Lecture Notes for STAT546: Computational Statistics

—Lecture 8: Monte Carlo

Faming Liang

Purdue University

September 11, 2024

Dynamic Weighting

The MH algorithm has a stringent requirement for the detailed balance condition. To move across an energy barrier, the expected waiting time is roughly exponential in the energy difference. Hence, the algorithm suffers from a *waiting time dilemma*: either to wait forever in a deep local energy minimum or to have an incorrect equilibrium distribution, in simulations from a complex system for which the energy landscape is rugged.

Wong and Liang (1997) proposed a way out of the waiting time dilemma, which can be described loosely as follows: If necessary, the system may make a transition against a steep probability barrier without a proportionally long waiting time. To account for the bias introduced thereby, an importance weight is computed and recorded along with the sampled values. This transition rule does not satisfy the detailed balance condition any more, but it satisfies what is called *invariance with respect to importance weights* (IWIW). At equilibrium, Monte Carlo approximations to integrals are obtained by the importance-weighted average of the sampled values, rather than the simple average as in the Metropolis-Hastings algorithm.

IWUW Principle

In dynamic weighting, the state of the Markov chain is augmented by an importance weight to (x, w), where the weight w carries the information of the past samples and can help the system escape from local-traps. Let (x_t, w_t) denote the current state of the Markov chain, a dynamic weighting transition involves the following steps:

1. Draw y from a proposal function $T(x_t, y)$.
2. Compute the dynamic ratio

$$r_d = w_t \frac{f(y)T(y, x_t)}{f(x_t)T(x_t, y)}.$$

3. Let θ_t be a non-negative number, which can be set as a function of (x_t, w_t). With probability $a = r_d/(\theta_t + r_d)$, set $x_{t+1} = y$ and $w_{t+1} = r_d/a$; otherwise set $x_{t+1} = x_t$ and $w_{t+1} = w_t/(1 - a)$.

This transition is called the R-type move in Wong and Liang (1997). It does not satisfy the detailed balance condition, but is *invariant with respect to the importance weight* (IWIW); that is, if

$$\int w_t g(x_t, w_t) dw_t \propto f(x_t) \tag{1}$$

holds for some constant c, then after one step of transition,

$$\int w_{t+1} g(x_{t+1}, w_{t+1}) dw_{t+1} \propto f(x_{t+1}) \tag{2}$$

also holds, where $g(x, w)$ denotes the joint density of (x, w).

Let $x = x_t$, $w = w_t$, $x' = x_{t+1}$ and $w' = w_{t+1}$. Then

$$\int_0^\infty w' g(x', w') dw'$$
$$= \int_{\mathcal{X}} \int_0^\infty [\theta_t + r_d(x, x', w)] g(x, w) T(x, x') \frac{r_d(x, x', w)}{\theta_t + r_d(x, x', w)} dwdx$$
$$+ \int_{\mathcal{X}} \int_0^\infty w[\theta_t + r_d(x', z, w)] g(x', w) T(x', z) \frac{\theta_t}{\theta_t + r_d(x', z, w)} dwdz$$
$$= \int_{\mathcal{X}} \int_0^\infty wg(x, w) \frac{f(x') T(x', x)}{f(x)} dwdx + \int_{\mathcal{X}} \int_0^\infty wg(x', w) T(x', z) dwdz$$
$$\propto f(x') \int_{\mathcal{X}} T(x', x) dx + f(x')$$
$$= 2f(x').$$

Hence, given a sequence of dynamic weighting samples (x_1, w_1), (x_2, w_2), ..., (x_n, w_n), the weighted average [of a state function $h(x)$ over the sample]

$$\hat{\mu} = \sum_{i=1}^{n} w_i h(x_i) / \sum_{i=1}^{n} w_i \tag{3}$$

The merit of dynamic weighting is as follows: If one trial is rejected, then the dynamic weight will be self-adjusted to a larger value by dividing the rejection probability of that trial, rendering a smaller total rejection probability in the next trial. Using importance weights provides a means for dynamic weighting to make transitions that are not allowed by the standard MH rule, and thus can traverse the energy landscape of the system more freely. But this advantage comes with a price: The importance weights have an infinite expectation, and the estimate (3) is of high variability and converges to the true values very slowly, seemingly at a rate of $\log(n)$ (Liu, Liang and Wong, 2001). In short, the infinite waiting time in the standard MH process now manifests itself as an infinite weight quantity in the dynamic weighting process.

