# Generalized Ensemble Sampling Methods

Scott C. Schmidler

Stat 863: Advanced Statistical Computing
Duke University
Fall 2018

## Generalized ensemble methods

Sampling a standard Boltzmann distn

$$\pi(x) = \frac{1}{Z(\beta)} e^{-\beta H(x)} \qquad \beta = \frac{1}{T}$$

can view as assigning each state $x$ weight $w(\beta, x) = e^{-\beta H(x)}$

Consider modifying $\tilde{\pi}(x) \propto \tilde{w}(x)\pi(x)$.

If $\tilde{x}$ smoother, may be easier to explore. (e.g. tempering)



## Generalized ensemble methods

One idea: try to choose $w(x)$ s.t. $U = H(x)$ approx uniformly distributed.

I.e. if $x \sim \pi$, the marginal

$$\pi_U(u) = \frac{1}{Z(\beta)} \Omega(u) e^{-\beta u}$$

where $\Omega(u)$ is the *density of states* (spectral density)

If we can sample

$$x \sim \tilde{\pi}(x) \propto e^{-S(H(x))}$$

where $S(u) = \log(\Omega(u))$, then

$$\tilde{\pi}_U \propto c$$



## Generalized ensemble methods

If we can sample

$$x \sim \tilde{\pi}(x) \propto e^{-S(H(x))}$$

where $S(u) = \log(\Omega(u))$, then

$$\tilde{\pi}_U \propto c$$

- Multi-canonical sampling
- Wang-Landau algorithm

iteratively uptdate approximation $\hat{\Omega}(u)$ to spectral density $\Omega(u)$ to achieve this.

Originally developed for finite state spaces (e.g. spin systems).



## Generalized Wang-Landau (Atchade & Liu, 2009)

Partition state space $\mathcal{X} = \mathcal{X}_0 \cup \ldots \cup \mathcal{X}_k$ according to predefined energy levels $-\infty \le e_0 < e_1 < \cdots < e_k \le \infty$.

Goal: Sample from $\tilde{\pi}(x) = \sum_{i=1}^{k} \frac{\pi(x)}{\pi(\mathcal{X}_i)} \mathbf{1}_{\mathcal{X}_i}(x)$ *uniform energy*

**Algorithm:** Adaptively estimate $\hat{\pi}_n(i) \approx \pi(\mathcal{X}_i)$ by SA:
$\{\gamma_n\}$ a sequence of decreasing positive numbers.
Initialize $\phi_0(i) > 0$ for $i = 1, \ldots, k$, and $\hat{\pi}_0(i) = \frac{\phi_0(i)}{\sum_j \phi_0(j)}$

(i) Sample $X_{n+1} \sim \sum_{i=1}^{k} \frac{\pi(x)}{\hat{\pi}_n(i)} \mathbf{1}_{\mathcal{X}_i}(x)$ by MH.

(ii) Set $\phi_{n+1}(i) = \phi_n(i)\left(1 + \gamma_{a_n} \mathbf{1}_{\{X_{n+1} \in \mathcal{X}_i\}}\right)$; $\hat{\pi}_{n+1}(i) = \frac{\phi_{n+1}(i)}{\sum_j \phi_{n+1}(j)}$.

(iii) If $\max_i \left| v_{\kappa,n+1}(i) - \frac{1}{k} \right| \le \frac{c}{k}$ where $v_{\kappa,n}(i) = \frac{1}{n-\kappa} \sum_{j=\kappa+1}^{n} \mathbf{1}_{\{X_j \in \mathcal{X}_i\}}$
then set $\kappa = n+1$ and $a_{n+1} = a_n + 1$, otherwise $a_{n+1} = a_n$.



## Improving on (generalized) Wang-Landau

Performance of the WL algorithm depends heavily on a good choice of the energy rings $E_0, \ldots, E_k$: number, spacing, max.

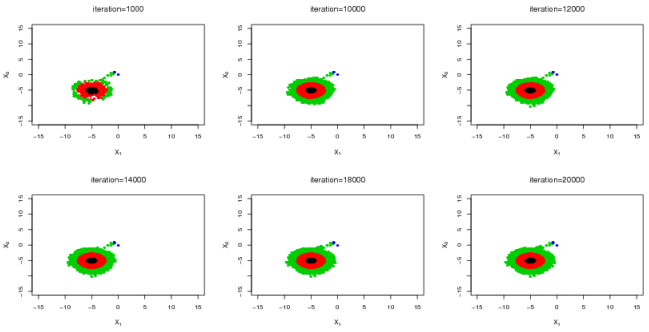
- *Adaptive-energy* GWL algorithm (AE-GWL), Wang & Schmidler (2011).

Monte-Carlo integration converges very slowly for WL

- Importance-resampling solution, Wang & Schmidler (2011).

