Convergence corollary for residual-multinomial resampling: a round-up of failed attempts

Suzie Brown

26 May 2021

1 Conditioning on the index set of deterministically-assigned offspring

- $R := N - \sum \lfloor N w_t^{(i)} \rfloor$
- $r_i := \frac{1}{R}(N w_t^{(i)} - \lfloor N w_t^{(i)} \rfloor)$
- Parent i is deterministically assigned $\lfloor N w_t^{(i)} \rfloor$ offspring, for each i, and the remaining R offspring are assigned to parents chosen independently $\sim$ Categorical$(r_{1:N})$
- Let $\mathcal{I} \subseteq [N]$ denote the index set of offspring that are assigned to the "deterministic slots"
- $|\mathcal{I}| = N - R = \sum \lfloor N w_t^{(i)} \rfloor$
- $\mathcal{I} \mid w_t^{(1:N)}$ is uniform over the $\binom{N}{R}$ possible subsets of size $N - R$, due to the Standing Assumption
- $a_t^{\mathcal{I}}$ and $a_t^{\mathcal{I}^c}$ are conditionally independent given $\mathcal{I}$, due to the Standing Assumption
- The assumed bounds on g_t imply almost surely $w_t^{(i)} \in [\frac{1}{a^2 N}, \frac{a^2}{N}]$, hence $\lfloor N w_t^{(i)} \rfloor \in [a^{-2}, a^2]$ and $|\mathcal{I}| = O(N)$

So...

$$\mathbb{P}[a_t^{(1:N)} = a_{1:N} \mid \mathcal{H}_t] = \sum_{\mathcal{I} \subseteq [N]} \mathbb{P}[\mathcal{I} \mid \mathcal{H}_t] \, \mathbb{P}[a_t^{\mathcal{I}} = a_{\mathcal{I}} \mid \mathcal{I}, \mathcal{H}_t] \, \mathbb{P}[a_t^{\mathcal{I}^c} = a_{\mathcal{I}^c} \mid \mathcal{I}, \mathcal{H}_t] \tag{1}$$

$\mathbb{P}[\mathcal{I} \mid \mathcal{H}_t]$ is not tractable, but will sum to one if the other terms can be bounded independently of $\mathcal{I}$.

$$\mathbb{P}[a_t^{\mathcal{I}} = a_{\mathcal{I}} \mid \mathcal{I}, \mathcal{H}_t] \propto \left(\prod_{i=1}^{N} \mathbb{1}_{\{|\{j \in \mathcal{I} : a_j = i\}| = \lfloor N w_t^{(i)} \rfloor\}} \right) \left(\prod_{i \in \mathcal{I}} q_{t-1}(X_t^{(a_i)}, X_{t-1}^{(i)}) \right) \tag{2}$$

Indicators ensure correct number of deterministic slots for each parent, q's incorporate probability of particular parent-offspring assignment.

$$\mathbb{P}[a_t^{\mathcal{I}^c} = a_{\mathcal{I}^c} \mid \mathcal{I}, \mathcal{H}_t] \propto \prod_{i \in \mathcal{I}^c} r_{a_i} q_{t-1}(X_t^{(a_i)}, X_{t-1}^{(i)}) \tag{3}$$

r's are the probabilities from the Categorical sampling of parents, q's as above.

2 Correct dominance of moments but with the wrong conditioning

With residual-multinomial resampling, for each i

$$\nu_t^{(i)} \mid w_t^{(1:N)} \overset{d}{=} \lfloor N w_t^{(i)} \rfloor + X_i$$

where $X_i \sim \text{Binomial}(R, r_i)$. As usual, $R := N - \sum_{i=1}^N \lfloor N w_t^{(i)} \rfloor$ and $r_i := (N w_t^{(i)} - \lfloor N w_t^{(i)} \rfloor)/R$. If $R = 0$ then $r_i = 0$ for all i and the following calculations remain correct. We can therefore compute

