Non-triviality condition (shortened version)

Suzie Brown

June 6, 2020

Multinomial resampling

Lemma 1. *For all $N \geq 2$, for all t,*

$$\mathbb{P}\left[c_N(t) > \frac{2}{N^2} \middle| \mathbf{w} = \left(\frac{1}{N}, \dots, \frac{1}{N}\right)\right] = 1 - \frac{N!}{N^N}.$$

Proof. Fix arbitrary t and $N \geq 2$. Since $2/(N)_2 > 2/(N^2)$ is the smallest possible non-zero value for $c_N(t)$,

$$\begin{aligned}\mathbb{P}\left[c_N(t) > \frac{2}{N^2} \mid \mathbf{w} = (1/N, \dots, 1/N)\right] &= 1 - \mathbb{P}[c_N(t) = 0 \mid \mathbf{w} = (1/N, \dots, 1/N)] \\ &= 1 - \mathbb{P}[\nu_t^{(1:N)} = (1, \dots, 1) \mid \mathbf{w} = (1/N, \dots, 1/N)].\end{aligned}$$

Conditional on the weights, $\nu_t^{(1:N)} \sim \text{Multinomial}(N, (1/N, \dots, 1/N))$, so the probability of interest is

$$\mathbb{P}[\nu_t^{(1:N)} = (1, \dots, 1) \mid \mathbf{w} = (1/N, \dots, 1/N)] = N! \prod_{i=1}^{N} \frac{1}{N} = \frac{N!}{N^N}.$$

□

Lemma 2. *In the neutral case (i.e. when all weights are equal at every time step) with multinomial resampling, there exists N_0 such that for all $N > N_0$, for all finite t, $\mathbb{P}[\tau_N(t) = \infty] = 0$.*

Proof. Let us rewrite the event of interest in a different way.

$$\begin{aligned}\mathbb{P}[\tau_N(t) = \infty] = 0 &\Leftrightarrow \mathbb{P}[\tau_N(t) < \infty] = 1 \\ &\Leftrightarrow \mathbb{P}\left[\min\left\{s > 1 : \sum_{r=1}^{s} c_N(r) > t\right\} < \infty\right] = 1 \\ &\Leftrightarrow \mathbb{P}\left[\exists s < \infty : \sum_{r=1}^{s} c_N(r) > t\right] = 1\end{aligned}$$

It is sufficient to show that, for all $N > N_0$, $c_N(r)$ is bounded away from zero infinitely often in r. We consider the sequence of events $E_r := \{c_N(r) > 2/N^2\}$ for $r \in \mathbb{N}$. In the neutral case, the resampled family sizes at each generation are independent, hence the events E_r are independent. By the second Borel-Cantelli lemma, E_r occurs infinitely often if $\sum_{r=1}^{\infty} \mathbb{P}(E_r) = \infty$. An expression for $\mathbb{P}(E_r)$ is given in Lemma 1. For any fixed $N \geq 2$, the probability is strictly positive and constant in r, so the Borel-Cantelli condition is satisfied, thus we conclude that E_r occurs infinitely often. Hence, taking $N_0 = 1$, we have that $\mathbb{P}[\tau_N(t) = \infty] = 0$ for all $N > N_0$ and all finite t, as required. □

Lemma 3. *For all $N \geq 2$, for all t, for any weight vector $(w_1, \dots, w_N) \in \mathcal{S}_{N-1}$,*

$$\mathbb{P}\left[c_N(t) > \frac{2}{N^2} \middle| \mathbf{w} = (w_1, \dots, w_N)\right] \geq \mathbb{P}\left[c_N(t) > \frac{2}{N^2} \middle| \mathbf{w} = \left(\frac{1}{N}, \dots, \frac{1}{N}\right)\right].$$

That is, the probability of having at least one merger is minimised by the vector of equal weights.

Proof. Fix arbitrary t and $N \geq 2$. Recall that

$$1 - \mathbb{P}\left[c_N(t) > \frac{2}{N^2} \mid \mathbf{w} = (w_1, \ldots, w_N)\right] = N! \prod_{i=1}^{N} w_i. \tag{1}$$

We will show that the global maximum of this function on the simplex $\mathcal{S}_{N-1}$ is attained at $\mathbf{w} = (1/N, \ldots, 1/N)$. This weight vector will therefore minimise the probability of the complementary event, implying the desired result.

