Augmented Ensemble MCMC with applications to Factorial HMMs

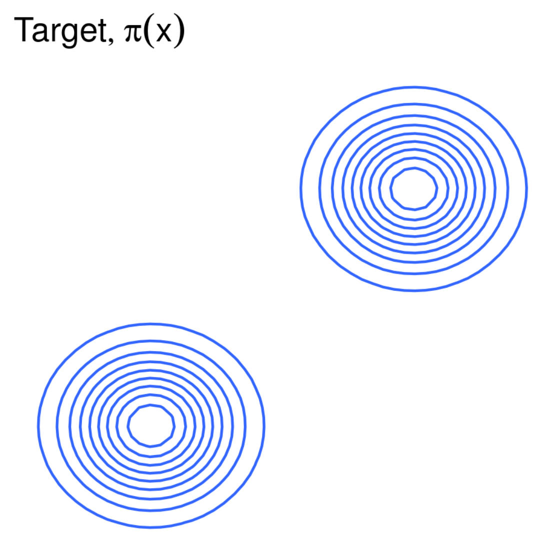
Kaspar Märtens, Michalis Titsias, Christopher Yau

UNIVERSITY OF OXFORD

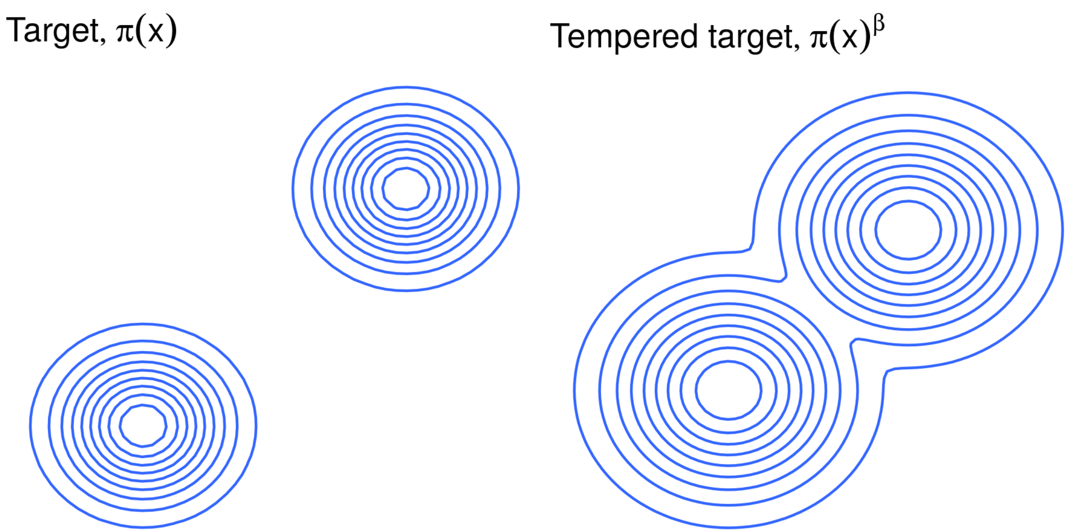
DEPARTMENT OF STATISTICS

Background

Bayesian inference for high-dimensional models is challenging: it is difficult to explore multimodal distributions.