$$\begin{aligned}
\mathbb{E}[(\nu_t^{(i)})_2 \mid w_t^{(1:N)}] &= \mathbb{E}\left[(\lfloor N w_t^{(i)} \rfloor + X_i)(\lfloor N w_t^{(i)} \rfloor + X_i - 1) \,\middle|\, w_t^{(1:N)} \right] \\
&= (\lfloor N w_t^{(i)} \rfloor)_2 + 2\lfloor N w_t^{(i)} \rfloor E[X_i \mid w_t^{(1:N)}] + \mathbb{E}[(X_i)_2 \mid w_t^{(1:N)}] \\
&= (\lfloor N w_t^{(i)} \rfloor)_2 + 2\lfloor N w_t^{(i)} \rfloor R r_i + (R)_2 r_i^2
\end{aligned}$$

using the moments of the Binomial distribution. We also have

$$\begin{aligned}
\mathbb{E}[(\nu_t^{(i)})_3 \mid w_t^{(1:N)}] &= \mathbb{E}\left[(\lfloor N w_t^{(i)} \rfloor + X_i)(\lfloor N w_t^{(i)} \rfloor + X_i - 1)(\lfloor N w_t^{(i)} \rfloor + X_i - 2) \,\middle|\, w_t^{(1:N)} \right] \\
&= \lfloor N w_t^{(i)} \rfloor^3 + \lfloor N w_t^{(i)} \rfloor^2 \mathbb{E}[3X_i - 3 \mid w_t^{(1:N)}] \\
&\quad + \lfloor N w_t^{(i)} \rfloor \mathbb{E}[X_i(X_i - 1) + X_i(X_i - 2) + (X_i - 1)(X_i - 2) \mid w_t^{(1:N)}] \\
&\quad + \mathbb{E}[(X_i)_3 \mid w_t^{(1:N)}] \\
&= \lfloor N w_t^{(i)} \rfloor^3 - 3\lfloor N w_t^{(i)} \rfloor^2 + 3\lfloor N w_t^{(i)} \rfloor^2 \mathbb{E}[X_i \mid w_t^{(1:N)}] \\
&\quad + \lfloor N w_t^{(i)} \rfloor \mathbb{E}[3X_i^2 - 6X_i + 2 \mid w_t^{(1:N)}] + E[(X_i)_3 \mid w_t^{(1:N)}] \\
&= \left(\lfloor N w_t^{(i)} \rfloor^3 - 3\lfloor N w_t^{(i)} \rfloor^2 + 2\lfloor N w_t^{(i)} \rfloor \right) + 3\left(\lfloor N w_t^{(i)} \rfloor^2 - \lfloor N w_t^{(i)} \rfloor \right) \mathbb{E}[X_i \mid w_t^{(1:N)}] \\
&\quad + 3\lfloor N w_t^{(i)} \rfloor \mathbb{E}[(X_i)_2 \mid w_t^{(1:N)}] + \mathbb{E}[(X_i)_3 \mid w_t^{(1:N)}] \\
&= (\lfloor N w_t^{(i)} \rfloor)_3 + 3(\lfloor N w_t^{(i)} \rfloor)_2 R r_i + 3\lfloor N w_t^{(i)} \rfloor (R)_2 r_i^2 + (R)_3 r_i^3 \\
&\leq \left(\lfloor N w_t^{(i)} \rfloor + R r_i \right) \left\{ (\lfloor N w_t^{(i)} \rfloor)_2 + 2\lfloor N w_t^{(i)} \rfloor R r_i + (R)_2 r_i^2 \right\} \\
&= N w_t^{(i)} \mathbb{E}[(\nu_t^{(i)})_2 \mid w_t^{(1:N)}] \\
&\leq a^2 \mathbb{E}[(\nu_t^{(i)})_2 \mid w_t^{(1:N)}],
\end{aligned}$$

using the almost sure bound $w_t^{(i)} \leq a^2/N$.

...