First, since we are working on the simplex, we encode the constraint $\sum w_j = 1$ by rewriting the function to optimise as

$$f(\mathbf{w}) := \prod_{i=1}^{N} w_i = \left(1 - \sum_{j=1}^{N-1} w_j\right) \prod_{i=1}^{N-1} w_i$$

where we have also dropped the constant positive factor $N!$. Note that this objective function is non-negative and obtains its minimal value zero whenever one or more of the weights is equal to zero; since we are looking for a maximum we can assume that $w_i > 0$ for all i. Now, for every $k \in \{1, \ldots, N-1\}$, we solve

$$\frac{\partial f(\mathbf{w})}{\partial w_k} = \left(1 - w_k - \sum_{j=1}^{N-1} w_j\right) \prod_{i \neq k}^{N-1} w_i = 0.$$

The product over $i \neq k$ is constant and positive for each k, so this amounts to solving

$$w_k = 1 - \sum_{j=1}^{N-1} w_j = w_N$$

for all k. The unique solution is $w_1 = w_2 = \cdots = w_N = 1/N$.

To verify that the critical point is a maximum, we evaluate the Hessian H:

$$H_{kl}(\mathbf{w}) = \begin{cases} -2 \prod_{i \neq k}^{N-1} w_i & k = l \\ \left(1 - w_k - w_l - \sum_{j=1}^{N-1} w_j\right) \prod_{i \neq k,l}^{N-1} w_i & k \neq l \end{cases}$$

$$H_{kl}(1/N, \ldots, 1/N) = \begin{cases} -2 \left(\frac{1}{N}\right)^{N-2} & k = l \\ -\left(\frac{1}{N}\right)^{N-2} & k \neq l \end{cases}$$

and show that H is negative definite at $(1/N, \ldots, 1/N)$: for any $\mathbf{x} \in \mathbb{R}^{N-1} \setminus \{\mathbf{0}\}$,

$$\begin{aligned} \mathbf{x}^T H\left(\frac{1}{N}, \ldots, \frac{1}{N}\right) \mathbf{x} &= \sum_{k=1}^{N-1} \left[-2\left(\frac{1}{N}\right)^{N-2} x_k^2 - \sum_{l \neq k}^{N-1} \left(\frac{1}{N}\right)^{N-2} x_k x_l\right] = \left(\frac{1}{N}\right)^{N-2} \left[-\sum_{k=1}^{N-1} 2x_k^2 - \sum_{k=1}^{N-1} \sum_{l \neq k}^{N-1} x_k x_l\right] \\ &= \left(\frac{1}{N}\right)^{N-2} \left[-\sum_{k=1}^{N-1} x_k^2 - \sum_{k=1}^{N-1} \sum_{l=1}^{N-1} x_k x_l\right] = \left(\frac{1}{N}\right)^{N-2} \left[-\sum_{k=1}^{N-1} x_k^2 - \left(\sum_{k=1}^{N-1} x_k\right)^2\right] < 0. \end{aligned}$$

□

Theorem 1. *With multinomial resampling, conditional on any sequence of weight vectors* $\mathbf{w}_r^{(1:N)} \in \mathcal{S}_{N-1}; r \in \mathbb{N}$, *there exists* N_0 *such that for all* $N > N_0$, *for all finite* t, $\mathbb{P}[\tau_N(t) = \infty] = 0$.

Proof. As in Lemma 2, denote the sequence of events $E_r := \{c_N(r) > 2/N^2\}$ for $r \in \mathbb{N}$. We know from Lemma 2 that, in the neutral case, E_r occurs infinitely often. Lemma 3 tells us that $\mathbb{P}[E_r \mid \mathbf{w} = (w_1, \ldots, w_N)] \geq \mathbb{P}[E_r \mid \mathbf{w} = (1/N, \ldots, 1/N)]$ for all r. Therefore, by a coupling argument, we conclude that E_r occurs infinitely often in the non-neutral case as well. □

Conditional SMC with multinomial resampling

Define $\mathbf{w}^* := \frac{1}{N-1}\left[(1, \ldots, 1) - \mathbf{e}_{i^*}\right]$, where i^* is the immortal index at generation t, and $\mathbf{e}_i$ denotes the i^{th} canonical basis vector.

Lemma 4. *For all $N \geq 2$, for all t,*

$$\mathbb{P}\left[c_N(t) > \frac{2}{N^2}\middle|\mathbf{w} = \mathbf{w}^*\right] = 1 - \frac{(N-1)!}{(N-1)^{N-1}}$$

Proof. Under $\mathbf{w}^*$, the immortal parent has zero weight and is therefore assigned exactly one offspring (the immortal particle). The remaining $N-1$ offspring are assigned to the remaining $N-1$ parents according to a Multinomial distribution with equal weights. We therefore have

$$\mathbb{P}\left[c_N(t) > \frac{2}{N^2} \mid \mathbf{w} = \mathbf{w}^*\right] = 1 - \mathbb{P}[\nu_t^{(1:N)} = (1,\dots,1) \mid \mathbf{w} = \mathbf{w}^*] = 1 - (N-1)! \prod_{i \neq i^*}^{N} \frac{1}{N-1} = 1 - \frac{(N-1)!}{(N-1)^{N-1}}.$$

