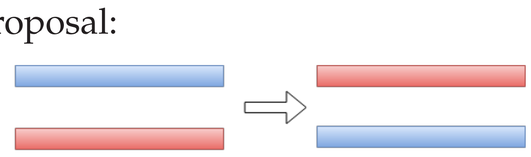
Tempered targets are less peaked. Therefore higher temperature chains explore the space well and do not get stuck.



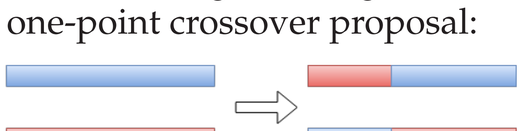
Here the key question is how to exchange information between the chains in the ensemble.

Existing approaches for **information exchange**:

Swap proposal:



Proposal schemes via genetic algorithms. For example, a one-point crossover proposal:



These are *proposal* mechanisms, i.e. an additional accept/reject step is needed.

Augmented Ensemble MCMC

We propose an augmented Gibbs sampler to exchange information between a pair of chains $\pi_i(\mathbf{x}_i)$ and $\pi_j(\mathbf{x}_j)$.

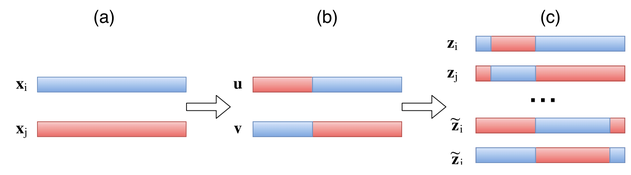
Let $\text{CR}(\mathbf{x}_i, \mathbf{x}_j)$ denote the set of all crossovers between the vectors $\mathbf{x}_i$ and $\mathbf{x}_j$. We introduce two auxiliary variables $\mathbf{u}$ and $\mathbf{v}$, that live in the same space as $\mathbf{x}_i$ and $\mathbf{x}_j$, drawn from an auxiliary distribution $p(\mathbf{u}, \mathbf{v}|\mathbf{x}_i, \mathbf{x}_j)$, which we define to be uniform on the set $\text{CR}(\mathbf{x}_i, \mathbf{x}_j)$.

Using the auxiliary variables we can exchange information between $\mathbf{x}_i$ and $\mathbf{x}_j$ through the intermediate step of sampling the auxiliary variables $(\mathbf{u}, \mathbf{v})$, based on the following two-step Gibbs procedure:

1. Generate $(\mathbf{u}, \mathbf{v}) \sim p(\mathbf{u}, \mathbf{v}|\mathbf{x}_i, \mathbf{x}_j)$
2. Generate $(\mathbf{x}_i, \mathbf{x}_j) \sim p(\mathbf{x}_i, \mathbf{x}_j|\text{rest})$, where

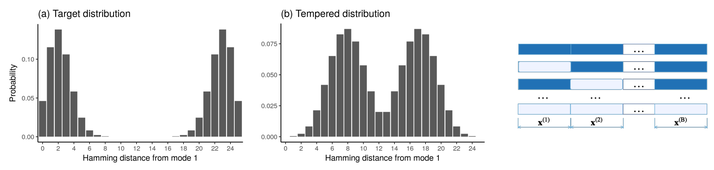
$$\begin{aligned} p(\mathbf{x}_i, \mathbf{x}_j|\text{rest}) &= \frac{1}{Z}\pi_i(\mathbf{x}_i)\pi_j(\mathbf{x}_j)p(\mathbf{u}, \mathbf{v}|\mathbf{x}_i, \mathbf{x}_j) \\ &= \frac{1}{Z}\pi_i(\mathbf{x}_i)\pi_j(\mathbf{x}_j)p(\mathbf{x}_i, \mathbf{x}_j|\mathbf{u}, \mathbf{v}) \\ &= \frac{1}{Z}\pi_i(\mathbf{x}_i)\pi_j(\mathbf{x}_j)I((\mathbf{x}_i, \mathbf{x}_j) \in \text{CR}(\mathbf{u}, \mathbf{v})) \end{aligned}$$

where the normalising constant $Z = Z(\mathbf{u}, \mathbf{v})$ is computed explicitly by summing over $\text{CR}(\mathbf{u}, \mathbf{v})$. Schematically:



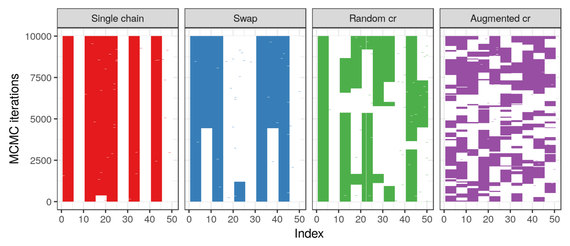
Toy example

We consider the following multimodal toy sampling problem, where the target distribution is binary and has multiple separated modes.



