Stratified resampling ++

Suzie Brown

21 May 2021

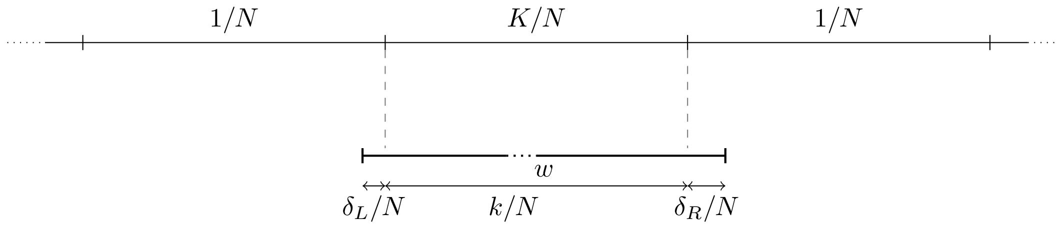
(A1) The conditional distribution of parental indices $a_t^{(1:N)}$ given offspring counts $\nu_t^{(1:N)}$ is uniform over all assignments such that $|\{j : a_t^{(j)} = i\}| = \nu_t^{(i)}$ for all i.

There are complex dependencies between the offspring counts, but we can still find some constraints on the distribution of each count conditional on the corresponding weight. Write the i^{th} weight in the form $w_t^{(i)} = (K + \delta)/N$, where $\delta \in [0, 1)$ and $K \in \{0, \dots, N-1\}$. The distribution of $\nu_t^{(i)}$ depends not only on $w_t^{(i)}$ but also on where the i^{th} weight interval falls with respect to the length-$(1/N)$ intervals for inversion sampling. There are two cases to consider, which are illustrated in Figure 1. Note that Case (b) cannot happen if $K = 0$.

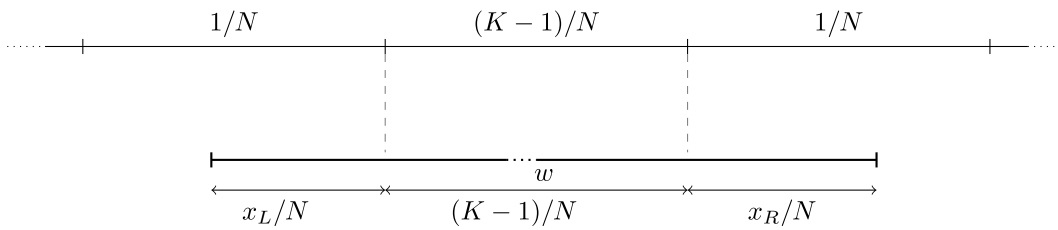
In any case $\nu_t^{(i)} \in \{k-1, k, k+1, k+2\}$ almost surely. To define a probability distribution over these four values, we introduce the notation $p_j := \mathbb{P}[\nu_t^{(i)} = \lfloor N w_t^{(i)} \rfloor + j \mid w_t^{(i)}]$, for $j = -1, 0, 1, 2$. Since the sample within each interval of length $1/N$ is uniform over that interval, we find the probabilities given in Table 1, in terms of δ and the other quantities $\delta_L, \delta_R \in [0, \delta]$ and $x_L, x_R \in [\delta, 1]$ defined in Figure 1. The probabilities do not depend on k, but of course the corresponding values of $\nu_t^{(i)}$ do. By definition $\delta_L + \delta_R = \delta$ and $x_L + x_R = 1 + \delta$.

	Case (a)	Case (b)	L.B.	U.B.
p_{-1}	0	$x_L x_R - \delta$	0	$1/4$
p_0	$1 - \delta + \delta_L \delta_R$	$1 + \delta - 2x_L x_R$	$(1-\delta)^2/2$	$1 - 3\delta/4$
p_1	$\delta - 2\delta_L \delta_R$	$x_L x_R$	$\delta/2$	$(1+\delta)/2$
p_2	$\delta_L \delta_R$	0	0	$1/4$

Table 1: Marginal probability distribution of $\nu_t^{(i)}$ conditional on $w_t^{(i)}$, in terms of δ and the quantities defined in Figure 1, along with upper and lower bounds on these in terms of δ only, which hold in both cases i.e. whenever $w_t^{(i)} = (K + \delta)/N$.