$\square$

Lemma 5. *In conditional SMC with multinomial resampling, in the optimal case where the weight vector is equal to $\mathbf{w}^*$ at every time step, there exists N_0 such that for all $N > N_0$, for all finite t, $\mathbb{P}[\tau_N(t) = \infty] = 0$.*

Proof. The proof is exactly the same as for Lemma 2; Lemma 4 provides the expression for $P(E_r)$ which is strictly positive and constant in r. $\square$

Lemma 6. *For all $N \geq 2$, for all t, for any weight vector $(w_1, \dots, w_N) \in \mathcal{S}_{N-1}$,*

$$\mathbb{P}\left[c_N(t) > \frac{2}{N^2}\middle|\mathbf{w} = (w_1,\dots,w_N)\right] \geq \mathbb{P}\left[c_N(t) > \frac{2}{N^2}\middle|\mathbf{w} = \mathbf{w}^*\right].$$

Proof. Recall that

$$1 - \mathbb{P}\left[c_N(t) > \frac{2}{N^2} \mid \mathbf{w} = (w_1,\dots,w_N)\right] = (N-1)! \prod_{i \neq i^*}^{N} w_i.$$

Ignoring the immortal particles, this is equivalent to multinomial resampling in the standard case (1), only with $N-1$ particles rather than N. As we saw in Lemma 3, this function is maximised at the vector of equal weights, in this case $\mathbf{w}_{-i^*} = \frac{1}{N-1}(1,\dots,1)$. This leaves zero weight for the immortal particle, so overall the maximum is attained at $\mathbf{w}^* = \frac{1}{N-1}\left[(1,\dots,1) - \mathbf{e}_{i^*}\right]$ as required. $\square$

Theorem 2. *In conditional SMC with multinomial resampling, conditional on any sequence of weight vectors $\mathbf{w}_r^{(1:N)} \in \mathcal{S}_{N-1}; r \in \mathbb{N}$, there exists N_0 such that for all $N > N_0$, for all finite t, $\mathbb{P}[\tau_N(t) = \infty] = 0$.*

Proof. As before, consider the sequence of events $E_r := \{c_N(r) > 2/N^2\}$ for $r \in \mathbb{N}$. We know from the argument behind Lemma 5 (which is completely analogous to Lemma 2) that, in the case $\mathbf{w} = \mathbf{w}^*$, E_r occurs infinitely often. Lemma 6 tells us that $\mathbb{P}[E_r \mid \mathbf{w} = (w_1,\dots,w_N)] \geq \mathbb{P}[E_r \mid \mathbf{w} = \mathbf{w}^*]$ for all r. Therefore, by a coupling argument, we conclude that E_r occurs infinitely often in the general case as well. $\square$

Stochastic rounding

Lemma 7. *Let $\mathbf{w} = (w_1,\dots,w_N) \in \mathcal{S}_{N-1}$ and resample by stochastic rounding.*

(i) If $w_i \geq 2/N$ for some i, then $\mathbb{P}[c_N(t) = 0|\mathbf{w}] = 0$.

(ii) If $w_i = 0$ for some i, then $\mathbb{P}[c_N(t) = 0|\mathbf{w}] = 0$.

Proof. In case (i) particle i is assigned at least two offspring, so $c_N(t)$ cannot be equal to zero. In case (ii) particle i is assigned zero offspring, so at least one other particle must be assigned more than one offspring, thus $c_N(t)$ cannot be equal to zero. $\square$

The upshot of Lemma 7 is that in these cases of "extreme weights" we have $c_N(t) > 2/N^2$ almost surely, so we can exclude these cases while we go about bounding $\mathbb{P}[c_N(t) > 2/N^2|\mathbf{w}]$ away from zero.

Lemma 8. *Define $\mathbf{w}^\delta := \frac{1}{N}\{(1,\dots,1) + \delta\mathbf{e}_i - \delta\mathbf{e}_j\}$ for any $i \neq j$ and $0 < \delta < 1$. Then $\mathbb{P}[c_N(t) > 2/N^2|\mathbf{w}^\delta] = \delta$.*

Proof.

$$\begin{aligned}\mathbb{P}[c_N(t) \le 2/N^2|\mathbf{w}^\delta] &= \mathbb{P}[\nu_t^{(1:N)} = (1,\dots,1) \mid \mathbf{w}^\delta] = \mathbb{P}[\nu_t^{(i)} = 1, \nu_t^{(j)} = 1 \mid \mathbf{w}^\delta] = \mathbb{P}[\nu_t^{(i)} = 1 \mid \mathbf{w}^\delta] \\ &= 1 - Nw_i^\delta + \lfloor Nw_i^\delta \rfloor = 1 - N(1+\delta)/N + 1 = 1 - \delta,\end{aligned}$$

since the offspring counts are deterministically equal to one apart from particles i and j, and it remains that $\nu_t^{(i)} = 1$ if and only if $\nu_t^{(j)} = 1$. The second line comes from the definition of stochastic rounding. □