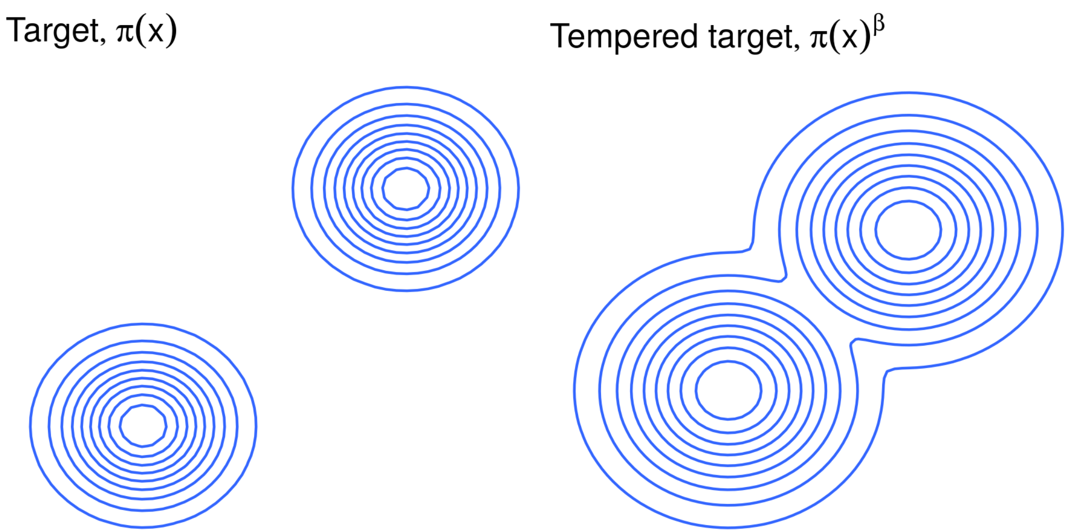
Parallel tempering



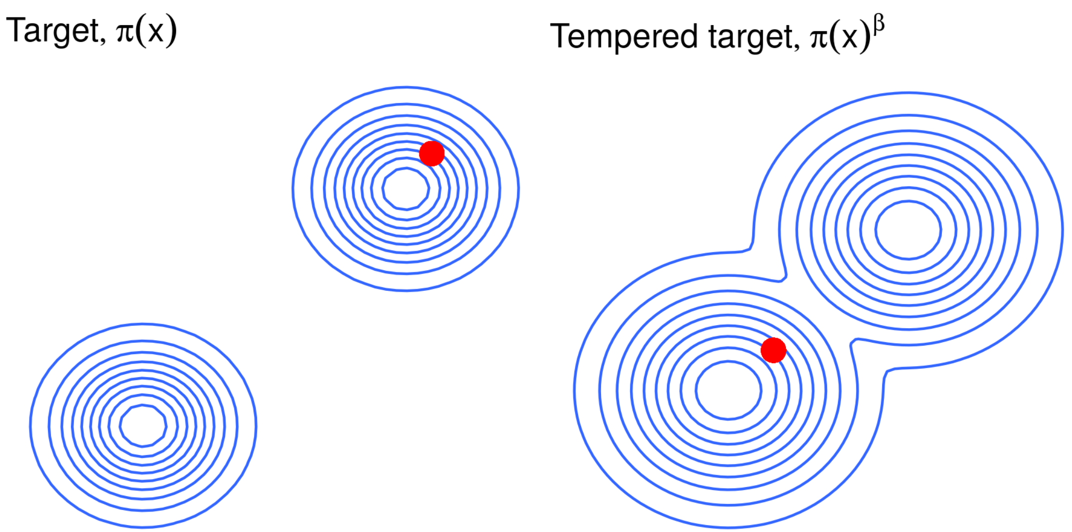
Parallel tempering

Instead of simply targeting $\pi(x)$, introduce a temperature ladder $T_1 < T_2 < \ldots < T_K$ and run multiple MCMC chains, with chain k targeting

$$\pi(x)^{\beta_k} \text{ where } \beta_k = 1/T_k$$



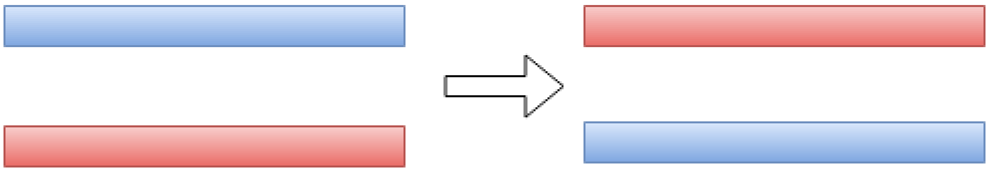
How to exchange information between an ensemble of chains



How to exchange information between an ensemble of chains

How to exchange information between an ensemble of chains

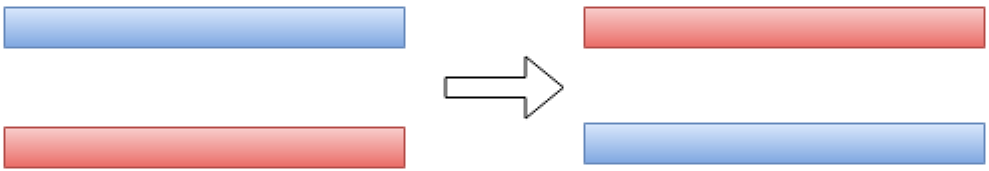
1. **Swap move**: Propose to swap (using Metropolis-Hastings to accept/reject)



Inefficient in a high-dimensional setting

How to exchange information between an ensemble of chains

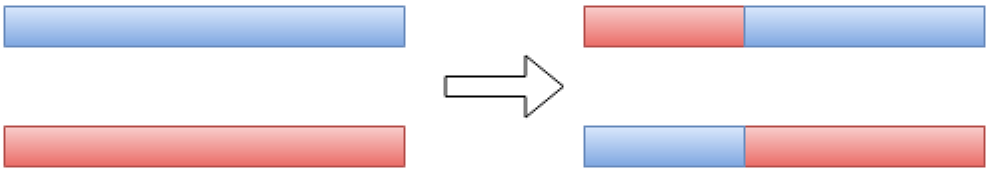
1. **Swap move**: Propose to swap (using Metropolis-Hastings to accept/reject)



   Inefficient in a high-dimensional setting

2. **Genetic algorithms**:

   One-point **crossover** (using Metropolis-Hastings to accept/reject)



