

Details for proofs of corollaries 1–3

Suzie Brown

February 4, 2020

Note: time is labelled in reverse throughout this document.

For any $N \in \mathbb{N}$, we define the conditioning set to be $\mathcal{H}_t^N = (\mathbf{X}_t, \mathbf{X}_{t-1}, \mathbf{w}_t, \mathbf{w}_{t-1})$. (The dependence on N is suppressed from the notation in the following so that I wouldn't have to update it everywhere.) This set works as a separatrix between \mathbf{a}_t and \mathcal{F}_{t-1} as desired. It is different from the one used in (Koskela et al 2019) because we include \mathbf{w}_t directly rather than its explicit expression in terms of $\mathbf{a}_{t+1}, \mathbf{X}_{t+1}, \mathbf{X}_t$ to simplify the notation.

In each case, the choice of parent depends on two factors: the conditional probability of choosing that parent under the given resampling scheme, and the conditional “probability” of a particle moving from that parent's position at time t to the offspring's position at time $t - 1$. We see the contribution of these two factors in the following.

Definition 1. For any N and any $i \in \{1, \dots, N\}$, a function $f : \{1, \dots, N\}^N \rightarrow \mathbb{R}$ is said to be i -increasing if it is an increasing function in $v_t^{(i)} = |\{j : a_t^{(j)} = i\}|$.

Multinomial resampling

Under multinomial resampling, the parental indices \mathbf{a}_t are conditionally independent given \mathcal{H}_t . The conditional law of each index is

$$\mathbb{P}[a_t^{(i)} = a_i \mid \mathcal{H}_t] \propto w_t^{(i)} q_{t-1}(X_t^{(a_i)}, X_{t-1}^{(i)}).$$

Since the indices are all independent, their joint conditional law is

$$\mathbb{P}[\mathbf{a}_t = \mathbf{a} \mid \mathcal{H}_t] \propto \prod_{i=1}^N w_t^{(i)} q_{t-1}(X_t^{(a_i)}, X_{t-1}^{(i)}).$$

Lemma 1. Let $a_t^{(i)}$ be the parental indices from a SMC algorithm with multinomial resampling. For any function f that is i -increasing,

$$\begin{aligned} \mathbb{E}[f(\mathbf{a}_t) \mid \mathcal{H}_t] &\leq \mathbb{E}[f(\mathbf{A}_1)] \\ \mathbb{E}[f(\mathbf{a}_t) \mid \mathcal{H}_t] &\geq \mathbb{E}[f(\mathbf{A}_2)] \end{aligned}$$

where the elements of $\mathbf{A}_1, \mathbf{A}_2$ are all mutually independent and independent of \mathcal{F}_∞ , and distributed according to

$$\begin{aligned} A_1^{(j)} &\sim \text{Categorical} \left(\left(\frac{a}{\varepsilon} \right)^{\mathbb{1}_{\{i=1\}} - \mathbb{1}_{\{i \neq 1\}}}, \dots, \left(\frac{a}{\varepsilon} \right)^{\mathbb{1}_{\{i=N\}} - \mathbb{1}_{\{i \neq N\}}} \right) \\ A_2^{(j)} &\sim \text{Categorical} \left(\left(\frac{\varepsilon}{a} \right)^{\mathbb{1}_{\{i=1\}} - \mathbb{1}_{\{i \neq 1\}}}, \dots, \left(\frac{\varepsilon}{a} \right)^{\mathbb{1}_{\{i=N\}} - \mathbb{1}_{\{i \neq N\}}} \right), \end{aligned}$$

where the arguments of Categorical and Multinomial distributions are given up to a normalising constant here and throughout this document.

