Tempered Simplex Sampler

Eleni Bakra[1] and Nial Friel[2]

[1]MRC Biostatistics Unit, Cambridge, U.K.
[2]University College Dublin, Dublin, Ireland.

August 30, 2009

Issues

Standard MCMC methods use a single Markov chain to explore the target distribution.

- Highly correlated variables $\longrightarrow$ highly correlated output, slow mixing of the chain.
- Multi-modal target $\longrightarrow$ difficult to move between the modes.
- Choice of the proposal distribution is crucial for the performance of the algorithm.

Population MCMC methods

Parallel MCMC where the chains interact with one another.

The idea is to consider an augmented state space with stationary distribution,

$$\pi(\mathbf{x}) = \prod_{i=1}^{n} \pi_i(\mathbf{x}_i),$$

where $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and one or more of the $\pi_i(\cdot)$'s is exactly $\pi(\cdot)$.

Evolutionary Monte Carlo - Liang & Wong (2003)

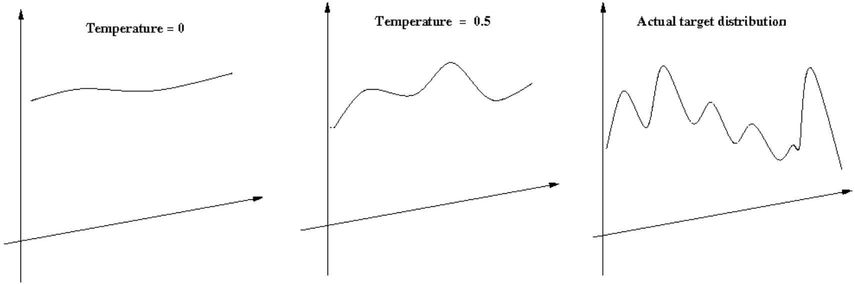
Here, $\pi_i(\cdot) \propto \{\pi(\cdot)\}^{T_i}$ for $i = 1, \ldots, n$ where $T_1 = 0, \ldots, T_n = 1$.

The intermediate distributions $\pi_i(\cdot)$ are flattened versions of the target.

The algorithm works by

1. Updating in turn each $\mathbf{x}_i$ from $\pi_i(\cdot)$.
2. Proposing to swap information from chain i to chain i^*, ie at time t

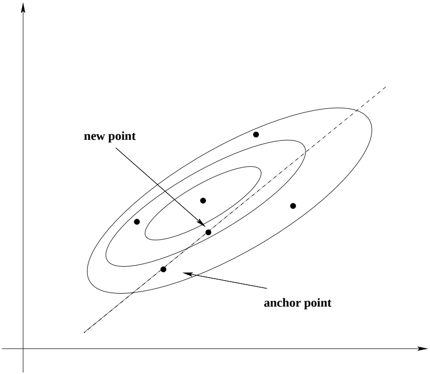
$$(\mathbf{x}_1^{(t)}, \ldots, \mathbf{x}_i^{(t)}, \ldots, \mathbf{x}_{i^*}^{(t)}, \ldots, \mathbf{x}_n^{(t)}) \rightarrow (\mathbf{x}_1^{(t)}, \ldots, \mathbf{x}_{i^*}^{(t)}, \ldots, \mathbf{x}_i^{(t)}, \ldots, \mathbf{x}_n^{(t)})$$


Temperature = 0
Temperature = 0.5
Actual target distribution

Snooker algorithm - Roberts & Gilks (1994)

Choose two points from current population.

Move one of the chosen points along the direction of the line formed by the two points.

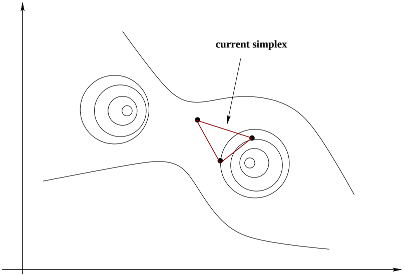


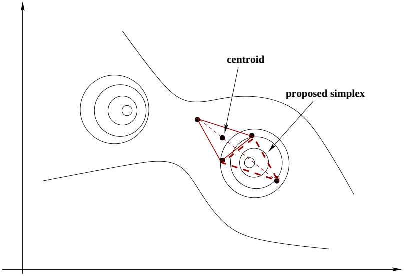
The simplex method - Nelder and Mead (1965)

Main idea is to move the vertex with the smallest target value from a low probability valley to a higher probability area.