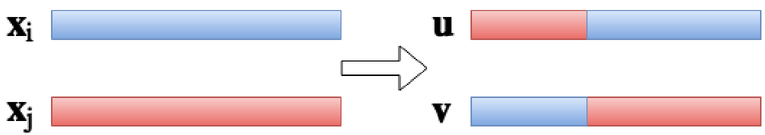
Augmented Ensemble MCMC

We construct an auxiliary variable Gibbs sampler, using a one-point crossover

Augmented Ensemble MCMC

We construct an auxiliary variable Gibbs sampler, using a one-point crossover

1. Generate auxiliary $(\mathbf{u}, \mathbf{v}) \sim p(\mathbf{u}, \mathbf{v} | \mathbf{x}_i, \mathbf{x}_j)$
    - Uniform distribution over all crossovers of $(\mathbf{x}_i, \mathbf{x}_j)$

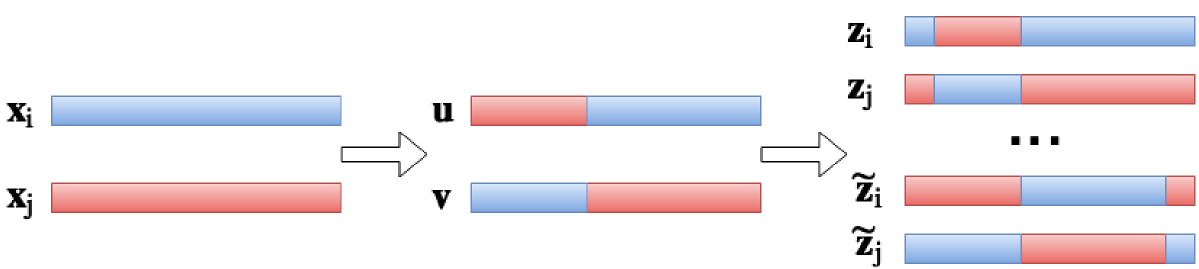


Augmented Ensemble MCMC

We construct an auxiliary variable Gibbs sampler, using a one-point crossover

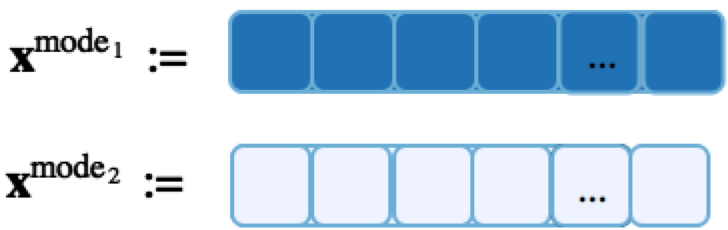
1. Generate auxiliary $(\mathbf{u}, \mathbf{v}) \sim p(\mathbf{u}, \mathbf{v} | \mathbf{x}_i, \mathbf{x}_j)$
   - Uniform distribution over all crossovers of $(\mathbf{x}_i, \mathbf{x}_j)$
2. Generate $(\mathbf{x}_i, \mathbf{x}_j) \sim p(\mathbf{x}_i, \mathbf{x}_j | \text{rest})$, where

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



Toy example

Consider the following distribution over binary vectors with two separated modes:

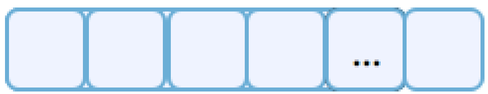


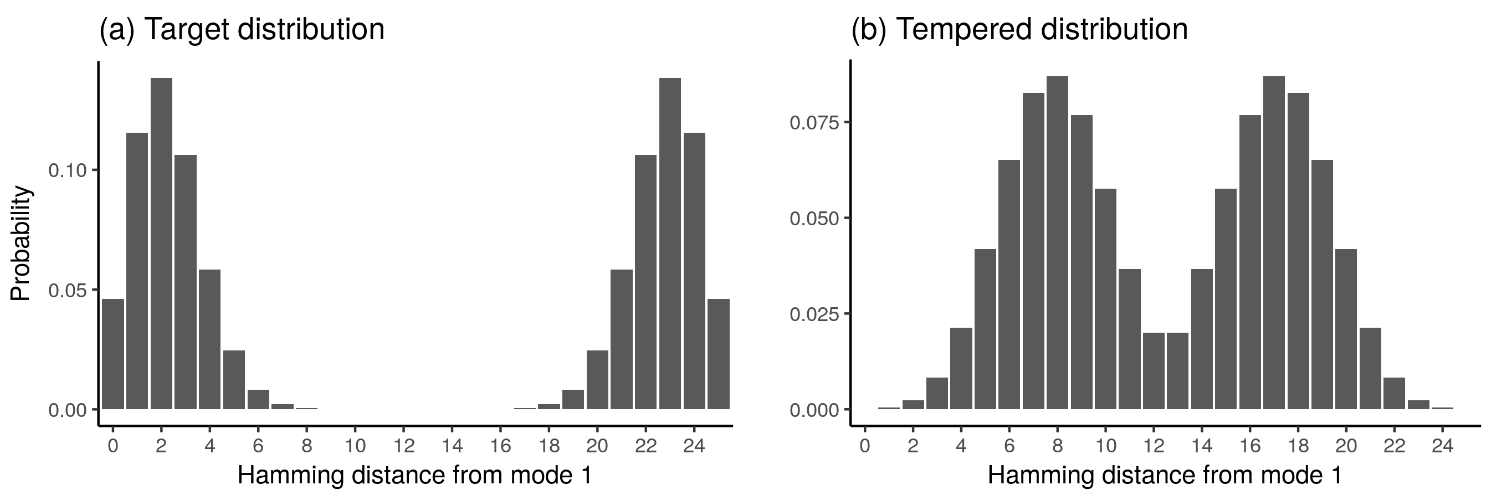
Toy example

Consider the following distribution over binary vectors with two separated modes:

$\mathbf{x}^{\text{mode}_1} :=$

$\mathbf{x}^{\text{mode}_2} :=$

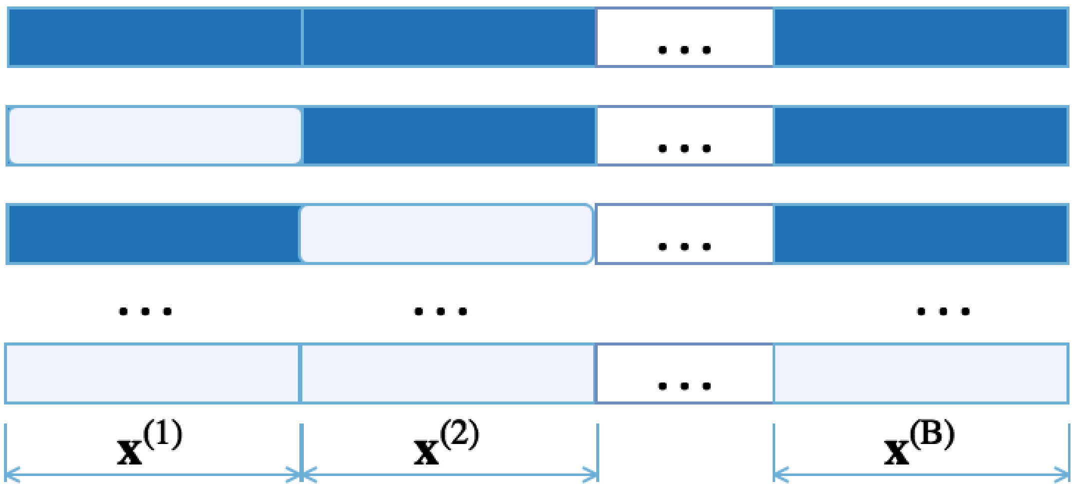




Toy example extended

Divide $\mathbf{x}$ into B blocks, within each block bimodal distribution.

Results in a total of 2^B modes:



Toy example: comparison of samplers

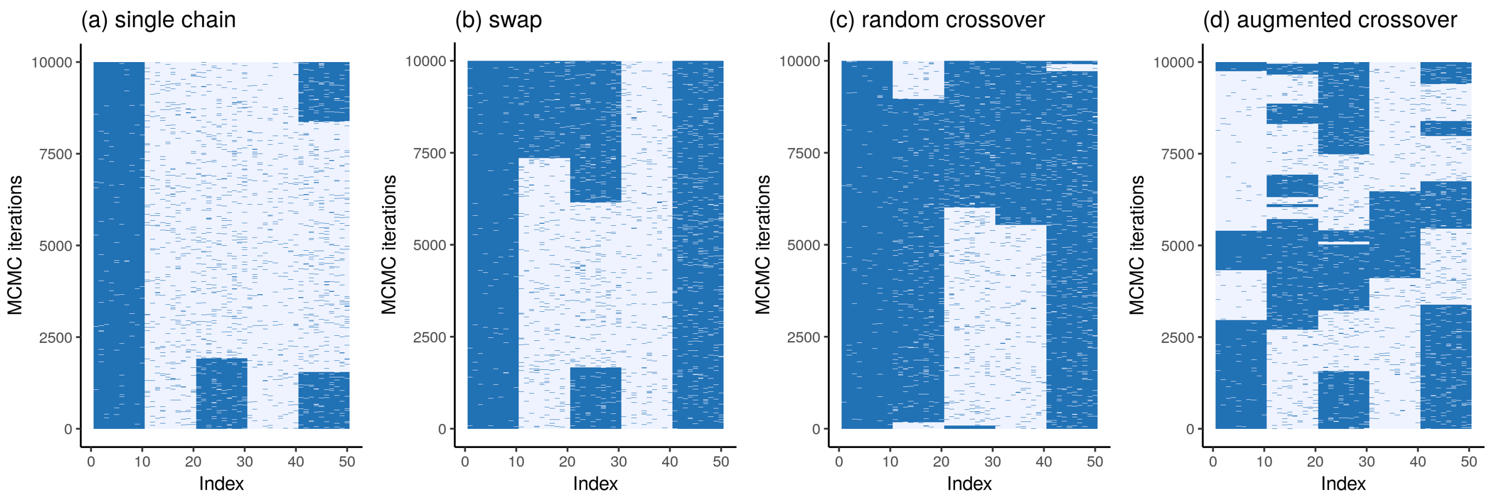
- Single chain Gibbs sampler
- Ensemble of Gibbs samplers (two chains: $T_1 = 1$, $T_2 = 4$),