(a) The parent under consideration is automatically assigned K offspring, plus up to two more. ($\delta_L + \delta_R = \delta$.)



(b) This case can only occur when $K \geq 1$. The parent under consideration is automatically assigned $K-1$ offspring, plus up to two more. ($x_L + x_R = 1 + \delta$.)

Figure 1: Cases for stratified resampling with a fixed weight $w = (K+\delta)/N$

Theorem 1. *Let $\nu_t^{(1:N)}$ denote the offspring numbers in an interacting particle system satisfying (A1) such that, for any N sufficiently large, $\mathbb{P}[\tau_N(t) = \infty] = 0$ for all finite t. Suppose that there exists a deterministic sequence $(b_N)_{N\geq 1}$ such that $\lim_{N\to\infty} b_N = 0$ and*

$$\frac{1}{(N)_3} \sum_{i=1}^{N} \mathbb{E}_t[(\nu_t^{(i)})_3] \leq b_N \frac{1}{(N)_2} \sum_{i=1}^{N} \mathbb{E}_t[(\nu_t^{(i)})_2] \tag{1}$$

for all N, uniformly in $t \geq 1$. Fix $n \leq N$ and consider a randomly chosen sample of n terminal particles. Then the resulting rescaled genealogical process $(G_{\tau_N(t)}^{(n,N)})_{t\geq 0}$ converges in the sense of finite-dimensional distributions to Kingman's n-coalescent as $N \to \infty$.

Corollary 2. *Consider an SMC algorithm using stratified resampling, such that (A1) is satisfied. Assume that there exists a constant $a \in [1, \infty)$ such that for all x, x', t,*

$$\frac{1}{a} \leq g_t(x, x') \leq a. \tag{2}$$

Assume that $\mathbb{P}[\tau_N(t) = \infty] = 0$ for all finite t. Let $(G_t^{(n,N)})_{t \geq 0}$ denote the genealogy of a random sample of n terminal particles from the output of the algorithm when the total number of particles used is N. Then, for any fixed n, the time-scaled genealogy $(G_{\tau_N(t)}^{(n,N)})_{t \geq 0}$ converges to Kingman's n-coalescent as $N \to \infty$, in the sense of finite-dimensional distributions.

Proof. Recall that the sequence of σ-algebras

$$\mathcal{H}_t := \sigma(X_{t-1}^{(1:N)}, X_t^{(1:N)}, w_{t-1}^{(1:N)}, w_t^{(1:N)}) \tag{3}$$

are such that $\nu_t^{(1:N)}$ is conditionally independent of the filtration $\mathcal{F}_{t-1}$ given $\mathcal{H}_t$. With stratified resampling, conditional on the weights each offspring count almost surely takes one of four values: $\nu_t^{(i)} \in \{\lfloor N w_t^{(i)} \rfloor - 1, \lfloor N w_t^{(i)} \rfloor, \lfloor N w_t^{(i)} \rfloor + 1, \lfloor N w_t^{(i)} \rfloor + 2\}$. Denote $p_j^{(i)} := \mathbb{P}[\nu_t^{(i)} = \lfloor N w_t^{(i)} \rfloor + j \mid \mathcal{H}_t]$ for $j = -1, 0, 1, 2$. Now

