



Resampling algorithms for high energy physics simulations

Jimmy Olsson^{1,a}, Simon Plätzer^{2,b}, Malin Sjö Dahl^{3,c}

¹ Department of Mathematics, KTH Royal Institute of Technology, Lindstedtsvägen 25, 100 44 Stockholm, Sweden

² Particle Physics, Faculty of Physics, University of Vienna, Boltzmanngasse 5, 1090 Vienna, Austria

³ Department of Astronomy and Theoretical Physics, Lund University, Sölvegatan 14A, 223 62 Lund, Sweden

Received: 21 December 2019 / Accepted: 26 September 2020 / Published online: 10 October 2020

© The Author(s) 2020

Abstract We demonstrate that the method of interleaved resampling in the context of parton showers can tremendously improve the statistical convergence of weighted parton shower evolution algorithms. We illustrate this by several examples showing significant statistical improvement.

1 Introduction

Event generators are indispensable simulation tools for understanding high energy collisions in particle physics. A core part of the event generators are parton showers where hard scattering events, calculated perturbatively, are dressed up with more and more partons in an iterative manner, typically starting from some high energy scale Q and evolving down towards smaller and smaller scales until an infrared cutoff Q_0 is reached.

Branchings within the parton shower evolution occur with a rate $P(q, z, x)$, which determines the probability density of the dynamic variables associated to the branching. These are the scale q , additional splitting variables z , and additional variables x on which a certain type of branching depends. The latter can be collections of discrete and continuous values. Branchings are ordered in decreasing scale q in the evolution interval Q to Q_0 . The probability for a certain splitting not to occur between the scales Q and q is given by the so-called Sudakov form factor $\Delta_P(q|Q, x)$,

$$\Delta_P(q|Q, x) = \exp\left(-\int_q^Q dk \int dz P(k, z, x)\right).$$

The density for the dynamic variables describing an individual branching which is to occur below a scale Q of a previous branching (or an initial condition set by the hard process), is then given by

^a e-mail: jimmyol@kth.se

^b e-mail: simon.plaetzer@univie.ac.at (corresponding author)

^c e-mail: malin.sjodahl@thep.lu.se

$$\frac{dS_P(q|Q, z, x)}{dq dz} = \Delta_P(Q_0|Q, x)\delta(q - Q_0)\delta(z - z_0) + \Delta_P(q|Q, x)P(q, z, x)\theta(Q - q)\theta(q - Q_0), \quad (1)$$

where $\Delta_P(Q_0|Q, x)\delta(q - Q_0)\delta(z - z_0)$ represents events which did not radiate until they reached the cut-off scale Q_0 and z_0 is an irrelevant parameter associating fixed branching variables to the cutoff scale, where no actual branching will occur. Different branching types x_1, \dots, x_n , with a total rate given by $\sum_{i=1}^n P(q, z, x_i)$, can be drawn algorithmically using the competition algorithm (see e.g. [1,2]), which we will not discuss in more detail in this letter.

An individual splitting kernel $P(q, z, x)$ cannot generally be integrated to an invertible function, which would allow to draw from dS_P using sampling by inversion, so what is typically done is to find an overestimate $R(q, z, x)$ (s.t. $R(q, z, x) \geq P(q, z, x)$ for all q, z, x), which can be integrated to an invertible function. Equating the normalized integrated splitting kernel to a random number $r \in (0, 1)$ and solving for q then generates a new scale q with the distribution for $R(q, z, x)$, and it can be proved that accepting the proposed branching variables with probability $P(q, z, x)/R(q, z, x)$ generates the desired density $P(q, z, x)$ instead [1–5].

In practice, in the case of a positive definite splitting kernel $P(q, z, x)$ with a known overestimate $R(q, z, x)$, the algorithm to draw from dS_P then is given by Algorithm 1. This algorithm is statistically perfect in the sense that it produces the correct distribution using only (positive) unit weights. In case of splitting kernels of indefinite sign, or in an attempt to include variations of the splitting rate P attached to a fixed sequence of branching variables, weights different from unity are unavoidable, see [2,5–10] for some of these applications.