It is interesting to point out that the usual MH transition can be regarded as a special type of IWIW transition: If we apply a MH transition to x and leave w unchanged, then the result satisfies IWIW. This can be shown as follows:

$$
\begin{aligned}
&\int w' g(x', w') dw' = \int w g(x, w) K(x \to x') dw dx \\
&\propto \int f(x) K(x \to x') dx = \int f(x') K(x' \to x) dx \\
&= f(x'),
\end{aligned}
$$

where $K(\cdot \to \cdot)$ denotes a MH transition kernel with $f(x)$ as its invariant distribution. Therefore, correctly weighted distributions will remain when dynamic weighting transitions and MH transitions are alternated in the same run of the Markov chain. This observation leads directly to the tempering dynamic weighting algorithm (Liang and Wong, 1999).

The tempering dynamic weighting (TDW) algorithm (Liang and Wong, 1999) is essentially the same as the simulated tempering algorithm except that a dynamic weight is now associated with the configuration (x, i) and the dynamic weighting rule is used to guide the transitions between adjacent temperature levels. Let $f_i(x)$ denote the trial distribution at level i, $i = 1, \ldots, N$. Let $0 < \alpha < 1$ be specified in advance and let (x_t, i_t, w_t) denote the current state of the Markov chain. One iteration of the TDW algorithm consists of the following steps:

Tempering Dynamic Weighting Algorithm

1. Draw U from the uniform distribution $U[0,1]$.
2. If $U \le \alpha$, set $i_{t+1} = i_t$ and $w_{t+1} = w_t$ and simulate x_{t+1} from $f_{i_t}(x)$ via one or several MH updates.
3. If $U > \alpha$, set $x_{t+1} = x_t$ and propose a level transition, $i_t \to i'$, from a transition function $q(i_t, i')$. Conduct a dynamic weighting transition to update (i_t, w_t):
   - Compute the dynamic weighting ratio

     $$r_d = w_t \frac{c_i f_{i'}(x_t) q(i', i_t)}{c_{i'} f_i(x_t) q(i_t, i')},$$

     where c_i denotes the pseudo-normalizing constant of $f_i(x)$.
   - Accept the transition with probability $a = r_d/(\theta_t + r_d)$, where θ_t can be chosen as a function of (i_t, w_t). If it is accepted, set $i_{t+1} = i'$ and $w_{t+1} = r_d/a$; otherwise, set $i_{t+1} = i_t$ and $w_{t+1} = w_t/(1-a)$.

Ising Model Simulation at Sub-Critical Temperature

Consider a 2-D Ising model with the Boltzmann density

$$f(\boldsymbol{x}) = \frac{1}{Z(K)} \exp\{K \sum_{i \sim j} x_i x_j\}, \tag{4}$$

where the spins $x_i = \pm 1$, $i \sim j$ denotes the nearest neighbors on the lattice, $Z(K)$ is the partition function, and K is the inverse temperature. When the temperature is at or below the critical point (K=0.4407), the model has two oppositely magnetized states separated by a very steep energy barrier.

Liang and Wong (1999) performed TDW simulations on the lattices of size 32^2, 64^2 and 128^2 using 6, 11, and 21 temperature levels (with the values of K being equally spaced between 0.4 and 0.5), respectively. At the same temperature level, the Gibbs sampler (Geman and Geman, 1984) is used to generate new configurations, meanwhile the weights were left unchanged. The dynamic weighting rule is only used to govern transitions between levels. After each sweep of Gibbs updates, it is randomly proposed to move to an adjacent temperature level with equal probability. The parameter θ_t is set to 1 if $w_t < 10^6$ and 0 otherwise. Five independent runs were performed for the model. In each run, the simulation continues until 10,000 configurations are obtained at the final temperature level. For the model of size 128^2, the average number of sweeps in each run is 776,547.