- **Simplex:** a $n+1$ polytope in n dimensional space, i.e a triangle in two dimensions, a tetrahedron in three dimensions etc.
- A local maximum is achieved by moving points of the simplex around the surface at each iteration in a deterministic way.

Reflection move of the simplex method.





The simplex sampler

Several MCMC chains that run in parallel and interact using ideas of the Nelder and Mead simplex method.

Three different type of moves are applied:

- Reflection
- Expansion and Contraction.
- Metropolis-Hastings update.

Reflection move

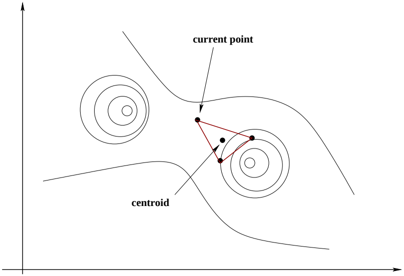
At time t

1. Choose $\mathbf{x}_i^{(t)}, i = 1, \ldots, n$ which tends to be the one with the smallest target probability compared to the rest of the population.
2. Estimate the centroid, $\bar{\mathbf{x}}^{(t)}$
3. Draw α from $p_\alpha(\cdot)$.
4. Estimate $\mathbf{y}_i = (1+\alpha)\,\bar{\mathbf{x}}^{(t)} - \alpha \mathbf{x}_i^{(t)}$
5. Accept the proposed point with probability

$$\alpha\left(\mathbf{x}_i^{(t)}, \mathbf{y}_i\right) = \min\left\{1, \frac{\pi(\mathbf{y}_i)}{\pi(\mathbf{x}_i^{(t)})} \frac{q(\mathbf{y}_i, \mathbf{x}_i^{(t)})}{q(\mathbf{x}_i^{(t)}, \mathbf{y}_i)}\right\}$$

where $q(\mathbf{x}_i^{(t)}, \mathbf{y}_i)$ is the probability of choosing the current point times the probability of choosing α.

Reflection move



Expansion/Contraction move

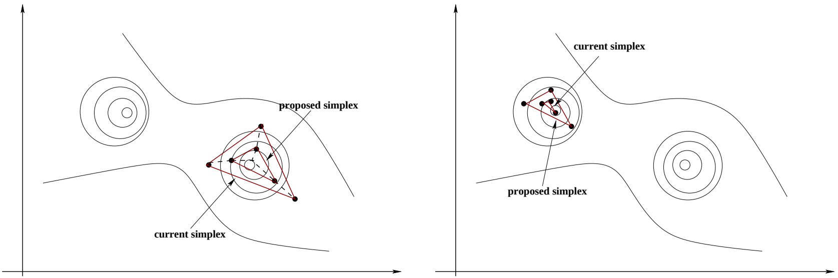
1. Choose the expansion/contraction coefficient β from some distribution $p_\beta(\cdot)$.
2. Propose new vertex of the simplex

$$\mathbf{y}_i = \beta(\mathbf{x}_i^{(t)} - \bar{\mathbf{x}}^{(t)}) + \bar{\mathbf{x}}^{(t)}, \quad i = 1, \ldots, n$$

3. Accept proposed move with probability

$$\min\left(1, \frac{\pi(\mathbf{y}) p_\beta(\beta^*)}{\pi(\mathbf{x}^{(t)}) p_\beta(\beta)}\right).$$