$$\begin{aligned}\mathbb{E}[(\nu_t^{(i)})_2 \mid \mathcal{H}_t] = p_{-1}^{(i)}(\lfloor N w_t^{(i)} \rfloor - 1)_2 + p_0^{(i)}(\lfloor N w_t^{(i)} \rfloor)_2 + p_1^{(i)}(\lfloor N w_t^{(i)} \rfloor + 1)_2 \\ + p_2^{(i)}(\lfloor N w_t^{(i)} \rfloor + 2)_2\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[(\nu_t^{(i)})_3 \mid \mathcal{H}_t] &= p_{-1}^{(i)}(\lfloor N w_t^{(i)} \rfloor - 1)_3 + p_0^{(i)}(\lfloor N w_t^{(i)} \rfloor)_3 + p_1^{(i)}(\lfloor N w_t^{(i)} \rfloor + 1)_3 \\ &\quad + p_2^{(i)}(\lfloor N w_t^{(i)} \rfloor + 2)_3 \\ &= p_{-1}^{(i)}(\lfloor N w_t^{(i)} \rfloor - 3)(\lfloor N w_t^{(i)} \rfloor - 1)_2 + p_0^{(i)}(\lfloor N w_t^{(i)} \rfloor - 2)(\lfloor N w_t^{(i)} \rfloor)_2 \\ &\quad + p_1^{(i)}(\lfloor N w_t^{(i)} \rfloor - 1)(\lfloor N w_t^{(i)} \rfloor + 1)_2 + p_2^{(i)} \lfloor N w_t^{(i)} \rfloor (\lfloor N w_t^{(i)} \rfloor + 2)_2 \\ &\leq \lfloor N w_t^{(i)} \rfloor \{ p_0^{(i)}(\lfloor N w_t^{(i)} \rfloor)_2 + p_1^{(i)}(\lfloor N w_t^{(i)} \rfloor + 1)_2 \} \\ &= \lfloor N w_t^{(i)} \rfloor \mathbb{E}[(\nu_t^{(i)})_2 \mid \mathcal{H}_t] \\ &\leq a^2 \mathbb{E}[(\nu_t^{(i)})_2 \mid \mathcal{H}_t]\end{aligned}$$

The last line uses the almost sure bound $w_t^{(i)} \leq a^2/N$ which follows from (2) along with the form of the weights in Algorithm **??**. Note that some terms in the above expressions may be equal to zero when $w_t^{(i)}$ is small enough, but the bound always holds nonetheless. Since the above holds for all i, applying the tower rule we have

$$\frac{1}{(N)_3} \sum_{i=1}^{N} \mathbb{E}_t[(\nu_t^{(i)})_3] \leq \frac{a^2}{N-2} \frac{1}{(N)_2} \sum_{i=1}^{N} \mathbb{E}_t[(\nu_t^{(i)})_2]$$

satisfying (1) with $b_N := a^2/(N-2) \to 0$. The result then follows by applying Theorem 1. ∎

Lemma 3. *Consider an SMC algorithm using stratified resampling. Suppose that*

$$\varepsilon \leq q_t(x, x') \leq \varepsilon^{-1}$$

uniformly in x, x' for some $\varepsilon \in (0, 1]$, and that there exist $\zeta > 0$ and $\delta \in (0, 1)$ such that

$$\mathbb{P}[\max_i w_t^{(i)} - \min_i w_t^{(i)} \geq 2\delta/N \mid \mathcal{F}_{t-1}] \geq \zeta$$

for infinitely many t. Then, for all $N > 1$, $\mathbb{P}[\tau_N(t) = \infty] = 0$ for all finite t.