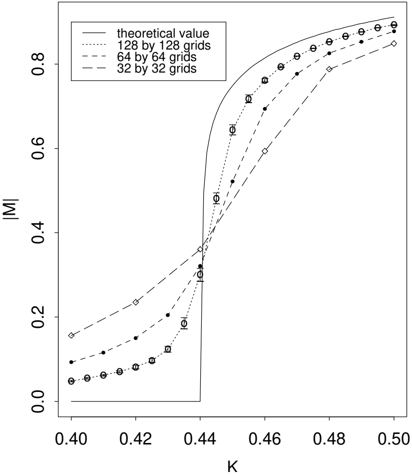

Figure 1: Expectation of the absolute spontaneous magnetization against the inverse temperature K for the lattices of size 32^2, 64^2 and 128^2. The points are averages over 5 independent runs. For clarity of picture, error bars are plotted only for the lattice of size 128^2. The smooth curve corresponds to the theoretical result of infinite lattice. (Liang and Wong, 1999)

Dynamic Weighting in Optimization

We use the traveling salesman problem (TSP) (Reinelt, 1994) to illustrate the use of dynamic weighting in optimization. Let n be the number of cities in an area and let d_{ij} denote the distance between city i and city j. The TSP is to find a permutation x of the cities such that the tour length

$$H(x) = \sum_{i=1}^{n-1} d_{x(i),x(i+1)} + d_{x(n),x(1)}, \tag{5}$$

is minimized. It is known that the TSP is a NP-complete problem.

Let V denote the set of cities, let A denote the set of cities that have been ordered, and let $A^c = V \setminus A$ denote the set of cities not yet ordered. Then the cities can be ordered as follows:

- Randomly select a city from V.
- Repeat steps (1) and (2) until A^c is empty:
  (1) set $k = \arg\max_{i \in A^c} \min_{j \in A} d_{ij}$;
  (2) set $A = A \cup \{k\}$ and $A^c = A^c \setminus \{k\}$.

This procedure ensures that each time the city added into A is the one having the maximum separation from the set of already ordered cities.

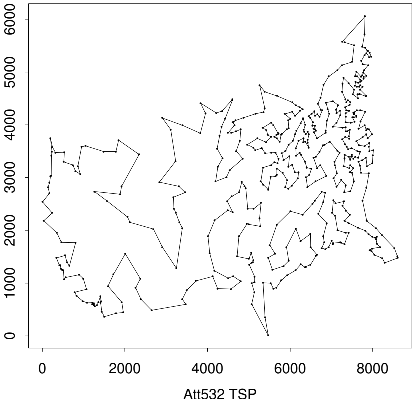

Figure 2: The best tour found by dynamic weighting for a 532-city TSP. The tour length is 27744 compared to the exact minimum of 27686. (Wong and Liang, 1997)

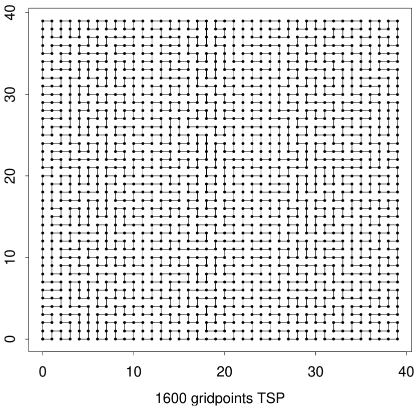

Figure 3: One of the exactly optimal tour found by dynamic weighting for a 1600-city TSP. (Wong and Liang, 1997)

dynamically weighted importance sampling (Liang, 2002)

Because the importance weight in dynamic weighting is of high variability, achieving a stable estimate requires techniques such as stratification and truncation (Liu, Liang and Wong, 2001). However, any attempt to shorten the tails of the weight distribution may lead to a biased estimate. To overcome this difficulty, Liang (2002b) proposed a population version of dynamic weighting, the so-called the dynamically weighted importance sampling (DWIS) algorithm, which has the importance weight successfully controlled to a desired range while keeping the estimate unbiased. In DWIS, the state space of the Markov chain is augmented to a population of size N, denoted by $(\boldsymbol{x}, \boldsymbol{w}) = \{x_1, w_1; \ldots; x_N, w_N\}$. With a slight abuse of notation, (x_i, w_i) is called an individual state of the population. Given the current population $(\boldsymbol{x}_t, \boldsymbol{w}_t)$, one iteration of DWIS (as illustrated by Figure 4) involves two steps:

1. *Dynamic weighting*: Each individual state of the current population is updated via one dynamic weighting transition step to form a new population.
2. *Population control*: Duplicate the samples with large weights and discard the samples with small weights. The bias induced thereby is counterbalanced by giving different weights to the new samples produced.