Toy example: comparison of samplers

- Single chain Gibbs sampler
- Ensemble of Gibbs samplers (two chains: $T_1 = 1$, $T_2 = 4$), using the following exchange moves
  - Swap (accept/reject)
  - Uniformly chosen crossover (accept/reject)
  - Augmented crossover

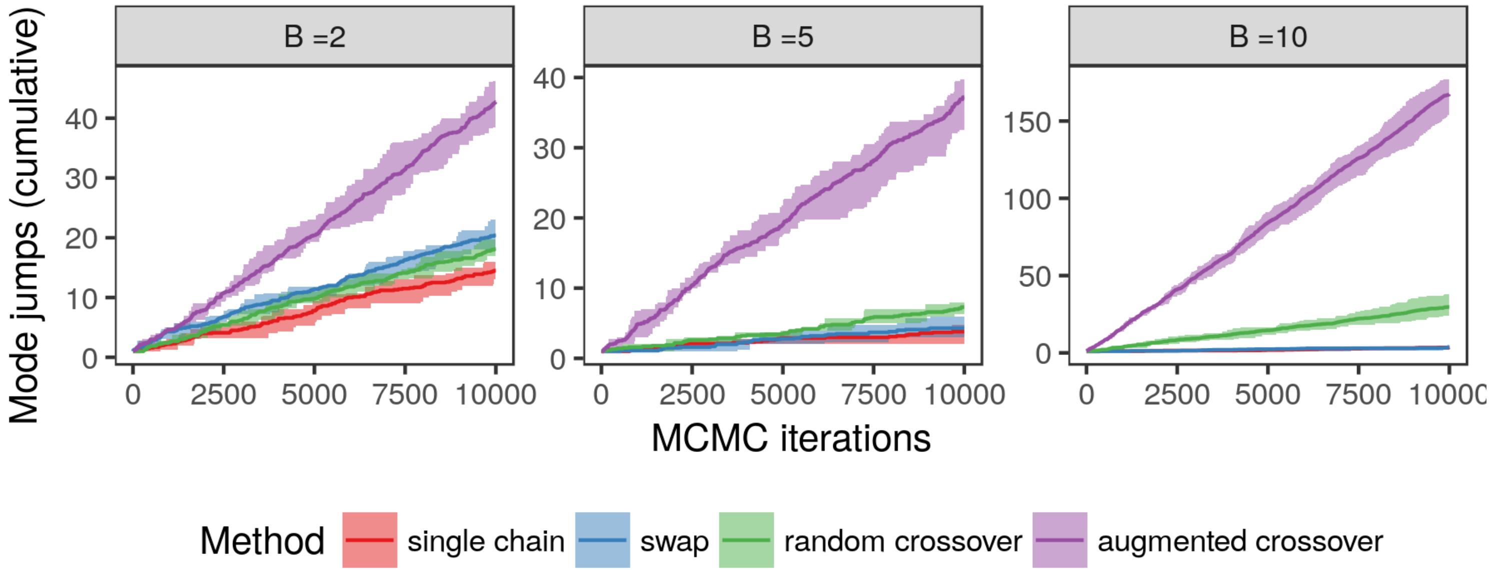
Toy example: comparison of samplers

- Single chain Gibbs sampler
- Ensemble of Gibbs samplers (two chains: $T_1 = 1$, $T_2 = 4$), using the following exchange moves
    - Swap (accept/reject)
    - Uniformly chosen crossover (accept/reject)
    - Augmented crossover