Expansion/Contraction move



Example 1 - Mixture of two normals with unknown means

$$wN(\mu_1, \sigma^2) + (1 - w)N(\mu_2, \sigma^2)$$

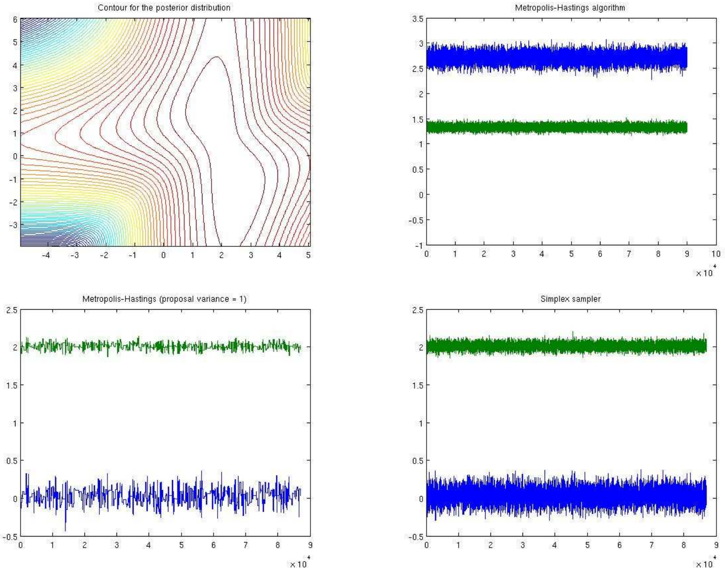
where μ_1, μ_2 are unknown and w, σ^2 are known.

Generate $1,000$ points from $0.2N(0,1) + 0.8N(2,1)$.

Prior Distribution: $N(1, 10)$.

Note: The total number of iterations for each algorithm was kept constant accounting for the differing number for iterations per sweep for each algorithms.

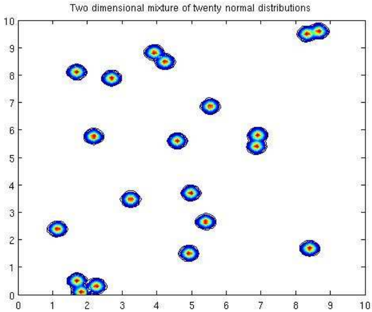
Results


Contour for the posterior distribution
Metropolis-Hastings algorithm
Metropolis-Hastings (proposal variance = 1)
Simplex sampler
$\times 10^4$
$\times 10^4$
$\times 10^4$

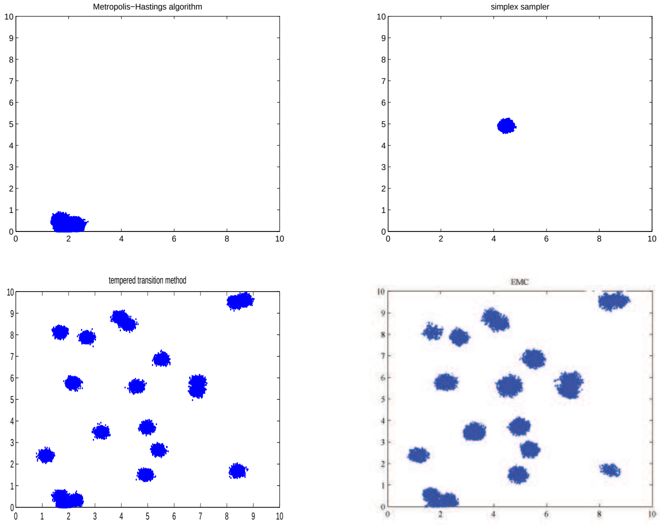
Mixture of **20** bivariate normals

$$f(\mathbf{x}) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{i=1}^{20} w_i \exp\left(-\frac{1}{2\sigma^2}(\mathbf{x}-\mu_i)'(\mathbf{x}-\mu_i)\right)$$

where $\sigma = 0.1$, $w_1 = \ldots = w_{20} = 0.05$ and $\mu_1, \mu_2, \ldots, \mu_{20}$ are known.



Results

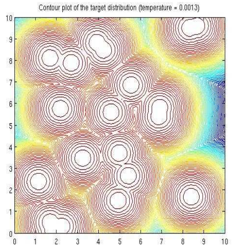

Metropolis−Hastings algorithm
simplex sampler
tempered transition method
EMC

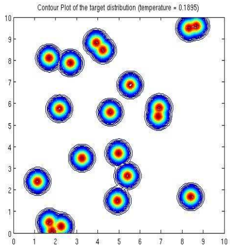
Conclusions

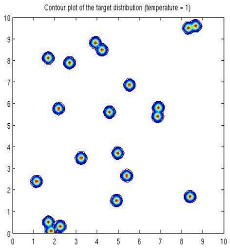
- High correlation $\longrightarrow$ Good mixing under the modes.
- Multi-modal target $\longrightarrow$ It gets trapped under a mode.