To complete the proof we need to exchange the conditioning on $w_t^{(1:N)}$ for conditioning on $\mathcal{H}_t$ so we can then invoke the D-separation and tower property to get:

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

Thus the main theorem condition is satisfied with $b_N = a^2/(N-2)$. The ε might want to get involved here as well once we switch the conditioning, coming from bounds on q_t (which would then have to be included in the statement of this corollary).

3 Comparing coalescence rates between residual-multinomial and multinomial resampling

The broad aim with this approach was to show a sort of c_N-ordering between residual-multinomial and multinomial resampling, conditional on the weight vector, which holds for any value of the weight vector and thus holds with any or no conditioning... Does it actually imply the same moment inequalities for the filtered expectations $\mathbb{E}_t[\cdot]$, i.e. when the conditioning is $\mathcal{F}_{t-1}$? I'm not sure.

3.1 Any N, one very large weight

It has been suggested that we should expect the expected coalescence rate to be always lower in the case of residual-multinomial, but this has not yet been proven. Some said that the proposed inequality could fail to hold when one or more weights are very large, causing a mega-coalescence to occur deterministically in the residual case, while multinomial resampling might chance to avoid the mega-coalescence. As a not-very-compelling-but-somewhat-informative example, consider a scenario where one of the weights is $(N-1)/N$ or more.

$$w_1 = \frac{N-1+\gamma}{N}$$

for some $\gamma \in [0,1]$. Thus the number of residual assignments is $R=1$ and the first residual is $r_1 = \gamma/1 = \gamma$.

$$\begin{aligned}
\mathbb{E}[c_N^{res} \mid w_{1:N}] &= \frac{1}{(N)_2}\{(N-1)_2 + (1)_2\}\mathbb{P}[\nu_1 = N-1 \mid w_{1:N}] + \frac{1}{(N)_2}(N)_2\mathbb{P}[\nu_1 = N \mid w_{1:N}] \\
&= \frac{(N-1)_2}{(N)_2}(1-\gamma) + \gamma \\
&= 1 - \frac{2(1-\gamma)}{N},
\end{aligned}$$

which, by the way, is equal to $2w_1 - 1$. Meanwhile under plain old multinomial resampling,

$$\begin{aligned}
\mathbb{E}[c_N^{mn} \mid w_{1:N}] = \sum_{i=1}^{N} w_i^2 &= \left(\frac{N-1+\gamma}{N}\right)^2 + \sum_{i=2}^{N} w_i^2 \\
&\geq \left(\frac{N-1+\gamma}{N}\right)^2 + (N-1)\left(\frac{1-\gamma}{N(N-1)}\right)^2 \\
&= \frac{1}{N^2}\left\{(N-1+\gamma)^2 + \frac{(1-\gamma)^2}{N-1}\right\} \\
&= \frac{1}{N^2}\left\{N^2 - 2(1-\gamma) + \frac{N}{N-1}(1-\gamma)^2\right\} \\
&\geq 1 - \frac{2(1-\gamma)}{N^2} \\
&\geq 1 - \frac{2(1-\gamma)}{N}.
\end{aligned}$$

Thus, even in this frightening case of huge weights, we see that the coalescence rate is (considerably) higher with multinomial resampling than with residual-multinomial resampling.

Alas, it is still unclear whether this c_N-ordering holds for *any* vector of weights, but this example might still our fears that it should fail when weights are large.

3.2 Any N, lots of quite large weights

One other special case we can construct in which $R = 1$ (and so direct computation is not too difficult) is that where $N - 1$ parents are deterministically assigned one offspring each:

$$w_1 = \frac{\gamma}{N}, \qquad w_i \in \left[\frac{1}{N}, \frac{2}{N}\right)$$

for each $i \neq 1$, where $\gamma \in [0, 1)$. Then, with residual-multinomial resampling,

$$\begin{aligned}
\mathbb{E}[c_N^{res} \mid w_{1:N}] &= 0\mathbb{P}[\nu_1 = 1 \mid w_{1:N}] + \sum_{i=2}^{N} \frac{2}{(N)_2} \mathbb{P}[\nu_i = 2 \mid w_{1:N}] \\
&= \frac{2}{(N)_2} \left(1 - \mathbb{P}[\nu_1 = 1 \mid w_{1:N}]\right) \\
&= \frac{2}{(N)_2}(1 - \gamma)
\end{aligned}$$

or, if you prefer, $= \frac{2}{N-1}((1/N) - w_1)$.