Lemma 9. *Denote* $\mathcal{S}_{N-1}^\delta := \{\mathbf{w} \in \mathcal{S}_{N-1} : \forall i,\, 0 < w_i < \frac{2}{N};\, \max_i w_i \ge \frac{1+\delta}{N}\}$ *for any* $\varepsilon \in (0,1)$*. Then, for all* $\mathbf{w} \in \mathcal{S}_{N-1}^\delta$*,* $\mathbb{P}[c_N(t) > 2/N^2|\mathbf{w}] \ge \mathbb{P}[c_N(t) > 2/N^2|\mathbf{w}^\delta]$*.*

Proof. Fix arbitrary $\mathbf{w} \in \mathcal{S}_{N-1}^\delta$. Let i^* be then index of the particle with the largest weight. Denote $\mathcal{I} := \{i \in \{1,\dots,N\} : w_i > 1/N\}$. Notice that

$$\mathbb{P}[c_N(t) \le 2/N^2|\mathbf{w}] = \mathbb{P}[\nu_t^{(i)} = 1 \,\forall i \in \{1,\dots,N\}|\mathbf{w}] = \mathbb{P}[\nu_t^{(i)} = 1 \,\forall i \in \mathcal{I}|\mathbf{w}].$$

This is true because all weights are in $(0, 2/N)$, so for $i \in \mathcal{I}, \nu_t^{(i)} \in \{1,2\}$, and for $i \notin \mathcal{I}, \nu_t^{(i)} \in \{0,1\}$; and the offspring counts must sum to N (a generalisation of the argument used in Lemma 8).

We can then decompose this probability into a product of conditional probabilities:

$$\begin{aligned}\mathbb{P}[\nu_t^{(i)} = 1 \,\forall i \in \mathcal{I}|\mathbf{w}] &= \prod_{i \in \mathcal{I}} \mathbb{P}[\nu_t^{(i)} = 1|\nu_t^{(j)} = 1 \,\forall j < i \in \mathcal{I}; \mathbf{w}] \\ &= \mathbb{P}[\nu_t^{(i^*)} = 1|\mathbf{w}] \prod_{i \neq i^* \in \mathcal{I}} \mathbb{P}[\nu_t^{(i)} = 1|\nu_t^{(i^*)} = 1; \nu_t^{(j)} = 1 \,\forall j < i \in \mathcal{I}; \mathbf{w}] \\ &\le \mathbb{P}[\nu_t^{(i^*)} = 1|\mathbf{w}].\end{aligned}$$

The last line is equal to the probability $\mathbb{P}[c_N(t) \le 2/N^2|\mathbf{w}]$ in the case where $|\mathcal{I}| = 1$, i.e. the only weight larger than $1/N$ is w_{i^*}.

In other words, $\mathbb{P}[c_N(t) > 2/N^2|\mathbf{w}]$ is minimised on $\mathcal{S}_{N-1}^\varepsilon$ by having only one weight larger than $1/N$, in which case the values of the other weights do not affect this probability.

We therefore find that a minimum of $\mathbb{P}[c_N(t) > 2/N^2|\mathbf{w}]$ on $\mathcal{S}_{N-1}^\delta$ is given by $\mathbf{w}^{\delta'}$, for some $\delta' \ge \delta$. It only remains to show that taking $\delta' > \delta$ does not decrease the probability. This is a consequence of Lemma 8, where we see that $\mathbb{P}[c_N(t) > 2/N^2|\mathbf{w}^{\delta'}]$ is monotonically increasing in δ'. Thus the minimum of $\mathbb{P}[c_N(t) > 2/N^2|\mathbf{w}]$ is attained at $\mathbf{w} = \mathbf{w}^\delta$, as required. (Although this minimum is not unique, we have shown explicitly that it is a global minimum on $\mathcal{S}_{N-1}^\delta$.) □

Combining the above three Lemmata we see that, for any $\mathbf{w} \in \mathcal{S}_{N-1}$ such that $\max_i w_i \ge \frac{1+\delta}{N}$, we have the bound $\mathbb{P}[c_N(t) > 2/N^2|\mathbf{w}] \ge \delta$. This will probably turn out to be useful, provided we can say something about whether we can expect the constraint $\max_i w_i \ge \frac{1+\delta}{N}$ to hold.