The two steps ensure that DWIS can move across energy barriers like dynamic weighting, but the weights are well controlled and have a finite expectation, and the resulting estimate can converge much faster than that of dynamic weighting.

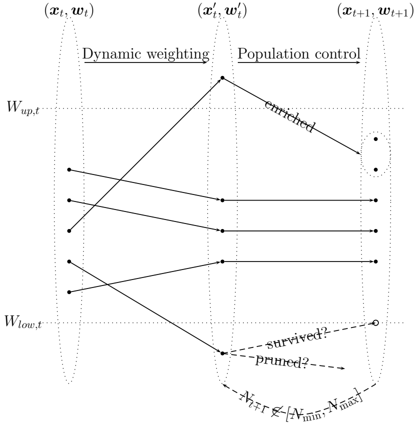

Figure 4: A diagram of the DWIS algorithm.

A Theory of DWIS

Let $g_t(x, w)$ denote the joint density of (x, w), an individual state of $(\boldsymbol{x}_t, \boldsymbol{w}_t)$, and let $f(x)$ denote the target distribution we are working with.

Definition 1

The distribution $g_t(x, w)$ defined on $\mathcal{X} \times (0, \infty)$ is called correctly weighted with respect to $f(x)$ if the following conditions hold,

$$\int w g_t(x, w) dw = c_{tx} f(x), \tag{6}$$

$$\frac{\int_{\mathcal{A}} c_{tx} f(x) dx}{\int_{\mathcal{X}} c_{tx} f(x) dx} = \int_{\mathcal{A}} f(x) dx, \tag{7}$$

where $\mathcal{A}$ is any Borel set, $\mathcal{A} \subseteq \mathcal{X}$.

Definition 2

If $g_t(x, w)$ is correctly weighted with respect to $f(x)$, and samples $(x_{t,i}, w_{t,i})$ are simulated from $g_t(x, w)$ for $i = 1, 2, \ldots, n_t$, then $(\boldsymbol{x}_t, \boldsymbol{w}_t) = (x_{t,1}, w_{t,1}; \cdots ; x_{t,n_t}, w_{t,n_t})$ is called a correctly weighted population with respect to $f(x)$.

Let $(\boldsymbol{x}_t, \boldsymbol{w}_t)$ be a correctly weighted population with respect to $f(x)$, and let $y_1, \ldots, y_m$ be distinct states in $\boldsymbol{x}_t$. Generate a random variable/vector Y such that

$$P\{Y = y_i\} = \frac{\sum_{j=1}^{n_t} w_j I(y_i = x_{t,j})}{\sum_{j=1}^{n_t} w_j}, \quad i = 1, 2, \ldots, m, \qquad (8)$$

where $I(\cdot)$ is an indicator function. Then Y is approximately distributed as $f(\cdot)$ if n_t is large. This can be summarized as the following theorem:

Theorem 3

The distribution of the random vector Y generated in (8) converges as $n_t \to \infty$ to $f(\cdot)$, if the importance weight has a finite expectation.