Proof. It is sufficient by a Borel-Cantelli argument, which is written somewhere else —SB to prove that under the stated conditions

$$\sum_{r=0}^{\infty} \mathbb{P}[c_N(r) > 2/N^2 \mid \mathcal{F}_{r-1}] = \infty.$$

Firstly,

$$\begin{aligned} \mathbb{P}[c_N(t) \leq 2/N^2 \mid \mathcal{H}_t] &= \mathbb{P}[c_N(t) = 0 \mid \mathcal{H}_t] = \mathbb{P}[\nu_t^{(i)} = 1 \, \forall i \in \{1, \dots, N\} \mid \mathcal{H}_t] \\ &\leq \mathbb{P}[\nu_t^{(i^\star)} = 1 \mid \mathcal{H}_t], \end{aligned} \tag{4}$$

where $i^\star := \operatorname{argmax}_i \{w_t^{(i)}\}$ (but note that the inequality holds when $i^\star$ is taken to be any particular index). Define for any $k \in \mathbb{Z}$

$$p_k^{(i)} := \mathbb{P}\left[\nu_t^{(i)} = \lfloor N w_t^{(i)} \rfloor + k \,\middle|\, \mathcal{H}_t\right].$$

Since stratified resampling almost surely results in $\nu_t^{(i)} \in \{\lfloor N w_t^{(i)} \rfloor - 1, \lfloor N w_t^{(i)} \rfloor, \lfloor N w_t^{(i)} \rfloor + 1, \lfloor N w_t^{(i)} \rfloor + 2\}$ we have that $p_k^{(i)} \equiv 0$ for $k \notin \{-1, 0, 1, 2\}$, and

$$\sum_{k=-1}^{2} p_k^{(i)} = \sum_{k=-1}^{2} \mathbb{P}\left[\nu_t^{(i)} = \lfloor N w_t^{(i)} \rfloor + k \,\middle|\, w_t^{(1:N)}\right] = 1.$$

Up to a proportionality constant C,

$$\begin{aligned} p_k^{(i)} &= C \, \mathbb{P}\left[\nu_t^{(i)} = \lfloor N w_t^{(i)} \rfloor + k \,\middle|\, w_t^{(1:N)}\right] \\ &\quad \times \sum_{\substack{a_{1:N} \in \{1, \dots, N\}^N: \\ |\{j : a_j = i\}| = \lfloor N w_t^{(i)} \rfloor + k}} \mathbb{P}\left[a_t^{(1:N)} = a_{1:N} \,\middle|\, \nu_t^{(i)}, w_t^{(1:N)}\right] \prod_{j=1}^{N} q_{t-1}(X_t^{(a_j)}, X_{t-1}^{(j)}) \end{aligned}$$

for each $k \in \{-1, 0, 1, 2\}$. We can bound each probability above and below using the almost sure bounds on q_{t-1} stated in the Lemma:

$$C \, \mathbb{P}\left[\nu_t^{(i)} = \lfloor N w_t^{(i)} \rfloor + k \,\middle|\, w_t^{(1:N)}\right] \varepsilon^N \leq p_k^{(i)} \leq C \, \mathbb{P}\left[\nu_t^{(i)} = \lfloor N w_t^{(i)} \rfloor + k \,\middle|\, w_t^{(1:N)}\right] \varepsilon^{-N}$$

then eliminate the constant C by normalising, to obtain lower bounds

$$\begin{aligned} p_k^{(i)} &\geq \frac{C\,\mathbb{P}[\nu_t^{(i)} = \lfloor N w_t^{(i)} \rfloor + k \mid w_t^{(1:N)}]\varepsilon^N}{\sum_{j=-1}^{2} C\,\mathbb{P}[\nu_t^{(i)} = \lfloor N w_t^{(i)} \rfloor + j \mid w_t^{(1:N)}]\varepsilon^{-N}} \\ &= \mathbb{P}[\nu_t^{(i)} = \lfloor N w_t^{(i)} \rfloor + k \mid w_t^{(1:N)}]\varepsilon^{2N}. \end{aligned} \tag{5}$$

Suppose that $\max_i w_t^{(i)} - \min_i w_t^{(i)} \geq 2\delta/N$. Then that at least one of $\{\max_i w_t^{(i)} \geq (1+\delta)/N\}$ and $\{\min_i w_t^{(i)} \leq (1-\delta)/N\}$ occurs. We will now examine each of these possibilities.