Nelder and Mead simplex method is a deterministic optimization algorithm...

Tempering ladder







Tempered simplex sampler

$\pi(\cdot)$ is a d dimensional distribution of interest.

$\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$: the current population.

$\mathbf{x}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{im})$ with $\mathbf{x}_{ij} \in \mathbb{R}^d$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$.

The temperature ladder is $T = (T_1, \ldots, T_n)$ with $T_1 = 0 \leq T_2 \leq \ldots \leq T_n = 1$.

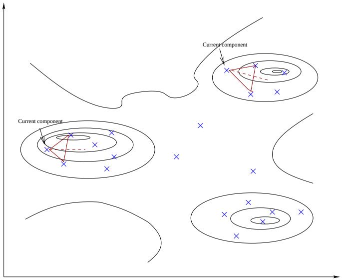
$$\pi(\mathbf{X}) = \prod_{i=1}^{n} \pi^{T_i}(\mathbf{x}_i)$$

and

$$\pi^{T_i}(\mathbf{x}_i) = \prod_{j=1}^{m} \pi^{T_i}(\mathbf{x}_{ij}).$$

Tempered simplex sampler - Step 1

- Update the simplex within each temperature, i.e $\mathbf{x}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{im})$ under temperature T_i.
  (i) Apply a reflection move.
  (ii) Apply an expansion/contraction move.
  (iii) Apply a Metropolis-Hastings move.



Tempered simplex sampler - Step 2

Current population $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_l, \ldots, \mathbf{x}_n)$ under $T = (T_1, \ldots, T_n)$ with $T_1 = 0 \le T_2 \le \ldots \le T_n = 1$.

- Exchange the population $\mathbf{x}_i$ under temperature T_i with another population $\mathbf{x}_l$ under temperature T_l.
  - (i) Set $\mathbf{y}_i = \mathbf{x}_l$ under T_i and $\mathbf{y}_l = \mathbf{x}_i$ under T_l.
  - (ii) Accept the proposed move with probability
$$\min\left\{1, \left(\frac{\pi(\mathbf{x}_i)}{\pi(\mathbf{x}_l)}\right)^{T_l - T_i}\right\}.$$

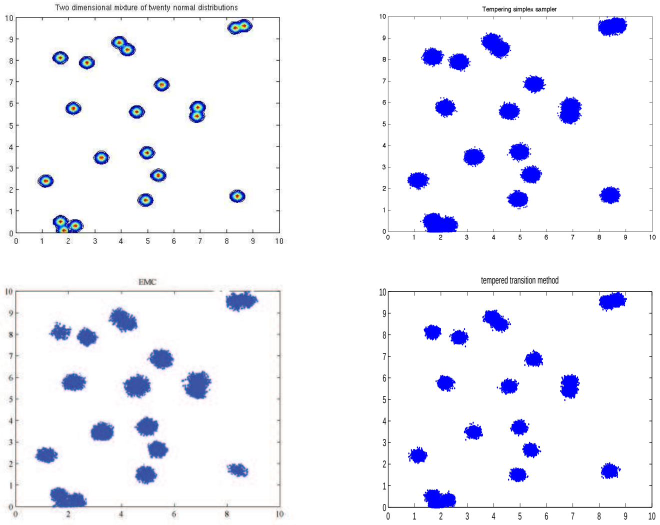
Then the new population is $(\mathbf{x}_1, \ldots, \mathbf{x}_l, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_n)$.

The tempered simplex sampler

The tempered simplex sampler borrows

- the good mixing of the simplex sampler in case of high correlated target distributions.
- the good mixing of the tempering methods that are able to traverse the entire sampling space without being trapped under a single maximum.