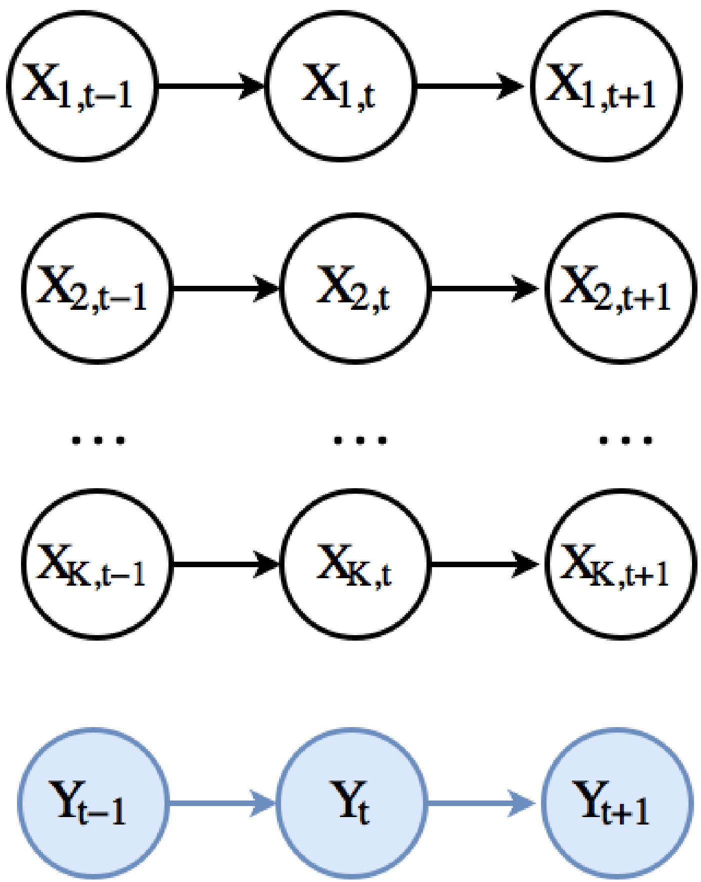
Toy example: comparison of samplers

Number of mode jumps (cumulative)