Proof. The result follows using the bounds given in equations (1), (2) with a balls-in-bins coupling, and cancelling h from the top and bottom. \square

Corollary 1. *Under the time scaling (??), supposing there exist constants $0 < \varepsilon \leq 1 \leq a < \infty$ such that*

$$\frac{1}{a} \leq g_t(x, x') \leq a \quad (1)$$

$$\varepsilon h(x') \leq q_t(x, x') \leq \frac{1}{\varepsilon} h(x'), \quad (2)$$

genealogies of SMC algorithms with multinomial resampling converge to Kingman's n -coalescent in the sense of finite-dimensional distributions as $N \rightarrow \infty$.

Proof. Define the “family sizes” $V_1^{(i)} := |\{j : A_1^{(j)} = i\}|$ and $V_2^{(i)} := |\{j : A_2^{(j)} = i\}|$ for $i = 1, \dots, N$. The distributions of $\mathbf{A}_1, \mathbf{A}_2$ imply the following:

$$\begin{aligned} \mathbf{V}_1 &\sim \text{Multinomial} \left(N, \left(\frac{a}{\varepsilon} \right)^{\mathbb{1}_{\{i=1\}} - \mathbb{1}_{\{i \neq 1\}}}, \dots, \left(\frac{a}{\varepsilon} \right)^{\mathbb{1}_{\{i=N\}} - \mathbb{1}_{\{i \neq N\}}} \right) \\ \mathbf{V}_2 &\sim \text{Multinomial} \left(N, \left(\frac{\varepsilon}{a} \right)^{\mathbb{1}_{\{i=1\}} - \mathbb{1}_{\{i \neq 1\}}}, \dots, \left(\frac{\varepsilon}{a} \right)^{\mathbb{1}_{\{i=N\}} - \mathbb{1}_{\{i \neq N\}}} \right), \end{aligned}$$

Notice that the function $f_i(\mathbf{a}_t) := (v_t^{(i)})_2$ is i -increasing for each $i = 1, \dots, N$. Applying Lemma 1 and the Multinomial moments formula (Mosimann, 1962), we obtain the following lower bound:

$$\begin{aligned} \mathbb{E}_t[f_i(\mathbf{a}_t)] &\geq \mathbb{E}[f_i(\mathbf{A}_2)] = \mathbb{E}[(V_2^{(i)})_2] \\ &= \frac{(N)_2 (\varepsilon/a)^2}{[(\varepsilon/a) + (N-1)(a/\varepsilon)]^2} \geq \frac{(N)_2 (\varepsilon/a)^2}{N^2 (a/\varepsilon)^2} = \frac{(N)_2}{N^2} \frac{\varepsilon^4}{a^4}. \end{aligned}$$

So we can lower bound the denominator by

$$\mathbb{E}_t[c_N(t)] = \frac{1}{(N)_2} \sum_{i=1}^N \mathbb{E}_t[(v_t^{(i)})_2] \geq \frac{N}{(N)_2} \frac{(N)_2}{N^2} \frac{\varepsilon^4}{a^4} = \frac{\varepsilon^4}{Na^4}.$$

To upper bound the numerator, consider the function $f_i(\mathbf{a}_t) := (v_t^{(i)})_3$, which is i -increasing for each $i = 1, \dots, N$. Again using Lemma 1 and (Mosimann, 1962), we obtain the following upper bound:

$$\begin{aligned} \mathbb{E}_t[f_i(\mathbf{a}_t)] &\leq \mathbb{E}[f_i(\mathbf{A}_1)] = \mathbb{E}[(V_1^{(i)})_3] \\ &= \frac{(N)_3 (a/\varepsilon)^3}{[(a/\varepsilon) + (N-1)(\varepsilon/a)]^3} \leq \frac{(N)_3 (a/\varepsilon)^3}{N^3 (\varepsilon/a)^3} = \frac{(N)_3}{N^3} \frac{a^6}{\varepsilon^6}. \end{aligned}$$

and the numerator is therefore bounded above by

$$\frac{1}{(N)_3} \sum_{i=1}^N \mathbb{E}_t[(v_t^{(i)})_3] \leq \frac{N}{(N)_3} \frac{(N)_3}{N^3} \frac{\varepsilon^6}{a^6} = \frac{\varepsilon^6}{N^2 a^6}.$$