Let $(\boldsymbol{x}_1, \boldsymbol{w}_1)$, ..., $(\boldsymbol{x}_N, \boldsymbol{w}_N)$ be a series of correctly weighted populations generated by a DWIS algorithm with respect to $f(x)$. Then the quantity $\mu = E_f h(x)$, provided its existence, can be estimated by

$$\widehat{\mu} = \frac{\sum_{t=1}^{N} \sum_{i=1}^{n_t} w_{t,i} h(x_{t,i})}{\sum_{t=1}^{N} \sum_{i=1}^{n_t} w_{t,i}}. \tag{9}$$

Let $U_t = \sum_{i=1}^{n_t} w_{t,i} h(x_{t,i})$, $S_t = \sum_{i=1}^{n_t} w_{t,i}$, $S = ES_t$, and $V_t = U_t - \mu S_t$. If the variance of U_t and V_t are both finite, then the standard error of $\widehat{\mu}$ can be calculated using the ratio estimate as in finite population sampling (Ripley, 1987, pp.158). As shown by Liang (2002b), $\widehat{\mu}$ is consistent and asymptotically normally distributed; that is,

$$\sqrt{N}(\widehat{\mu} - \mu) \to N(0, \sigma^2), \tag{10}$$

where σ^2 is defined as $\mathrm{Var}(V_t)/S^2$.

Definition 4

A transition rule for a population $(\boldsymbol{x}, \boldsymbol{w})$ is said to be *invariant with respect to the importance weight* (IWIW_p) if the joint density of $(\boldsymbol{x}, \boldsymbol{w})$ remains correctly weighted whenever the initial joint density is correctly weighted.

To make this rule distinguishable from the IWIW rule given in (1) and (2), which is defined for the single Markov chain, we denote it by IWIW_p with the subscript representing for population.

Some IWIW_p Transition Rules

Many transition rules are IWIW_p exactly or approximately. The following are some useful examples. Let $(\boldsymbol{x}_t, \boldsymbol{w}_t)$ denote the current population, let (x, w) denote an individual state of $(\boldsymbol{x}_t, \boldsymbol{w}_t)$, and let (x', w') denote the individual state transmitted from (x, w) in one transition step.

A general dynamic weighting sampler

1. Draw y from some transition proposal distribution $T(x, y)$, and compute the dynamic ratio

$$r_d = w \frac{f(y)T(y,x)}{f(x)T(x,y)}.$$

2. Choose $\theta_t = \theta(\boldsymbol{x}_t, \boldsymbol{w}_t) \geq 0$ and draw $U \sim unif(0, 1)$. Update (x, w) as (x', w')

$$(x', w') = \begin{cases} (y, (1+\delta_t) r_d / a) & \text{if } U \leq a \\ (x, (1+\delta_t) w / (1-a)) & \text{otherwise.} \end{cases}$$

where $a = r_d/(r_d + \theta_t)$; and θ_t and δ_t are both functions of $(\boldsymbol{x}_t, \boldsymbol{w}_t)$, but they remain constants for each individual state of the same population.

Adaptive pruned-enriched population control scheme

Given the current population $(\boldsymbol{\theta}_t, \boldsymbol{w}_t)$, adjusting the values of the importance weights and the population size to suitable ranges involves the following scheme. Let $(x_{t,i}, w_{t,i})$ be the ith individual state of the population, let n_t and n'_t denote the current and new population sizes, let $W_{low,t}$ and $W_{up,t}$ denote the lower and upper weight control bounds, let $n_{\min}$ and $n_{\max}$ denote the minimum and maximum population size allowed by the user, and let n_{low} and n_{up} denote the lower and upper reference bounds of the population size.

1. (Initialization) Initialize the parameters $W_{low,t}$ and $W_{up,t}$ by

$$W_{low,t} = \sum_{i=1}^{n_t} w_{t,i}/n_{up}, \quad W_{up,t} = \sum_{i=1}^{n_t} w_{t,i}/n_{low}.$$

   Set $n'_t = 0$ and $\lambda > 1$. Do steps 2–4 for $i = 1, 2, \cdots, n_t$.

2. (Pruned) If $w_{t,i} < W_{low,t}$, prune the state with probability $q = 1 - w_{t,i}/W_{low,t}$. If it is pruned, drop $(x_{t,i}, w_{t,i})$ from $(\boldsymbol{x}_t, \boldsymbol{w}_t)$; otherwise, update $(x_{t,i}, w_{t,i})$ as $(x_{t,i}, W_{low,t})$ and set $n'_t = n'_t + 1$.