With plain old multinomial resampling,

$$\begin{aligned}
\mathbb{E}[c_N^{mn} \mid w_{1:N}] &= \sum_{i=1}^{N} w_i^2 = \left(\frac{\gamma}{N}\right)^2 + \sum_{i=2}^{N} w_i^2 \\
&\geq \left(\frac{\gamma}{N}\right)^2 + (N-1)\left(\frac{1}{N}\right)^2 \\
&= \frac{1}{N^2}(\gamma^2 + N - 1) \\
&= \frac{1}{(N)_2}\frac{N-1}{N}(\gamma^2 + N - 1) \\
&\geq \frac{1}{(N)_2}(\gamma^2 + N - 2) \\
&\geq \frac{1}{(N)_2}(2 - 2\gamma + N - 4) \\
&\geq \frac{2}{(N)_2}(1 - \gamma)
\end{aligned}$$

whenever $N \geq 4$. We see in this case also (though it was never in much doubt) that residual-multinomial resampling yields a lower expected coalescence rate than multinomial resampling.

We also have some results like the above that hold for *any* weight vector, but so far only in the cases $N = 2, 3$ (and the limiting case $N = \infty$). These special cases are included below. Now we just need to show that the same inequality holds for $N = 4, 5, \dots$ — ha ha.

3.3 Case $N = 2$

Lemma 1. *For all weight vectors $w_t^{(1:2)}$, $\mathbb{E}[c_2^m(t)|w_t^{(1:2)}] \geq \mathbb{E}[c_2^r(t)|w_t^{(1:2)}]$.*

Proof. With only $N = 2$ particles, the coalescence rate becomes

$$\mathbb{E}[c_N(t)|w_t^{(1:2)}] = \frac{1}{(N)_2}\sum_{i=1}^{N} \mathbb{E}\left[(\nu_t^{(i)})_2|w_t^{(1:2)}\right] = \mathbb{P}[\nu_t^{(1)} = 0] + \mathbb{P}[\nu_t^{(1)} = 2].$$

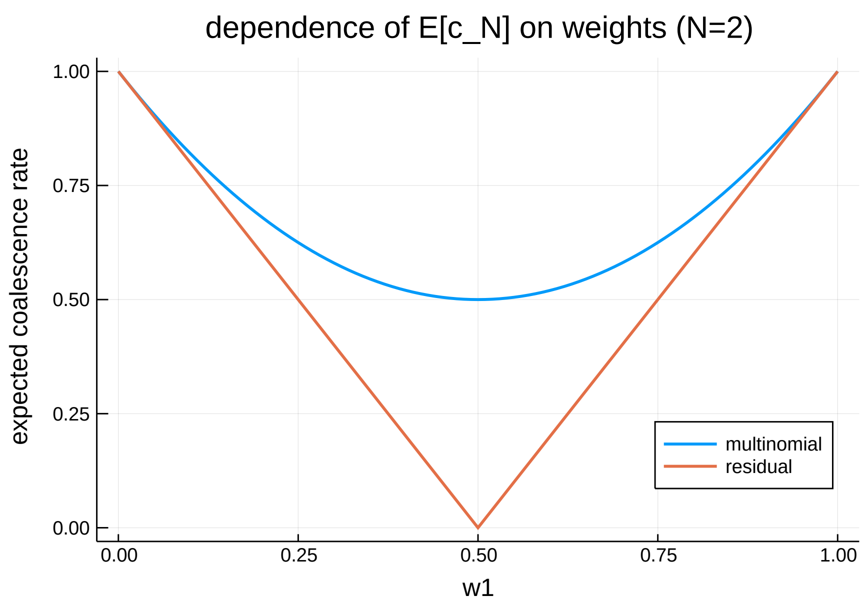
For residual resampling,

$$\mathbb{E}[c_2^r(t)|w_t^{(1:2)}] = \mathbb{1}_{\{w_t^{(1)} \geq 1/2\}}(2w_t^{(1)} - 1) + \mathbb{1}_{\{w_t^{(1)} < 1/2\}}(2w_t^{(2)} - 1)$$

