Bits of maths that didn't make it into the thesis

Suzie Brown

12 August 2021

1 Resampling parental indices or offspring counts?

Another "property" of resampling schemes. Does it directly sample the parental indices (and then offspring counts can be calculated), or does it directly sample offspring counts (and parental indices can then be assigned e.g. uniformly under these counts)?

Resampling scheme	Directly samples
`multi`	parental indices
`star`	parental indices
`strat`	parental indices
`strat-roulette`	parental indices
`syst`	parental indices
`res-multi`	offspring counts
`res-star`	offspring counts
`res-strat`	offspring counts
`res-syst`	offspring counts
`ssp`	offspring counts
`mvb`	offspring counts

Schemes based on inversion sampling (even roulette-style) transform each input point u_i into a parental index. Under star resampling, the parental indices and offspring counts contain the same information, so it can be equivalently formulated either way, but I've defined it as sampling one parent and then assigning all the offspring to that parent. The determinstic part of residual schemes does not specify particular parents, even if the stochastic part might. The SSP scheme is presented in terms of an algorithm for calculating offspring counts. The MVB scheme doesn't specify a particular algorithm but is defined in terms of the distribution of offspring counts, not parental indices.

2 Probability of killing a low-weight particle

This is another property of resampling schemes which I didn't include in the thesis. However it is very tractable and it leads to an alternative partial ordering (cf. thesis Figure 2.9) with a reasonable interpretation.

Suppose $w_t^{(i)} = \delta/N$ where $\delta \in [0,1)$. What is the probability that $\nu_t^{(i)} = 0$ under each resampling scheme?

Multinomial

$$\mathbb{P}[\nu_t^{(i)} = 0 \mid w_t] = \mathbb{P}[\text{Binomial}(N, \delta/N) = 0] = \left(1 - \frac{\delta}{N}\right)^N.$$

Residual-multinomial

Deterministic part is zero.

$$\mathbb{P}[\nu_t^{(i)} = 0 \mid w_t] = \mathbb{P}[\text{Binomial}(R, r_i) = 0] = (1 - r_i)^R = \left(1 - \frac{\delta}{R}\right)^R.$$

Star

$$\mathbb{P}[\nu_t^{(i)} = 0 \mid w_t] = \mathbb{P}[\text{Bernoulli}(\delta/N) = 0] = 1 - \frac{\delta}{N}.$$

Residual-star

Deterministic part is zero.

$$\mathbb{P}[\nu_t^{(i)} = 0 \mid w_t] = \mathbb{P}[\text{Bernoulli}(r_i) = 0] = 1 - r_i = 1 - \frac{\delta}{R}.$$

Stratified

Using Table 2.2 in thesis, with $\delta_L \in [0, \delta]$,

$$\mathbb{P}[\nu_t^{(i)} = 0 \mid w_t] = 1 - \delta + \delta_L(\delta - \delta_L) \in \left[1 - \delta, 1 - \frac{3}{4}\delta\right].$$

Residual-stratified

Same as stratified, since there is no deterministic part:

$$\mathbb{P}[\nu_t^{(i)} = 0 \mid w_t] = 1 - \delta + \delta_L(\delta - \delta_L) \in \left[1 - \delta, 1 - \frac{3}{4}\delta\right].$$

Stochastic rounding

$$\mathbb{P}[\nu_t^{(i)} = 0 \mid w_t] = \mathbb{P}[\nu_t^{(i)} = \lfloor Nw_t^{(i)} \rfloor] \mid w_t] = 1 - \delta.$$