The ratio is therefore bounded above by

$$\frac{\frac{1}{(N)_3} \sum_{i=1}^N \mathbb{E}_t[(v_t^{(i)})_3]}{\frac{1}{(N)_2} \sum_{i=1}^N \mathbb{E}_t[(v_t^{(i)})_2]} \leq \frac{N}{(N)_3} \frac{(N)_3}{N^3} \frac{\varepsilon^6}{a^6} = \frac{\varepsilon^6}{N^2 a^6} \frac{Na^4}{\varepsilon^4} = \frac{a^{10}}{N \varepsilon^{10}} =: b_N \xrightarrow{N \rightarrow \infty} 0.$$

We can thus conclude the proof of Corollary 1 by applying Theorem 1. \square

Conditional SMC with multinomial resampling

We can apply the same technique to tackle conditional SMC, but it requires an adjustment of the bounding distributions. We assume wlog that the immortal particle always takes index 1. The parental indices are still independent, but we now have to treat $i = 1$ as a special case. We have the following conditional law:

$$\begin{aligned}\mathbb{P}[a_t^{(1)} = a_1 \mid \mathcal{H}_t] &= \mathbb{I}\{a_1 = 1\} \\ \mathbb{P}[a_t^{(i)} = a_i \mid \mathcal{H}_t] &\propto w_t^{(i)} q_{t-1}(X_t^{(a_i)}, X_{t-1}^{(i)}) \quad i = 2, \dots, N.\end{aligned}$$

The joint conditional law is therefore

$$\mathbb{P}[\mathbf{a}_t = \mathbf{a} \mid \mathcal{H}_t] \propto \mathbb{I}\{a_1 = 1\} \prod_{i=2}^N w_t^{(i)} q_{t-1}(X_t^{(a_i)}, X_{t-1}^{(i)}).$$

Lemma 2. *Let $a_t^{(i)}$ be the parental indices from a conditional SMC algorithm with multinomial resampling, such that (1) and (2) are satisfied. For any function f that is i -increasing,*

$$\begin{aligned}\mathbb{E}[f(\mathbf{a}_t) \mid \mathcal{H}_t] &\leq \mathbb{E}[f(\mathbf{A}_1)] \\ \mathbb{E}[f(\mathbf{a}_t) \mid \mathcal{H}_t] &\geq \mathbb{E}[f(\mathbf{A}_2)]\end{aligned}$$

where the elements of $\mathbf{A}_1, \mathbf{A}_2$ are all mutually independent and independent of \mathcal{F}_∞ , and distributed according to

$$\begin{aligned}A_1^{(j)} &\sim \begin{cases} \delta_1 & j = 1 \\ \text{Categorical}\left(\left(\frac{a}{\varepsilon}\right)^{\mathbb{1}_{\{i=1\}} - \mathbb{1}_{\{i \neq 1\}}}, \dots, \left(\frac{a}{\varepsilon}\right)^{\mathbb{1}_{\{i=N\}} - \mathbb{1}_{\{i \neq N\}}}\right) & j \neq 1 \end{cases} \\ A_2^{(j)} &\sim \begin{cases} \delta_1 & j = 1 \\ \text{Categorical}\left(\left(\frac{\varepsilon}{a}\right)^{\mathbb{1}_{\{i=1\}} - \mathbb{1}_{\{i \neq 1\}}}, \dots, \left(\frac{\varepsilon}{a}\right)^{\mathbb{1}_{\{i=N\}} - \mathbb{1}_{\{i \neq N\}}}\right) & j \neq 1. \end{cases}\end{aligned}$$

Corollary 2. *Under the time scaling (??) and conditions (1) and (2), genealogies of conditional SMC algorithms with multinomial resampling converge to Kingman's n -coalescent in the sense of finite-dimensional distributions as $N \rightarrow \infty$.*