And for multinomial resampling,

$$\begin{aligned}\mathbb{E}[c_2^m(t)|w_t^{(1:2)}] &= (w_t^{(1)})^2 + (w_t^{(2)})^2 \\ &= \mathbb{1}_{\{w_t^{(1)} \geq 1/2\}}((w_t^{(1)})^2 + (w_t^{(2)})^2) + \mathbb{1}_{\{w_t^{(1)} < 1/2\}}((w_t^{(1)})^2 + (w_t^{(2)})^2) \\ &\geq \mathbb{1}_{\{w_t^{(1)} \geq 1/2\}}(w_t^{(1)})^2 + \mathbb{1}_{\{w_t^{(1)} < 1/2\}}(w_t^{(2)})^2\end{aligned}$$

Then since $(w_t^{(i)} - 1)^2 = (w_t^{(i)})^2 - 2w_t^{(i)} + 1 \geq 0$, we have that $(w_t^{(i)})^2 \geq 2w_t^{(i)} - 1$ and hence we conclude the proof. $\square$



3.4 Case $N = 3$

Given a weight vector $(w_t^{(1)}, w_t^{(2)}, w_t^{(3)})$, let $w_{(1)} \geq w_{(2)} \geq w_{(3)}$ denote the weights sorted from high to low. With $N = 3$ there are many more cases than with $N = 2$, and these are described below, using the sorted weights.

In each case for the conditions on the sorted weights, the possible offspring count vectors (sorted in the same order as the weights) are listed, along with the probability of each (conditional on the given case). Finally, using these outcomes and associated probabilities, the conditional expectation of interest is calculated.

Case	Weights	Offspring counts	Conditional probabilities	$\mathbb{E}[c_2^r(t)\|w_t^{(1:3)}]$
(A)	$w_{(1)} = 1$	$(3,0,0)$	1	1
(B)	$2/3 < w_{(1)} < 1$	$(3,0,0)$ $(2,1,0)$ $(2,0,1)$	$3w_{(1)} - 2$ $3w_{(2)}$ $3w_{(3)}$	$2w_{(1)} - 1$
(C)	$w_{(1)} = 2/3$	$(2,1,0)$ $(2,0,1)$	$3w_{(2)}$ $3w_{(3)}$	$1/3$
(D1)	$1/3 < w_{(1)} < 2/3$ and $1/3 \leq w_{(2)} < 2/3$	$(2,1,0)$ $(1,2,0)$ $(1,1,1)$	$3w_{(1)} - 1$ $3w_{(2)} - 1$ $3w_{(3)}$	$1/3 - w_{(3)}$
(D2)	$1/3 < w_{(1)} < 2/3$ and $w_{(2)} < 1/3$	$(3,0,0)$ $(2,1,0)$ $(2,0,1)$ $(1,2,0)$ $(1,0,2)$ $(1,1,1)$	$(3/2)^2(w_{(1)} - 1/3)^2$ $(3/2)^2 2(w_{(1)} - 1/3)w_{(2)}$ $(3/2)^2 2(w_{(1)} - 1/3)w_{(3)}$ $(3/2)^2 w_{(2)}^2$ $(3/2)^2 w_{(3)}^2$ $(3/2)^2 2w_{(2)}w_{(3)}$	$(1/4)(3w_{(1)} - 1)(w_{(1)} + 1)$
(E)	$w_{(1)} = 1/3$	$(1,1,1)$	1	0

Lemma 2. *For all weight vectors $w_t^{(1:3)}$, $\mathbb{E}[c_3^r(t)|w_t^{(1:3)}] \leq \mathbb{E}[c_3^m(t)|w_t^{(1:3)}]$.*