Especially when extending parton showers to include corrections beyond leading order [11–13] as well as sub-leading N_c effects and in attempts to perform the evolution at the amplitude level [10,14–20], negative contributions to the

Algorithm 1 The Sudakov veto algorithm, defined using an overestimate R such that $R(q, z, x) \geq P(q, z, x)$ for all q, z, x .

```

 $Q' \leftarrow Q$ 
loop
   $r \leftarrow \text{rnd}$ 
  if  $r \leq \Delta_R(Q_0|Q', x)$  then
    return  $Q_0$ 
  else
    solve  $r = \Delta_R(q|Q', x)$  for  $q$ 
    select  $z$  in proportion to  $R(q, z, x)$ 
    return  $q$  with probability  $P(q, z, x)/R(q, z, x)$ 
  end if
 $Q' \leftarrow q$ 
end loop

```

splitting rates arise. These contributions require the weighted Sudakov veto algorithm, Algorithm 2, in order to be included in the simulation.

Algorithm 2 The weighted Sudakov veto algorithm from [7], starting with an initial non-zero weight w_0 and a scale Q . Here S_R can be any density according to which we can generate branching variables, not necessarily defined by an overestimate of the target splitting kernel P . The acceptance probability, $0 < \epsilon < 1$, may in principle depend on the splitting variables, but for the purpose of this letter we will treat it as a constant, though this is not a conceptual limitation.

```

 $Q' \leftarrow Q, w \leftarrow w_0$ 
loop
  A trial splitting scale and variables,  $q, z$ , are generated according to  $S_R(q|Q', z, x)$ , for example using Algorithm 1.
  if  $q = Q_0$  then
    There is no emission and the cut-off scale  $Q_0$  is returned while the event weight is kept at  $w$ .
  else
    if  $\text{rnd} \leq \epsilon$  then
      The trial splitting variables  $q, z$  are accepted, and
      
$$w \leftarrow w \times \frac{1}{\epsilon} \times \frac{P(Q', z, x)}{R(Q', z, x)}. \quad (2)$$

    else
      The emission is rejected, and the algorithm continues with
      
$$w \leftarrow w \times \frac{1}{1 - \epsilon} \times \left(1 - \frac{P(q, z, x)}{R(q, z, x)}\right) \quad (3)$$

       $Q' \leftarrow q$ .
    end if
  end if
end loop

```

While the weighted Sudakov veto algorithm can be shown to correctly account for negative contributions, an issue with this algorithm and in general with weighting procedures, is largely varying weight distributions which accumulate mul-

tiplicatively during the simulation of an event, especially in presence of the competition algorithm. As a consequence, the Monte Carlo variance of the algorithm increases approximately geometrically fast with the number of weight changes caused by each emission attempt. In [10] the weight degeneration was partly reduced by only keeping event weights down to the winning scale in the competition step. This greatly reduced the weight spreading, but did not fully resolve the issue, in the sense that incorporating yet more emissions with negative weights would have produced unmanageable weights.

Approaching the parton shower precision era, we expect the issue of negative or varying weights to be a severe obstacle that has to be overcome, preferably by more general methods than to case by case monitor weights and adjust weighting algorithms. We therefore suggest to utilize the method of *resampling* [21, 22] in the context of Monte Carlo event generators. We introduce the new method in this letter as follows: in Sect. 2, we introduce the basics of the resampling method, while a benchmark example, illustrating the resampling power by comparing the proposed approach to a simplified parton shower is given in Sect. 4. We also show the efficiency of resampling by taking a more realistic parton shower [23] with unit weights, destroying its statistical convergence beyond recognition using the weighted Sudakov algorithm, Algorithm 2, and recovering good statistical convergence using the resampling method. Finally we make an outlook and discuss improvements in Sect. 5.