As before, we can define “family sizes” $V_1^{(i)} := |\{j : A_1^{(j)} = i\}|$ and $V_2^{(i)} := |\{j : A_2^{(j)} = i\}|$ for $i = 1, \dots, N$. They now have the following distributions:

$$\begin{aligned}\mathbf{V}_1 &\stackrel{d}{=} (1, 0, \dots, 0) + \text{Multinomial}\left(N - 1, \left(\frac{a}{\varepsilon}\right)^{\mathbb{1}_{\{i=1\}} - \mathbb{1}_{\{i \neq 1\}}}, \dots, \left(\frac{a}{\varepsilon}\right)^{\mathbb{1}_{\{i=N\}} - \mathbb{1}_{\{i \neq N\}}}\right) \\ \mathbf{V}_2 &\stackrel{d}{=} (1, 0, \dots, 0) + \text{Multinomial}\left(N - 1, \left(\frac{\varepsilon}{a}\right)^{\mathbb{1}_{\{i=1\}} - \mathbb{1}_{\{i \neq 1\}}}, \dots, \left(\frac{\varepsilon}{a}\right)^{\mathbb{1}_{\{i=N\}} - \mathbb{1}_{\{i \neq N\}}}\right).\end{aligned}$$

Now consider again the i -increasing function $f_i(\mathbf{a}_t) := (v_t^{(i)})_2$. In the conditional SMC case, we can apply Lemma 2 to obtain the lower bound

$$\begin{aligned}\frac{1}{(N)_2} \sum_{i=1}^N \mathbb{E}_t[(v_t^{(i)})_2] &\geq \frac{1}{(N)_2} \sum_{i=1}^N \mathbb{E}[(V_2^{(i)})_2] = \frac{1}{(N)_2} \left[\mathbb{E}[(V_2^{(1)})_2] + \sum_{i=2}^N \mathbb{E}[(V_2^{(i)})_2] \right] \\ &= \frac{1}{(N)_2} \left[\frac{(N-1)_2(\varepsilon/a)^2}{[(\varepsilon/a) + (N-2)(a/\varepsilon)]^2} + 2 \frac{(N-1)(\varepsilon/a)}{(\varepsilon/a) + (N-2)(a/\varepsilon)} + \sum_{i=2}^N \frac{(N-1)_2(\varepsilon/a)^2}{[(\varepsilon/a) + (N-2)(a/\varepsilon)]^2} \right] \\ &= \frac{1}{(N)_2} \left[2 \frac{(N-1)(\varepsilon/a)}{(\varepsilon/a) + (N-2)(a/\varepsilon)} + \sum_{i=1}^N \frac{(N-1)_2(\varepsilon/a)^2}{[(\varepsilon/a) + (N-2)(a/\varepsilon)]^2} \right]\end{aligned}$$

using the Multinomial moments as before, along with the identity $(X+1)_2 \equiv 2(X)_1 + (X)_2$. This is further bounded by

$$\frac{1}{(N)_2} \sum_{i=1}^N \mathbb{E}_t[(v_t^{(i)})_2] \geq \frac{1}{(N)_2} \left[\frac{2(N-1)(\varepsilon/a)}{(N-1)(a/\varepsilon)} + \frac{(N)_3(\varepsilon/a)^2}{(N-1)^2(a/\varepsilon)^2} \right] = \frac{1}{(N)_2} \left[\frac{2\varepsilon^2}{a^2} + \frac{(N)_3}{(N-1)^2} \frac{\varepsilon^4}{a^4} \right]$$