Proof. We have the following expression in the case of residual resampling:

$$\begin{aligned}\mathbb{E}[c_3^r(t)|w_t^{(1:3)}] = (2w_{(1)} - 1)&\mathbb{1}_{\{2/3 \leq w_{(1)} \leq 1\}}\\ &+ (1/3 - w_{(3)})\mathbb{1}_{\{1/3 \leq w_{(1)} < 2/3\}}\mathbb{1}_{\{1/3 \leq w_{(2)} < 2/3\}}\\ &+ (1/4)(3w_{(1)} - 1)(w_{(1)} + 1)\mathbb{1}_{\{1/3 \leq w_{(1)} < 2/3\}}\mathbb{1}_{\{w_{(2)} < 1/3\}}\end{aligned}$$

compared to the following in the case of multinomial resampling:

$$\begin{aligned}\mathbb{E}[c_3^m(t)|w_t^{(1:3)}] &= w_{(1)}^2 + w_{(2)}^2 + w_{(3)}^2\\ &= (w_{(1)}^2 + w_{(2)}^2 + w_{(3)}^2)\mathbb{1}_{\{2/3 \leq w_{(1)} \leq 1\}}\\ &\quad + (w_{(1)}^2 + w_{(2)}^2 + w_{(3)}^2)\mathbb{1}_{\{1/3 \leq w_{(1)} < 2/3\}}\mathbb{1}_{\{1/3 \leq w_{(2)} < 2/3\}}\\ &\quad + (w_{(1)}^2 + w_{(2)}^2 + w_{(3)}^2)\mathbb{1}_{\{1/3 \leq w_{(1)} < 2/3\}}\mathbb{1}_{\{w_{(2)} < 1/3\}}.\end{aligned}$$

Hence it suffices to show the following:

(i) $(2w_{(1)} - 1) \leq (w_{(1)}^2 + w_{(2)}^2 + w_{(3)}^2)$

(ii) $(1/3 - w_{(3)}) \leq (w_{(1)}^2 + w_{(2)}^2 + w_{(3)}^2)$

(iii) $(1/4)(3w_{(1)} - 1)(w_{(1)} + 1) \leq (w_{(1)}^2 + w_{(2)}^2 + w_{(3)}^2)$

First consider (ii). Since $w_{(3)}$ is defined as the smallest of the three weights, we know that $w_{(3)} \in [0, 1/3]$. Meanwhile, the RHS is the sum of the squared weights, which is always between $1/3$ and 1. Therefore (ii) is true.

Now let us consider (i). We have the identity $(w_{(1)} - 1)^2 = w_{(1)}^2 - 2w_{(1)} + 1$, which implies that $w_{(1)}^2 \geq 2_w(1) - 1$. Since $w_{(2)}^2 + w_{(3)}^2 \geq 0$ we can therefore conclude that (i) is true.

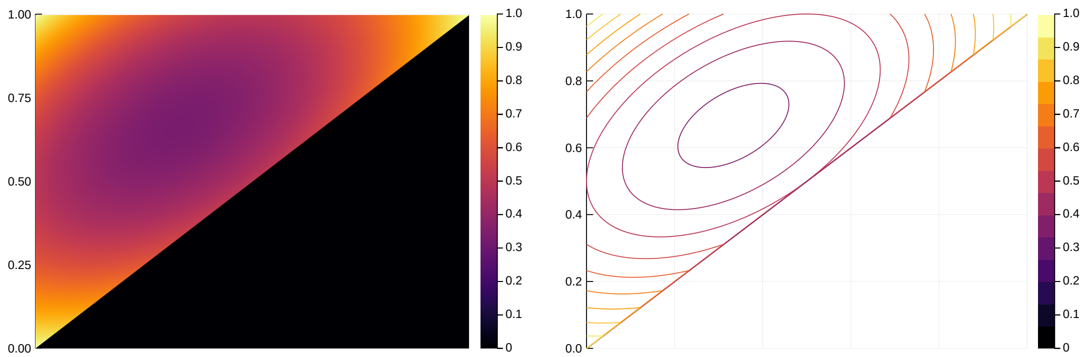
Finally consider (iii). In this case it is sufficient to show that $(1/4)(3w_{(1)} - 1)(w_{(1)} + 1) \leq w_{(1)}^2$. Note that