2.1 Interpretation

The above quantities admit some inequalities, implying a partial ordering among resampling schemes. These inequalities are represented in the following graph, where $\mathbb{P}[\nu_t^{(i)} = 0 \mid w_t]$ is non-increasing along arrows. The dashed arrow indicates an inequality that holds iff $R > 1$ (if $R = 1$ the inequality goes the other way). The quantity under consideration is equal for straitfied and residual-stratified resampling. The missing arrows (those which are not implied by paths of arrows) represent quantities which do not admit an inequality holding for all δ, N.
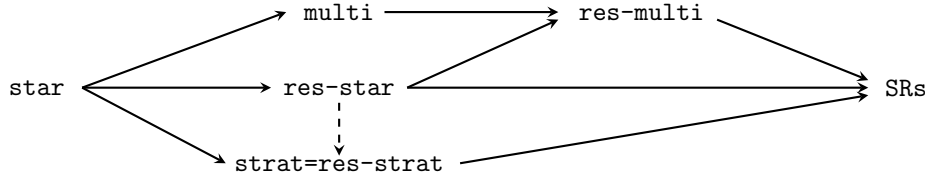
Since $\mathbb{E}[\nu_t^{(i)} \mid w_t]$ is equal across all resampling schemes (ensured by the unbiasedness property), the probability under consideration is capturing the spread around this mean value. Larger values of $\mathbb{P}[\nu_t^{(i)} = 0 \mid w_t]$ must be compensating for higher mass on large values of $\nu_t^{(i)}$ (in the sense of its conditional PMF). It is not surprising then that SRs come out smaller than everything else (since they give the lowest possible conditional variance of $\nu_t^{(i)}$, with mass only on 0 and 1) and that star resampling comes out bigger than everything else (since it gives the highest possible conditional variance of $\nu_t^{(i)}$, with mass only on 0 and N).

3 Properties of stratified roulette resampling

I have noted that Whitley's "roulette wheel" is an equivalent definition to the usual inversion sampling one in the case of systematic (or multinomial) resampling, but that these representations are not equivalent for stratified resampling. Indeed, the roulette wheel version is always "worse" because it unnecessarily adds extra randomness. In fairness, Whitley never proposed to use the roulette wheel for stratified resampling.

I didn't include the roulette version of stratified resampling in the analysis of resampling schemes, but I initially did some of the calculations as if I were going to include it. Here are those.

Degenerate under equal weights?

Standard stratified resampling: YES
Roulette stratified resampling: NO

When all weights are multiples of $1/N$, the sampling intervals still do not generally line up with the weight intervals (whereas they do in standard stratified resampling). It is therefore possible for the offspring counts to deviate from their means by 1 in either direction.

Marginal support of offspring counts

Standard stratified resampling: $\{K-1, K, K+1, K+2\}$
Roulette stratified resampling: $\{K-1, K, K+1, K+2\}$
(given $w_t^{(i)} = \frac{K+\delta}{N}$)

This follows from the same reasoning as usual (just draw a picture).

Marginal variance of offspring counts

Standard stratified resampling: ≤ 2
Roulette stratified resampling: ≤ 2

I haven't got the calculations here, but according to my notes the best upper bound I could find was the same in each case. This doesn't mean the actual variance is the same for both, just the bounds on the variance. I haven't attempted a direct comparison for say the same vector of weights under both resampling schemes, but I imagine there are some complications with doing such a calculation.

Monte Carlo variance

I would expect that perhaps the roulette version has higher variance, because of the additional randomness involved. I haven't proved it though.

Exchangeability

Standard stratified resampling: NO
Roulette stratified resampling: NO

It is not as obvious with the roulette wheel, because for instance offspring 1 is no more likely to choose parent 1 than offspring N is. But if you look at the joint distribution of parental indices you can see that it does not leave the offspring exchangeable. For instance, consecutive offspring tend to have parents that are near to each other, while two offspring with distant indices do not.

Permutation sensitivity

Standard stratified resampling: YES
Roulette stratified resampling: ?

I think probably yes, but I haven't proved it.

Computational complexity

Standard stratified resampling: $O(N)$
Roulette stratified resampling: $O(N)$

Adding a phase to each u_i requires at most $O(N)$ operations. I suspect that in practice the two algorithms in fact differ by only a constant number of operations.

Negative association

Standard stratified resampling: YES
Roulette stratified resampling: ?

Star discrepancy

Standard stratified resampling: $\in [1/(2N), 1/N]$
Roulette stratified resampling: $\in [1/(2N), 2/N]$

The best case is the same: in theory both standard and roulette versions can attain the regular grid with discrepancy $1/(2N)$. But the worst case for roulette version is much worse than the standard case, because it is possible for example for two points to fall arbitrarily close to zero. This is really just a feature of the definition of star discrepancy, which is more compatible with the standard inversion sampling than with the roulette wheel; one could define a star discrepancy on the wheel and then we would see no difference between the two formulations of stratified resampling.

Conditional independence

Standard stratified resampling: YES
Roulette stratified resampling: ?

Do you lose conditional independence because all sampled points share a common phase? Or are the points still independent because we can treat the phase as part of the original distribution of sampled points?