We can always write the maximum weight as $w_t^{(i^\star)} = \frac{1+\gamma}{N}$ for some $\gamma \geq 0$. Then, using (4),

$$\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{H}_t] \geq 1 - \mathbb{P}[\nu_t^{(i^\star)} = 1 \mid \mathcal{H}_t] = \begin{cases} 0 & \text{if } \gamma = 0 \\ 1 - p_0^{(i^\star)} & \text{if } \gamma \in (0,1) \\ 1 - p_{-1}^{(i^\star)} & \text{if } \gamma \in [1,2) \\ 1 & \text{if } \gamma \geq 2. \end{cases}$$

If $\gamma \in (0,1)$ then

$$1 - p_0^{(i^\star)} \geq \frac{3\gamma\varepsilon^{2N}}{4}$$

using (5) and Table 1 (p_0, U.B.). Similarly, if $\gamma \in [1,2)$ then by Table 1 (p_{-1}, U.B.),

$$1 - p_{-1}^{(i^\star)} \geq \left(1 - \frac{1}{4}\right)\varepsilon^{2N} \geq \frac{3\varepsilon^{2N}}{4}.$$

So overall, under the constraint $\max_i w_t^{(i)} \geq (1+\delta)/N$, we have

$$\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{H}_t] \geq \min_{\gamma \geq \delta} \left\{ \frac{3\gamma\varepsilon^{2N}}{4} \mathbb{1}_{\{\gamma \in [0,1)\}} + \frac{3\varepsilon^{2N}}{4} \mathbb{1}_{\{\gamma \in [1,2)\}} + \mathbb{1}_{\{\gamma \geq 2\}} \right\} = \frac{3\delta\varepsilon^{2N}}{4}.$$

Now for the minimum weight. Let $j^\star := \operatorname{argmin}_i\{w_t^{(i)}\}$ and write $w_t^{(j^\star)} = \frac{1-\gamma}{N}$, for some $\gamma \in [0,1]$. Then we have

$$\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{H}_t] \geq 1 - \mathbb{P}[\nu_t^{(j^\star)} = 1 \mid \mathcal{H}_t] = \begin{cases} 1 - p_1^{(j^\star)} & \text{if } \gamma \in (0,1] \\ 0 & \text{if } \gamma = 0. \end{cases}$$

If $\gamma \in (0,1]$ then

$$1 - p_1^{(j^\star)} \geq \left(1 - \frac{1 + (1-\gamma)}{2}\right)\varepsilon^{2N} = \frac{\gamma\varepsilon^{2N}}{2},$$

again using Table 1 (p_1, U.B.). Therefore, under the constraint $\min_i w_t^{(i)} \leq (1-\delta)/N$, we have

$$\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{H}_t] \geq \min_{\gamma \geq \delta} \left\{ \frac{\gamma\varepsilon^{2N}}{2} \right\} = \frac{\delta\varepsilon^{2N}}{2}.$$

Combining both cases, we find for arbitrary r

$$\mathbb{P}[c_N(r) > 2/N^2 \mid \mathcal{H}_r] \geq \frac{\delta\varepsilon^{2N}}{2} \mathbb{1}_{\{\max_i w_r^{(i)} - \min_i w_r^{(i)} \geq 2\delta/N\}}$$

so

$$\begin{aligned}\mathbb{P}[c_N(r) > 2/N^2 \mid \mathcal{F}_{r-1}] &\geq \frac{\delta \varepsilon^{2N}}{2} \mathbb{P}[\max_i w_r^{(i)} - \min_i w_r^{(i)} \geq 2\delta/N \mid \mathcal{F}_{r-1}] \\ &\geq \zeta \frac{\delta \varepsilon^{2N}}{2} > 0\end{aligned}$$

for infinitely many r. Hence

$$\sum_{r=0}^{\infty} \mathbb{P}[c_N(r) > 2/N^2 \mid \mathcal{F}_{r-1}] = \infty$$

as required. ■