3. (Enriched) If $w_{t,i} > W_{up,t}$, set $d = [w_{t,i}/W_{up,t} + 1]$, $w'_{t,i} = w_{t,i}/d$, replace $(x_{t,i}, w_{t,i})$ by d identical states $(x_{t,i}, w'_{t,i})$, and set $n'_t = n'_t + d$, where $[z]$ denotes the integer part of z.

4. (Unchanged) If $W_{low,t} \leq w_{t,i} \leq W_{up,t}$, keep $(x_{t,i}, w_{t,i})$ unchanged, and set $n'_t = n'_t + 1$.

5. (Checking) If $n'_t > n_{\max}$, set $W_{low,t} \leftarrow \lambda W_{low,t}$, $W_{up,t} \leftarrow \lambda W_{up,t}$ and $n'_t = 0$, do step 2–4 again for $i = 1, 2, \cdots, n_t$. If $n'_t < n_{\min}$, set $W_{low,t} \leftarrow W_{low,t}/\lambda$, $W_{up,t} \leftarrow W_{up,t}/\lambda$ and $n'_t = 0$, do step 2–4 again for $i = 1, 2, \cdots, n_t$. Otherwise, stop.

Dynamically Weighted Importance Sampling (scheme-R)

- *Dynamic weighting*: Apply the R-type move to the population $(\boldsymbol{x}_{t-1}, \boldsymbol{w}_{t-1})$. If $W_{up,t-1} \leq W_c$, then set $\theta_t = 1$. Otherwise, set $\theta_t = 0$. The new population is denoted by $(\boldsymbol{x}'_t, \boldsymbol{w}'_t)$.
- *Population Control*: Apply APEPCS to $(\boldsymbol{x}'_t, \boldsymbol{w}'_t)$. The new population is denoted by $(\boldsymbol{x}_t, \boldsymbol{w}_t)$.

Dynamically Weighted Importance Sampling (scheme-W)

- *Preweight control*: If $n_{t-1} < n_{low}$, then set $W_{low,t} = W_{low,t-1}/\lambda$ and $W_{up,t} = W_{up,t-1}/\lambda$. If $n_{t-1} > n_{up}$, then set $W_{low,t} = \lambda W_{low,t-1}$ and $W_{up,t} = \lambda W_{up,t-1}$. Otherwise, set $W_{low,t} = W_{low,t-1}$ and $W_{up,t} = W_{up,t-1}$.
- *Dynamic weighting*: Apply the W-type move to the population $(\boldsymbol{x}_{t-1}, \boldsymbol{w}_{t-1})$ with $\delta_t = 1/(\alpha + \beta W_{up,t}^{1+\epsilon})$ for some $\epsilon > 0$. The new population is denoted by $(\boldsymbol{x}'_t, \boldsymbol{w}'_t)$.
- *Population Control*: Apply APEPCS to $(\boldsymbol{x}'_t, \boldsymbol{w}'_t)$. The new population is denoted by $(\boldsymbol{x}_t, \boldsymbol{w}_t)$.

Weight Behavior Analysis

To analyze the weight behavior of the DWIS, we first introduce the following lemma.

Lemma 5

Let $\pi(x_0)$ denote the marginal equilibrium distribution under transition T and let $\pi(x_0, x_1) = \pi(x_0) \times T(x_0, x_1)$ and $r(x_0, x_1) = \pi(x_1, x_0)/\pi(x_0, x_1)$ be the MH ratio. Then

$$e_0 = E_\pi \log r(x_0, x_1) \leq 0,$$

where the equality holds when it induces a reversible Markov chain, that is, $\pi(x_0, x_1) = \pi(x_1, x_0)$.

For simplicity, let (x_t, w_t) denote an individual state of the population $(\boldsymbol{x}_t, \boldsymbol{w}_t)$. When $\delta_t \equiv 0$ and $\theta_t \equiv 0$, the weights of the W-type move and the R-type move evolve as

$$\log w_t = \log w_{t-1} + \log r(x_{t-1}, x_t),$$

which results in

$$\log w_t = \log w_0 + \sum_{s=1}^{t} \log r(x_{s-1}, x_s). \tag{11}$$

Following from Lemma 5 and the ergodicity theorem (under stationarity),

$$\frac{1}{t} \sum_{s=1}^{t} \log r(x_{s-1}, x_s) \rightarrow e_0 < 0, \quad a.s. \tag{12}$$

as $t \rightarrow \infty$. Hence, w_t will go to 0 almost surely as $t \rightarrow \infty$.