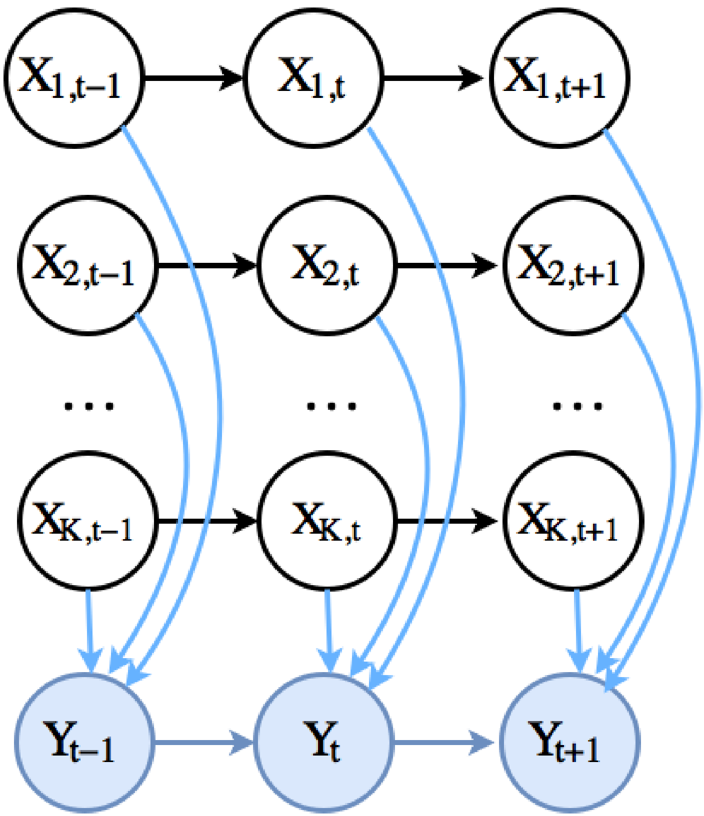


Factorial HMM

Extension of HMM: multiple latent sequences



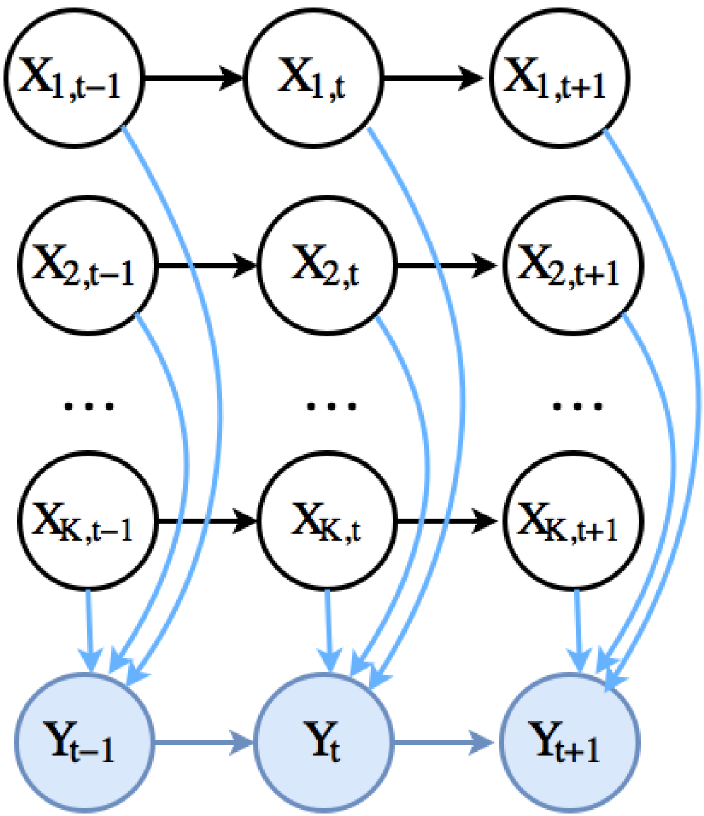
Factorial HMM



$$(w_1, \ldots, w_K) \sim \mathrm{Dir}(\alpha, \ldots, \alpha)$$

$$Y_t | X_{:,t}, w \sim \mathcal{N}\left(\sum_{k=1}^{K} w_k X_{k,t},\ \sigma^2\right)$$

Factorial HMM



$$(w_1, \ldots, w_K) \sim \mathrm{Dir}(\alpha, \ldots, \alpha)$$

$$Y_t | X_{:,t}, w \sim \mathcal{N}\left(\sum_{k=1}^{K} w_k X_{k,t},\ \sigma^2\right)$$

Inference by Gibbs sampling:

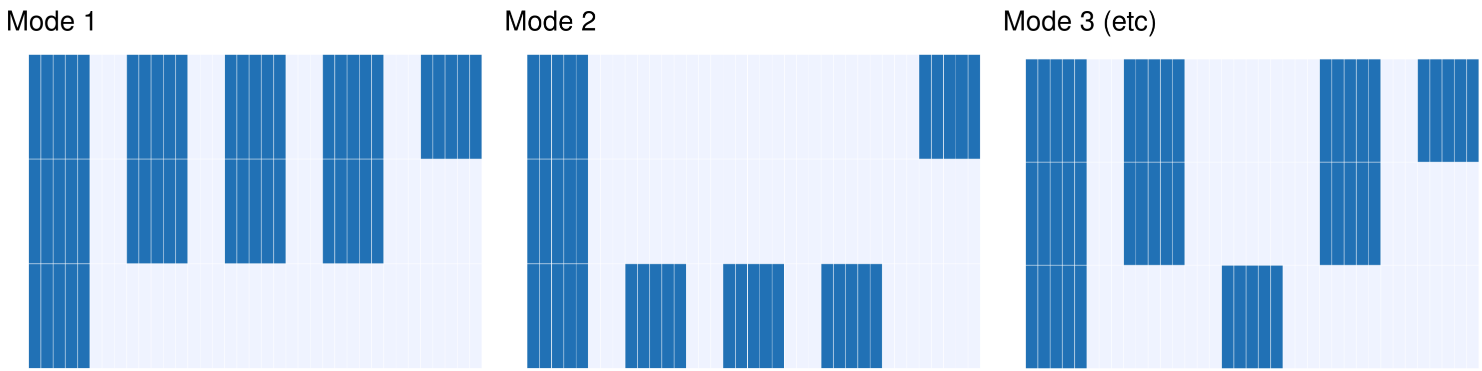
1. Sample $\mathbf{w}|Y, X$
   - e.g. by random walk MH
2. Sample $X|\mathbf{w}, Y$
   - Full FB recursion becomes quickly infeasible, due to complexity $O(2^{2K}T)$
   - Solutions:
     - Update one row of X conditional on the rest
     - Hamming Ball sampling (Titsias and Yau 2016)