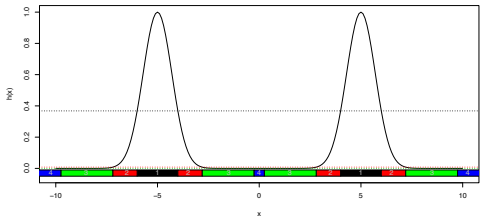
## Example

Figure: Example 2, modes at (-5,-5) and (5,5)






## Slow mixing of generalized Wang-Landau




(b) $d = 4$, fixed energy levels

Theorem (SW11b): GWL slowly mixing for geometric energy-levels.



## Energy level adaptation scheme

Performance of the WL algorithm depends heavily on a good choice of the energy rings $E_0, \ldots, E_k$.

We introduce an adaptive scheme to make updating energy levels fully automatic:

1. Initialize by a geometric progression:

   $e_0 = \inf_x E(x) = 0,\ e_1 = 1,\ e_2 = r_e, \ldots, E_{k-1} = r_e^{k-2}, E_k = infty.$
2. Every $n_{\text{split}}$ iterations: if any $|\log(\phi_i) - \log(\phi_{i+1})| > E$, divide the $i$-th energy ring by adding a new $e^*_{i+1} = e_i \times \sqrt{\frac{e_{i+1}}{e_i}}$, again using geometric progression. Set $\log(\phi^*_{i+1}) = 0$.
3. Also update the second largest $e_i$;

$$E^*_{k-1} = \frac{E^2_{k-1}}{E_k}$$

Set $\log(\phi^*_k) = 0$.



## Adaptive Energy Generalized Wang-Landau (AE-GWL)

**Algorithm:** Adaptively estimate $\hat\pi_n(i) \approx \pi(\mathcal{X}_i)$ by SA:
$\{\gamma_n\}$ a sequence of decreasing positive numbers.
Initialize $\phi_0(i) > 0$ for $i = 1, \ldots, k$, and $\hat\pi_0(i) = \frac{\phi_0(i)}{\sum_j \phi_0(j)}$

(i) Sample $X_{n+1} \sim \sum_{i=1}^k \frac{\pi(x)}{\hat\pi_n(i)} \mathbf{1}_{\mathcal{X}_i}(x)$ by MH.

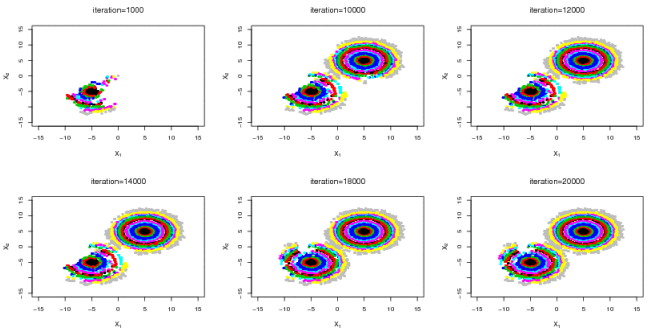
(ii) Set $\phi_{n+1}(i) = \phi_n(i)\left(1 + \gamma_{a_n} \mathbf{1}_{\{X_{n+1} \in \mathcal{X}_i\}}\right)$ and $\hat\pi_{n+1}(i) = \frac{\phi_{n+1}(i)}{\sum_j \phi_{n+1}(j)}$.

(iii) If $\max_i \left|v_{\kappa,n+1}(i) - \frac{1}{k}\right| \le \frac{c}{k}$ where $v_{\kappa,n}(i) = \frac{1}{n-\kappa} \sum_{j=\kappa+1}^{n} \mathbf{1}_{\{X_j \in \mathcal{X}_i\}}$ then set $\kappa = n+1$ and $a_{n+1} = a_n + 1$, otherwise $a_{n+1} = a_n$.

**(iv)* For every $n_{\text{split}}$ iterations, adaptively update $E = \{E_i\}$.**



## Example




(c) $d = 4$, update internal energy levels



## $\frac{1}{k}$-ensemble method

- Tries to make entropy variable uniformly distributed.
- Estimates $k(H(x))$ the number states with *smaller or equal* energy

## Generalized ensemble sampling

Note that all these methods *modify the target distn* $\pi(x)$ to speed up sampling.

How to recover samples/integrals of interest?

Reweighting: importance sampling estimators.

But may be highly variable...

Can we use a similar idea to speed up sampling *of* $\pi$ *itself*?



## Equi-Energy Sampler [Kou et al., 2006]

Constructs $I$ processes $X^{(i)}$ with tempered target densities $\pi^{(i)} \propto \pi^{\beta_i}$ for inverse temperatures $1 = \beta_1 > \ldots > \beta_I \geq 0$.

For each $i$, bin sample history $(X^{(i)}_{0:n})$ according to energy.

Process $X^{(i)}$ occasionally proposes to move to a state previously visited by $X^{(i+1)}$ lying in same energy bin.

These "equi-energy" moves can be non-local in the state space, potentially enabling transitions between distinct modes of $\pi$.