Scheme-R

When $\theta_t \equiv 1$, the expectation of w_t, conditional on x_{t-1}, x_t and w_{t-1}, can be calculated as

$$\begin{aligned} E[w_t|x_{t-1}, x_t, w_{t-1}] &= (r_d + 1)\frac{r_d}{r_d + 1} + w_{t-1}(r_d + 1)\frac{1}{r_d + 1} \\ &= r_d + w_{t-1} \\ &= w_{t-1}[1 + r(x_{t-1}, x_t)]. \end{aligned} \tag{13}$$

Since $r(x_{t-1}, x_t) \geq 0$, the weight process $\{w_t\}$ is driven by an inflation drift. Hence, w_t will go to infinity almost surely as t becomes large.

To prevent the weight process from going to 0 or ∞, scheme-R alters the use of $\theta_t = 0$ and $\theta_t = 1$. When $W_{up,t} > W_c$, θ_t is set to 0, so the weight process of scheme-R can be bounded above by

$$\log w_t = \log w_0 + \sum_{s=1}^{t} \log r(x_{s-1}, x_s) - \sum_{s=1}^{t} \log(d_s),$$

provided that $\theta_1 = \cdots = \theta_t = 0$, where d_s is a positive integer as defined in the APEPCS.

Moments of DWIS weights

The weight process of the two DWIS schemes can be characterized by the following process:

$$Z_t = \begin{cases} Z_{t-1} + \log r(x_{t-1}, x_t) - \log(d_t), & \text{if } Z_{t-1} > 0 \ , \\ 0, & \text{if } Z_{t-1} < 0 \ , \end{cases} \tag{14}$$

where there exists a constant C such that $|Z_t - Z_{t-1}| \leq C$ almost surely. Let $T_0 = 0$, $T_i = \min\{t : t > T_{i-1}, Z_t = 0\}$, and $L_i = T_i - T_{i-1}$ for $i \geq 1$. From (12) and the fact that $d_t \geq 1$, it is easy to see that L_i is almost surely finite; that is, there exists an integer M such that $P(L_i < M) = 1$. This implies that for any fixed $\eta > 0$,

$$E \exp(\eta Z_t) \leq \exp(\eta MC), \quad a.s.$$

This leads to the following theorem:

Theorem 6

*Under the assumption (**??**), the importance weight in DWIS almost surely has finite moments of any order.*

Suppose we are interested in simulating from a distribution $f \propto (1, 1000, 1, 2000)$ with the transition matrix

$$T = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{2}{3} & 0 & \frac{1}{3} & 0 \\ 0 & \frac{4}{7} & 0 & \frac{3}{7} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

This example characterizes the problem of simulation from a multimodal distribution.

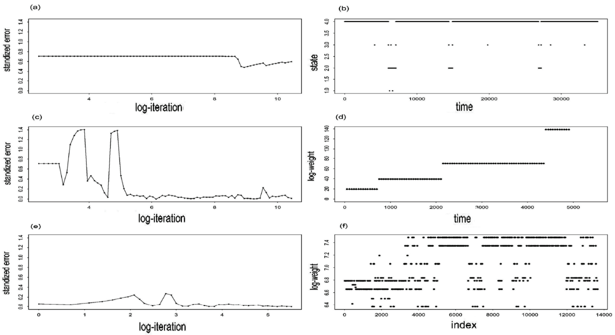

Figure 5: Comparison of MH, dynamic weighting (DW) and DWIS for the 4-state example. (a) Time plot of standardized errors of $\hat{f}_{MH}$ against log-iterations. (b) Time plot of state transitions of the MH run. (c) Time plot of standardized errors of $\hat{f}_{DW}$ against log-iterations. (d) Time plot of the log-weights collected at state 4 in the run of dynamic weighting. (e) Time plot of standardized errors of $\hat{f}_{DWIS}$ against log-iterations. (f) Time plot of the log-weights collected at state 4 in the run of DWIS. (Liang, 2002b)