Factorial HMM example

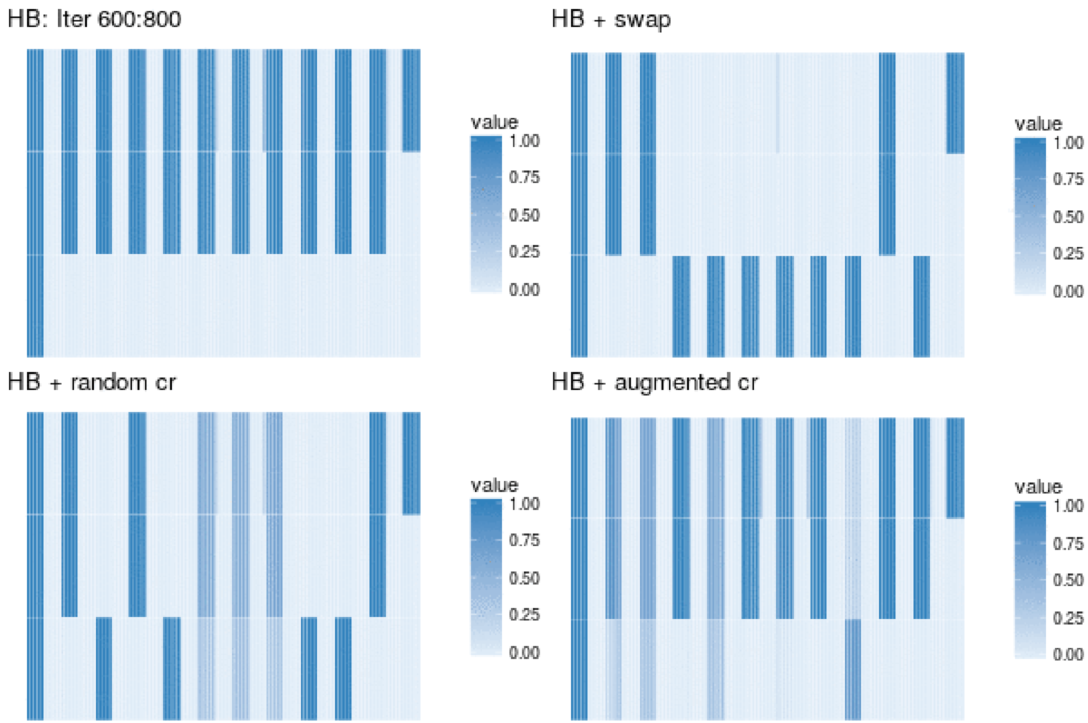
Motivated by a cancer genomics application.

Consider the following example ($K = 3$) with multiple underlying explanations, where $w_1 + w_2 \approx w_3$

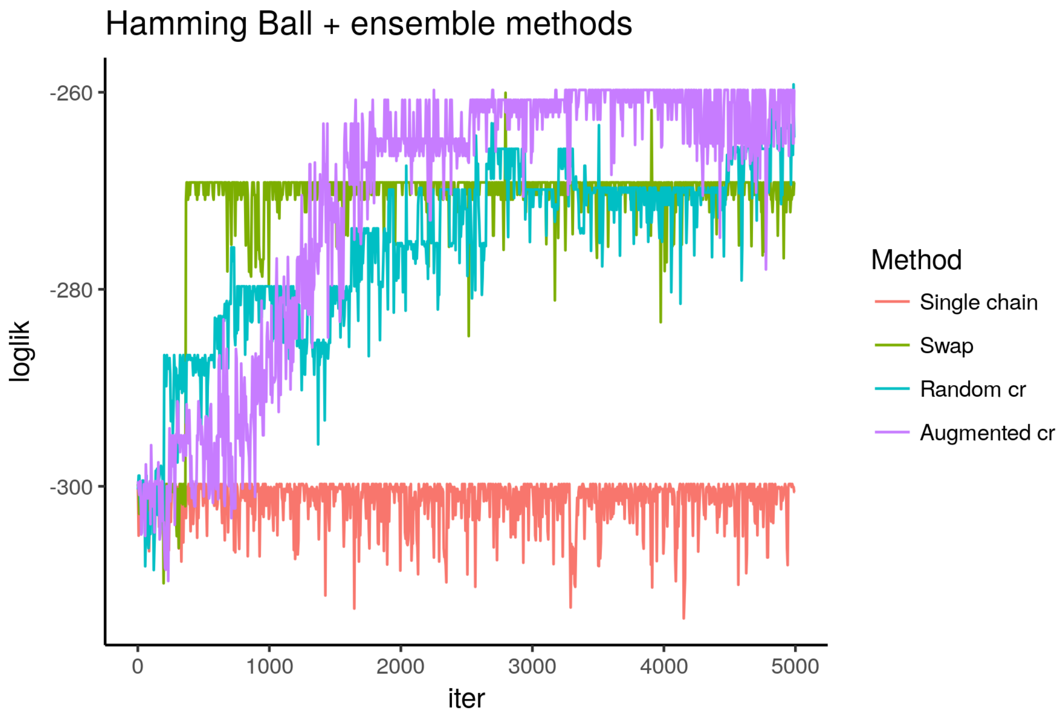
$$Y_t | X_{:,t}, w \sim \mathcal{N}\left(\sum_{k=1}^{K} w_k X_{k,t},\ \sigma^2\right)$$



Factorial HMM example



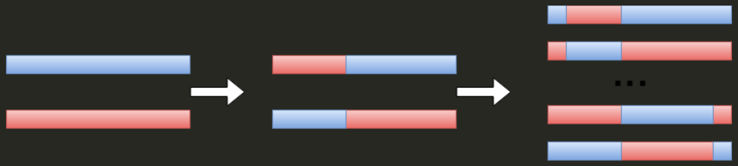
Factorial HMM example



Augmented Ensemble MCMC with applications to Factorial HMMs

Kaspar Märtens, Michalis Titsias, Christopher Yau

https://arxiv.org/abs/1703.08520



kaspar.martens@gmail.com

@kasparmartens