2 Resampling

The key idea behind *resampling* is to, rather than *reweight*¹ N weighted events, select n events among the N weighted ones in proportion to their weights. This is done in an interleaved procedure which is supposed to keep the weight distribution as narrow as possible at each intermediate step. After each weight change (in our case induced by the shower), all selected events are assigned the weight N/n and are further evolved separately.² It is important to mention that this method will introduce correlated events; some of them will be duplicated or have partly identical evolution history. The number n of selected events may well be chosen equal to the number N of events, but can also be chosen differently.

Note that the resampling procedure will not alter the flexibility of the simulation when it comes to predicting arbitrarily

¹ This is typically referred to as 'unweighting' in the context of high energy physics event simulation.

² The extension to an initially weighted sample of events (where the weight average is not necessarily one) is trivial: save the average event weight, and set (assuming N kept events) each event weight to this average after finishing the simulation.

many differential cross sections in one run of the program; in fact, the appearance of correlated events is also known in approaches using Markov Chain Monte Carlo (MCMC) methods as explored for the hard process cross sections in [24].

In the more commonly used *reweighting* procedure, events with small weights are selected with a low probability; however, if selected, they are assigned the same nominal weight (typically 1) as other events. Using *resampling*, also the large weights are kept under control. Instead of being allowed to increase without limit, all event weights can be kept at unity through duplication of events with large weights. This greatly improves the statistical convergence properties, and mitigates the risks of obtaining distributions that appear statistically wrong for a small number of events.

In an implementation of a parton shower, a natural way of implementing the resampling procedure is to resample after each emission step, i.e. to let each event radiate according to whatever versions of Sudakov factors and splitting kernels are used, and to then, among the say N' evolved events, pick N' events in proportion to their weights w_i , meaning that some events will be duplicated and some will be forgotten. This can even be extended to apply to intermediate proposals within the veto and competition algorithm itself; we will explore both options in Sect. 4. In any way, the resampling mechanism is guiding the gradual evolution of the events by imposing a selection pressure for fitness as measured by the weights.

Events which have reached the cut-off scale Q_0 can be ignored in the resampling procedure, since these are not evolved and will hence not acquire any further weights in the shower evolution. Rather, resampling among events which already have unit weights can only impair the statistical convergence, since some events will be “forgotten”, whereas others will be duplicated. After resampling, all the resampled events are assigned the same weight; this weight is given by the average weight of the evolved events among which the resampled events are drawn. As a consequence, resampling will not introduce any statistical bias, in the sense that distributions of observables before and after resampling are statistically equivalent; we refer to Appendix A for further mathematical details. Since the repeated resampling operations continuously uniformise the weights of the evolving events, they serve to stabilise numerically Algorithm 2, leading to a dramatic reduction of variance.

In the broader context of Monte Carlo simulation, resampling was first introduced in [21, 25] as a means for transforming Monte Carlo samples of weighted simulations into uniformly weighted samples. Later, [22] proposed resampling as a tool for controlling weight degeneracy in *sequential importance sampling methods* [26]. In sequential importance sampling, weighted Monte Carlo samples targeting sequences of probability distributions are formed by means of recursive

sampling and multiplicative weight updating operations. The observation in [22], that recurrent resampling guarantees the numerical stability of sequential importance sampling estimators over large time horizons—a finding that was later confirmed theoretically in [27]—lead to an explosive interest in such *sequential Monte Carlo methods* during the last decades. Today sequential Monte Carlo methods constitute a standard tool in the statistician’s tool box and are applied in a variety of scientific and engineering disciplines such as computer vision, robotics, machine learning, automatic control, image/signal processing, optimisation, and finance; see e.g. the collection [28].