$$\begin{aligned}
&w_{(1)}^2 - \frac{1}{4}(3w_{(1)} - 1)(w_{(1)} + 1) \\
&= \frac{1}{4}w_{(1)}^2 - \frac{1}{2}w_{(1)} + \frac{1}{4} \\
&= \frac{1}{4}(w_{(1)} - 1)^2 \geq 0.
\end{aligned}$$

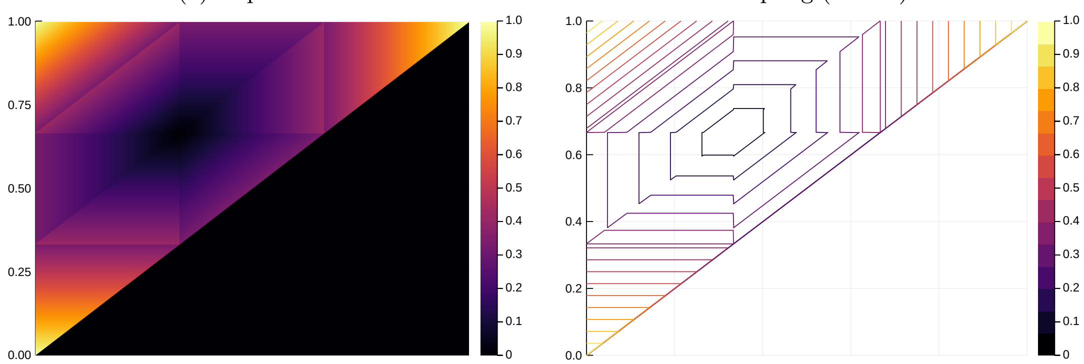
Therefore (iii) is also true, concluding the proof. □

Remark 1. *Examining the table, we can see that the function* $\mathbb{E}[c_3^r(t)|w_t^{(1:3)}]$ *is continuous in* $w_{(1)}$*. However, the plot clearly shows discontinuities. These occur at the boundaries between different orderings when we sort the weights from high to low.*

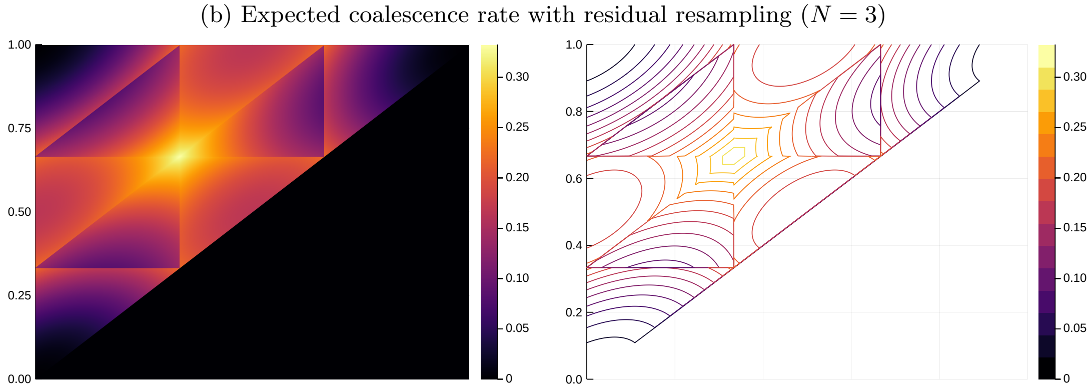
4 Coupling between multinomial and residual-multinomial resampling



(a) Expected coalescence rate with multinomial resampling ($N = 3$)



(b) Expected coalescence rate with residual resampling ($N = 3$)



(c) Difference between expected coalescence rate with multinomial or residual resampling ($N = 3$)