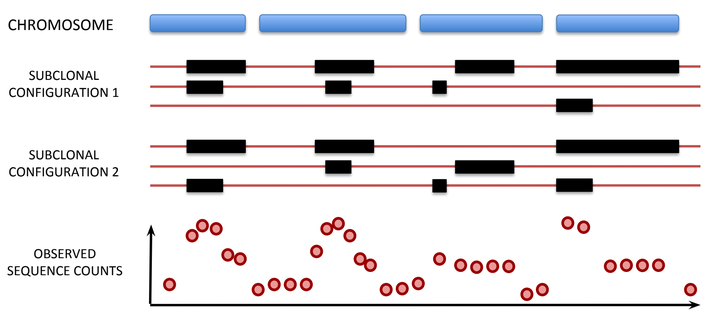
We compare our augmentation scheme (*augmented crossover*) with a *single-chain* sampler and two additional exchange moves for ensemble samplers: *swap* move, and a uniformly proposed crossover (*random cr*).

Heatmap for the traces of $\mathbf{x}$ for this toy example:



Tumor deconvolution

Factorial HMM is naturally suited for the cancer genomics application below, where our goal is to infer subpopulations among tumor cells.



We note that a poorly mixing sampler which is exploring only one of the possible latent explanations could lead to misleading conclusions regarding the subclonal architecture of a tumor.

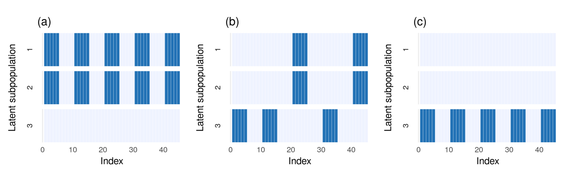
Model

Emission likelihood for the factorial HMM:

$$y_t|\mathbf{x}_t, \mathbf{w}, h \sim \mathcal{N}\left(h \sum_{k=1}^{K} w_k x_{k,t},\ \sigma^2\right)$$

Simulation study

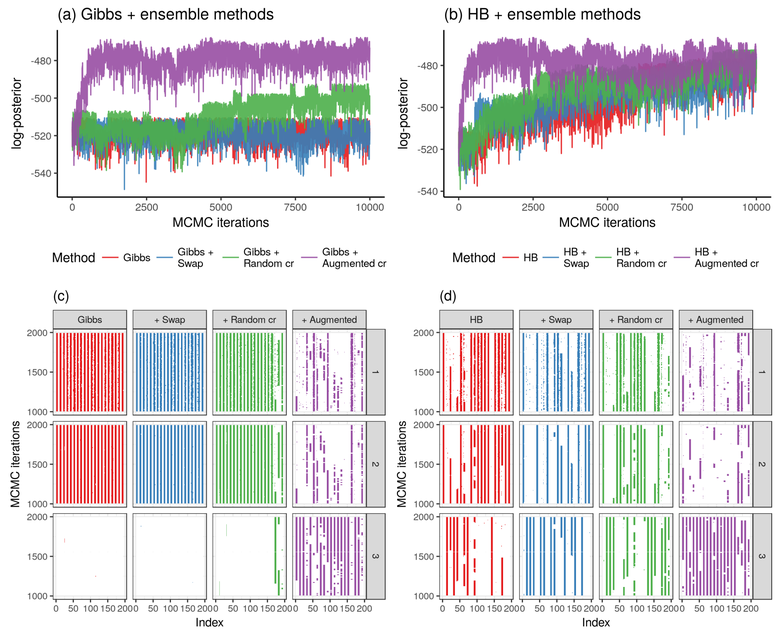
First, we investigated the performance of sampling schemes for FHMMs in the presence of multimodality in a controlled setting. We generated data from the model with $K = 3$ and weights such that $w_1 + w_2 \approx w_3$. As a result, all of the following configurations of $\mathbf{X}$ are plausible:



For inference in FHMMs, we considered two single-chain samplers

- One-row updates of $\mathbf{X}$ while keeping the rest fixed ("Gibbs")
- Hamming Ball sampling of $\mathbf{X}$ ("HB") (Titsias and Yau, 2014; 2017)

and ensemble versions of both of these (*swap*, *random cr*, and *augmented cr* exchange moves).



Real tumor data analysis

Next we illustrate the utility of our sampling approach on the whole-genome tumor sequencing data for bladder cancers.