Example - Mixture of twenty normal distributions



Results - Mixture of twenty normal distributions

Modes	A	B	C	D
target	5%	15%	5%	10%
tempered simplex sampler	4.81%	14.18%	5.03%	10.57 %
EMC	1.17%	7.82%	5.33%	8.24%
tempered transitions	2.48%	16.48%	6.88%	3.68%

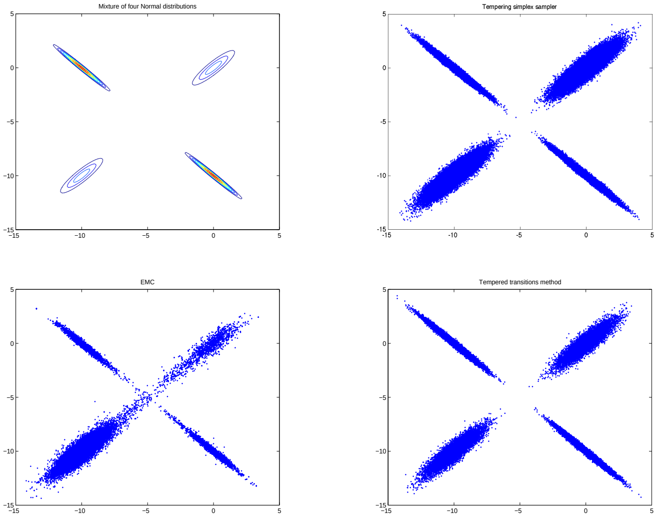
Table: Percentage of observations under different modes using the different methods.

Results - Mixture of twenty normal distributions

Methods	IA	ESS
tempered simplex sampler	60	17
EMC	90	5
tempered transitions	67	6

Table: Integrated autocorrelation and efficient sample size for the mixture of twenty normal distributions using each of the different methods.

Results - Mixture of four bivariate Normal distributions



Results - Mixture of four bivariate Normal distributions

Modes	A	B	C	D
target	25%	25%	25%	25%
tempered simplex sampler	26.68%	22.72%	26.30%	24.31 %
EMC	28.73%	51.34%	14.41%	5.52%
tempered transitions	32.67%	20.44%	24.36%	22.54%

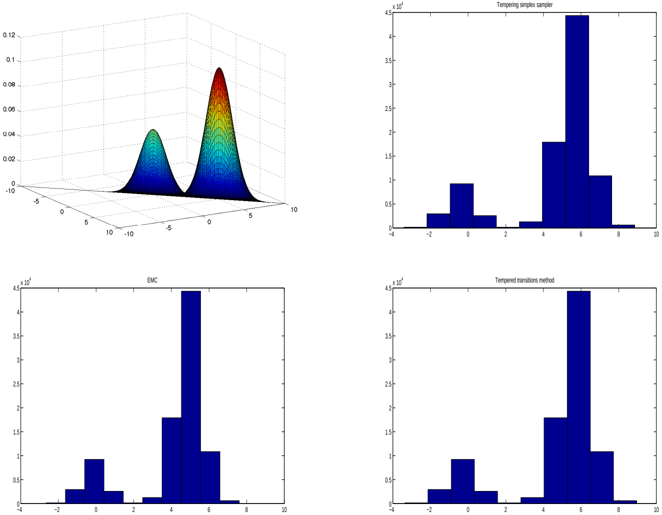
Table: Percentage of observations under each mode for the mixture of four Normal distributions using the different methods.

Results - Mixture of four bivariate Normal distributions

Methods	IA	ESS
tempered simplex sampler	41	24
EMC	66	6
tempered transitions	54	7

Table: Integrated autocorrelation and efficient sample size for the mixture of four normal distributions using each of the different methods.

Results - Mixture of two ten-dimensional normal distributions. Marginal histogram of the sample.



Results - Mixture of two ten-dimensional normal distributions

Methods	IA	ESS
tempered simplex sampler	4106	0.24
EMC	5002	0.09
tempered transitions	5459	0.07

Table: Integrated autocorrelation and efficient sample size for the mixture of the two bimodal ten-dimensional Normal distributions using each of the different methods.

Conclusions

- Samples efficiently from multi-modal target distributions.
- Mixes well under the modes.