Similarly, we derive an upper bound on $f_i(\mathbf{a}_t) := (v_t^{(i)})_3$, this time applying the identity $(X+1)_3 \equiv 3(X)_2 + (X)_3$:

$$\begin{aligned} \frac{1}{(N)_3} \sum_{i=1}^N \mathbb{E}_t[(v_t^{(i)})_3] &\leq \frac{1}{(N)_3} \left[\mathbb{E}[(V_2^{(1)})_3] + \sum_{i=2}^N \mathbb{E}[(V_2^{(i)})_3] \right] \\ &\leq \frac{1}{(N)_3} \left[3 \frac{(N-1)_2 (a/\varepsilon)^2}{[(a/\varepsilon) + (N-2)(\varepsilon/a)]^2} + \sum_{i=1}^N \frac{(N-1)_3 (a/\varepsilon)^3}{[(a/\varepsilon) + (N-2)(\varepsilon/a)]^3} \right] \\ &\leq \frac{1}{(N)_3} \left[\frac{3(N-1)_2 (a/\varepsilon)^2}{(N-1)^2 (\varepsilon/a)^2} + \frac{(N)_4 (a/\varepsilon)^3}{(N-1)^3 (\varepsilon/a)^3} \right] \\ &= \frac{1}{(N)_3} \left[\frac{3(N-1)_2}{(N-1)^2} \frac{a^4}{\varepsilon^4} + \frac{(N)_4}{(N-1)^3} \frac{a^6}{\varepsilon^6} \right]. \end{aligned}$$

We therefore have the following upper bound on the ratio:

$$\begin{aligned} \frac{\frac{1}{(N)_3} \sum_{i=1}^N \mathbb{E}_t[(v_t^{(i)})_3]}{\frac{1}{(N)_2} \sum_{i=1}^N \mathbb{E}_t[(v_t^{(i)})_2]} &\leq \frac{(N)_2}{(N)_3} \frac{\frac{(N)_4}{(N-1)^3} \frac{a^6}{\varepsilon^6} + \frac{3(N-1)_2}{(N-1)^2} \frac{a^4}{\varepsilon^4}}{\frac{(N)_3}{(N-1)^2} \frac{\varepsilon^4}{a^4} + \frac{2\varepsilon^2}{a^2}} \leq \frac{1}{N-2} \left[\frac{\frac{(N)_4}{(N-1)^3} \frac{a^6}{\varepsilon^6} + \frac{3(N-1)_2}{(N-1)^2} \frac{a^4}{\varepsilon^4}}{\frac{(N)_3}{(N-1)^2} \frac{\varepsilon^4}{a^4} + \frac{2\varepsilon^2}{a^2}} \right] \\ &= \frac{1}{N-2} \left[\frac{N-3}{N-1} \frac{a^{10}}{\varepsilon^{10}} + \frac{3}{2} \frac{N-2}{N-1} \frac{a^6}{\varepsilon^6} \right] \leq \frac{1}{N-2} \left[\frac{a^{10}}{\varepsilon^{10}} + \frac{3}{2} \frac{a^6}{\varepsilon^6} \right] =: b_N \rightarrow 0. \end{aligned}$$

Stochastic rounding case

If we resample using a stochastic rounding, we lose the independence between parental indices. The set of valid assignments is much smaller, because each family size can vary by no more than one from its expected value.

Defining the family sizes $v_t^{(i)} := |\{j : a_t^{(j)} = i\}|$ as functions of \mathbf{a}_t , we have the constraint $v_t^{(i)} \in \{\lfloor Nw_t^{(i)} \rfloor, \lfloor Nw_t^{(i)} \rfloor + 1\}$. This is encoded in the law of \mathbf{a}_t via indicator functions. We can find the probability of each of the two possible values up to a constant:

$$\begin{aligned} \mathbb{P}[v_t^{(i)} = \lfloor Nw_t^{(i)} \rfloor \mid \mathcal{H}_t] &= C \cdot (1 - Nw_t^{(i)} + \lfloor Nw_t^{(i)} \rfloor) \sum_{\substack{J \subset [N]: \\ |J| = \lfloor Nw_t^{(i)} \rfloor}} \left(\frac{N}{\lfloor Nw_t^{(i)} \rfloor} \right)^{-1} \prod_{j \in J} q_{t-1}(X_t^{(i)}, X_{t-1}^{(j)}) \\ \mathbb{P}[v_t^{(i)} = \lfloor Nw_t^{(i)} \rfloor + 1 \mid \mathcal{H}_t] &= C \cdot (Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor) \sum_{\substack{J \subset [N]: \\ |J| = \lfloor Nw_t^{(i)} \rfloor + 1}} \left(\frac{N}{\lfloor Nw_t^{(i)} \rfloor + 1} \right)^{-1} \prod_{j \in J} q_{t-1}(X_t^{(i)}, X_{t-1}^{(j)}) \end{aligned}$$

We assume uniform boundedness of the transition densities, which is stronger than the condition used in the other corollaries. That is, there exists $\varepsilon \in (0, 1)$ such that, almost surely,

$$\varepsilon \leq q_{t-1}(x, x') \leq \frac{1}{\varepsilon}. \quad (3)$$

We can use this to bound the probabilities above and below:

$$\begin{aligned} C \cdot (1 - Nw_t^{(i)} + \lfloor Nw_t^{(i)} \rfloor) \varepsilon^{\lfloor Nw_t^{(i)} \rfloor} &\leq \mathbb{P}[v_t^{(i)} = \lfloor Nw_t^{(i)} \rfloor \mid \mathcal{H}_t] \leq C \cdot (1 - Nw_t^{(i)} + \lfloor Nw_t^{(i)} \rfloor) \varepsilon^{-\lfloor Nw_t^{(i)} \rfloor} \\ C \cdot (Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor) \varepsilon^{(\lfloor Nw_t^{(i)} \rfloor + 1)} &\leq \mathbb{P}[v_t^{(i)} = \lfloor Nw_t^{(i)} \rfloor + 1 \mid \mathcal{H}_t] \leq C \cdot (Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor) \varepsilon^{-(\lfloor Nw_t^{(i)} \rfloor + 1)} \end{aligned}$$

As these are the only two possibilities, we can easily find bounds on the normalised probabilities. We use that, for $A, B > 0$,

$$\frac{A}{A+B} = \frac{1}{1+(B/A)} \leq \frac{1}{1+(B^-/A^+)} = \frac{A^+}{A^+ + B^-} \quad (4)$$

where A^+ is an upper bound on A and B^- is a lower bound on B . This yields slightly tighter bounds than just bounding the numerator A above and the denominator $A + B$ below. A similar result is used for the lower bound.

For the first upper bound we get:

$$\begin{aligned}\mathbb{P}[v_t^{(i)} = \lfloor Nw_t^{(i)} \rfloor \mid \mathcal{H}_t] &\leq \frac{(1 - Nw_t^{(i)} + \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{-\lfloor Nw_t^{(i)} \rfloor}}{(1 - Nw_t^{(i)} + \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{-\lfloor Nw_t^{(i)} \rfloor} + (Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{(\lfloor Nw_t^{(i)} \rfloor + 1)}} \\ &= (1 - Nw_t^{(i)} + \lfloor Nw_t^{(i)} \rfloor) \left[1 - (1 - \varepsilon^{(2\lfloor Nw_t^{(i)} \rfloor + 1)})(Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor) \right]^{-1} \\ &\leq (1 - Nw_t^{(i)} + \lfloor Nw_t^{(i)} \rfloor) \left[1 - (1 - \varepsilon^{(2\lfloor Nw_t^{(i)} \rfloor + 1)}) \right]^{-1} \\ &= (1 - Nw_t^{(i)} + \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{-(2\lfloor Nw_t^{(i)} \rfloor + 1)}\end{aligned}$$

where we use that $x - \lfloor x \rfloor \in [0, 1)$. A similar calculation gives us the corresponding lower bound:

$$\begin{aligned}\mathbb{P}[v_t^{(i)} = \lfloor Nw_t^{(i)} \rfloor \mid \mathcal{H}_t] &\geq \frac{(1 - Nw_t^{(i)} + \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{\lfloor Nw_t^{(i)} \rfloor}}{(1 - Nw_t^{(i)} + \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{\lfloor Nw_t^{(i)} \rfloor} + (Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{-(\lfloor Nw_t^{(i)} \rfloor + 1)}} \\ &= (1 - Nw_t^{(i)} + \lfloor Nw_t^{(i)} \rfloor) \left[1 + (\varepsilon^{-(2\lfloor Nw_t^{(i)} \rfloor + 1)} - 1)(Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor) \right]^{-1} \\ &\geq (1 - Nw_t^{(i)} + \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{(2\lfloor Nw_t^{(i)} \rfloor + 1)}.\end{aligned}$$

Also the upper bound on the other probability:

$$\begin{aligned}\mathbb{P}[v_t^{(i)} = \lfloor Nw_t^{(i)} \rfloor + 1 \mid \mathcal{H}_t] &\leq \frac{(Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{-(\lfloor Nw_t^{(i)} \rfloor + 1)}}{(Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{-(\lfloor Nw_t^{(i)} \rfloor + 1)} + (1 - Nw_t^{(i)} + \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{\lfloor Nw_t^{(i)} \rfloor}} \\ &= (Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor) \left[\varepsilon^{(2\lfloor Nw_t^{(i)} \rfloor + 1)} + (1 - \varepsilon^{(2\lfloor Nw_t^{(i)} \rfloor + 1)})(Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor) \right]^{-1} \\ &= (Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{-(2\lfloor Nw_t^{(i)} \rfloor + 1)},\end{aligned}$$

and the corresponding lower bound:

$$\begin{aligned}\mathbb{P}[v_t^{(i)} = \lfloor Nw_t^{(i)} \rfloor + 1 \mid \mathcal{H}_t] &\geq \frac{(Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{(\lfloor Nw_t^{(i)} \rfloor + 1)}}{(Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{(\lfloor Nw_t^{(i)} \rfloor + 1)} + (1 - Nw_t^{(i)} + \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{-\lfloor Nw_t^{(i)} \rfloor}} \\ &= (Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor) \left[\varepsilon^{-(2\lfloor Nw_t^{(i)} \rfloor + 1)} - (\varepsilon^{-(2\lfloor Nw_t^{(i)} \rfloor + 1)} - 1)(Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor) \right]^{-1} \\ &= (Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{(2\lfloor Nw_t^{(i)} \rfloor + 1)}.\end{aligned}$$

In summary, we have:

$$\begin{aligned}(1 - Nw_t^{(i)} + \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{(2\lfloor Nw_t^{(i)} \rfloor + 1)} &\leq \mathbb{P}[v_t^{(i)} = \lfloor Nw_t^{(i)} \rfloor \mid \mathcal{H}_t] \leq (1 - Nw_t^{(i)} + \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{-(2\lfloor Nw_t^{(i)} \rfloor + 1)} \\ (Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{(2\lfloor Nw_t^{(i)} \rfloor + 1)} &\leq \mathbb{P}[v_t^{(i)} = \lfloor Nw_t^{(i)} \rfloor + 1 \mid \mathcal{H}_t] \leq (Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{-(2\lfloor Nw_t^{(i)} \rfloor + 1)}\end{aligned}$$

Now we can get at the expectations of interest:

$$\begin{aligned}\mathbb{E}[(v_t^{(i)})_2 \mid \mathcal{H}_t] &= (\lfloor Nw_t^{(i)} \rfloor)_2 \mathbb{P}[v_t^{(i)} = \lfloor Nw_t^{(i)} \rfloor \mid \mathcal{H}_t] + (\lfloor Nw_t^{(i)} \rfloor + 1)_2 \mathbb{P}[v_t^{(i)} = \lfloor Nw_t^{(i)} \rfloor + 1 \mid \mathcal{H}_t] \\ &\geq (\lfloor Nw_t^{(i)} \rfloor)_2 (1 - Nw_t^{(i)} + \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{(2\lfloor Nw_t^{(i)} \rfloor + 1)} + (\lfloor Nw_t^{(i)} \rfloor + 1)_2 (Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{(2\lfloor Nw_t^{(i)} \rfloor + 1)} \\ &= \lfloor Nw_t^{(i)} \rfloor (2Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor - 1) \varepsilon^{(2\lfloor Nw_t^{(i)} \rfloor + 1)}\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[(v_t^{(i)})_3 \mid \mathcal{H}_t] &= (\lfloor Nw_t^{(i)} \rfloor)_3 \mathbb{P}[v_t^{(i)} = \lfloor Nw_t^{(i)} \rfloor \mid \mathcal{H}_t] + (\lfloor Nw_t^{(i)} \rfloor + 1)_3 \mathbb{P}[v_t^{(i)} = \lfloor Nw_t^{(i)} \rfloor + 1 \mid \mathcal{H}_t] \\ &\leq (\lfloor Nw_t^{(i)} \rfloor)_3 (1 - Nw_t^{(i)} + \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{-(2\lfloor Nw_t^{(i)} \rfloor + 1)} + (\lfloor Nw_t^{(i)} \rfloor + 1)_3 (Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{-(2\lfloor Nw_t^{(i)} \rfloor + 1)} \\ &= (\lfloor Nw_t^{(i)} \rfloor - 1)\lfloor Nw_t^{(i)} \rfloor \varepsilon^{-(2\lfloor Nw_t^{(i)} \rfloor + 1)} (3Nw_t^{(i)} - 2\lfloor Nw_t^{(i)} \rfloor - 2) \\ &\leq (\lfloor Nw_t^{(i)} \rfloor - 1)\lfloor Nw_t^{(i)} \rfloor (2Nw_t^{(i)} - 1 - \lfloor Nw_t^{(i)} \rfloor) \varepsilon^{-(2\lfloor Nw_t^{(i)} \rfloor + 1)}.\end{aligned}$$

In particular, we now have

$$\mathbb{E}[(v_t^{(i)})_3 \mid \mathcal{H}_t] \leq (\lfloor Nw_t^{(i)} \rfloor - 1) \varepsilon^{-(4\lfloor Nw_t^{(i)} \rfloor + 2)} \mathbb{E}[(v_t^{(i)})_2 \mid \mathcal{H}_t].$$

Now we use the bounds on the potential functions to take care of the terms in $w_t^{(i)}$. Using the bounds (1) along with the form of the weights in Algorithm 1, we have almost surely for each i

$$\frac{1}{Na^2} \leq w_t^{(i)} \leq \frac{a^2}{N}.$$

Applying these bounds along with the simple inequality $\lfloor x \rfloor \leq x$, we have

$$\mathbb{E}[(v_t^{(i)})_3 \mid \mathcal{H}_t] \leq (a^2 - 1) \varepsilon^{-(4a^2 + 2)} \mathbb{E}[(v_t^{(i)})_2 \mid \mathcal{H}_t].$$

Finally, since this bound applies uniformly for each i , the ratio of interest is bounded by

$$\frac{\frac{1}{(N)_3} \sum_{i=1}^N \mathbb{E}_t[(v_t^{(i)})_3]}{\frac{1}{(N)_2} \sum_{i=1}^N \mathbb{E}_t[(v_t^{(i)})_2]} \leq \frac{1}{N-2} (a^2 - 1) \varepsilon^{-(4a^2 + 2)} =: b_N \rightarrow 0$$

as required, so we can apply Theorem 1 to obtain the result.

References

Mosimann, J. E. (1962), ‘On the compound multinomial distribution, the multivariate β -distribution, and correlations among proportions’, *Biometrika* **49**(1/2), 65–82.