We emphasise that when implemented properly, the resampling procedure does not involve any significant time-penalty. The selection of n events among N ones in proportion to their weights w_i is equivalent to drawing n independent and identically distributed indices i_1, \dots, i_n among $\{1, \dots, N\}$ such that the probability that $i_\ell = j$ is $w_j / \sum_{\ell=1}^N w_\ell$. Naively, one may, using the “table-look-up” method, generate each such index i_ℓ by simulating a uniform (i.e., a uniformly distributed number) u_ℓ over $(0, 1)$, finding k such that

$$\sum_{i=1}^{k-1} \frac{w_i}{\sum_{j=1}^N w_j} < u_\ell \leq \sum_{i=1}^k \frac{w_i}{\sum_{j=1}^N w_j},$$

and letting $i_\ell = k$. Assume that $n = N$ for simplicity; then, since using a simple binary search tree determining k requires, on the average, $\log_2 N$ comparisons, the overall complexity for generating $\{i_\ell\}$ is $N \log_2 N$. However, it is easily seen that if the uniforms $\{u_\ell\}$ are *ordered*, $u_{(1)} \leq u_{(2)} \leq \dots \leq u_{(N)}$, generating the N indices $\{i_\ell\}$ using the table-look-up method requires only N comparisons. Hence, once a sample $\{u_{(\ell)}\}$ of N ordered uniforms can be simulated at a cost increasing only linearly in N , the overall complexity of the resampling procedure is linear in N .

In the following we review one such simulation approach, which is based on the observation that the distribution of the order statistics $(u_{(1)}, \dots, u_{(n)})$ of n independent uniforms coincides with the distribution of

$$\left(\prod_{i=1}^n \sqrt[i]{v_i}, \prod_{i=2}^n \sqrt[i]{v_i}, \dots, n^{-1} \sqrt[n-1]{v_{n-1}} \sqrt[n]{v_n}, \sqrt[n]{v_n} \right),$$

where the n random variables $\{v_i\}$ are again independent and uniformly distributed over $(0, 1)$; see [29, Chapter 5]. As a consequence, a strategy for simulating $\{u_{(\ell)}\}$ is given by the Algorithm 3 having indeed a linear complexity in n .

Alternatively, one may apply the method of *uniform spacings*; see again [29, Chapter 5] for details. A caveat of the resampling procedure that one must be aware of is that it typically leads to event path depletion for long simulations. Indeed, when the number of iterations is large the ancestral paths of the events will typically, if resampling is performed

Algorithm 3 Algorithm generating a sample $\{u_{(\ell)}\}$ of n ordered uniforms. The uniforms $\{v_i\}$ are supposed to be independent.

```

Generate  $v_n$  from the uniform distribution over  $(0, 1)$ .
Set  $u_{(n)} = \sqrt[n]{v_n}$ .
for  $i = n - 1 \rightarrow 1$  do
    Generate  $v_i$  from the uniform distribution over  $(0, 1)$ .
    Set  $u_{(i)} = u_{(i+1)}\sqrt[n]{v_i}$ .
end for
    
```

systematically, coincide before a certain random point; in other words, the event paths will form an ancestral tree with a common ancestor. For $n = N$ one may establish a bound on the expected height of the “crown” of this ancestral tree, i.e., the expected time distance from the last generation back to the most recent common ancestor, which is proportional to $N \ln N$; see [30]. Thus, when the number of iterations increases, the ratio of the height of the “crown” to that of the “trunk” tends to zero, implying high variance of path space Monte Carlo estimators.

The most trivial remedies for the path depletion phenomenon is to increase the sample size N or to reduce the number of resampling operations. In [31] it is suggested that resampling should be triggered adaptively and only when the *coefficient of variation*

$$\begin{aligned}
 C_V^2 &= N \sum_{i=1}^N \left(\frac{w_i}{\sum_{\ell=1}^N w_\ell} - \frac{1}{N} \right)^2 \\
 &= N \sum_{i=1}^N \left(\frac{w_i}{\sum_{\ell=1}^N w_\ell} \right)^2 - 1
 \end{aligned} \tag{4}$$

exceeds a given threshold. The coefficient of variation detects weight skewness in the sense that C_V^2 is zero if all weights are equal. On the other hand, if the total weight is carried by a single draw (corresponding to the situation of maximal weight skewness), then C_V^2 is maximal and equal to $N - 1$. A closely related measure of weight skewness is the *effective sample size*

$$\text{ESS} = \frac{N}{1 + C_V^2}.$$

Heuristically, ESS measures, as its name suggests, the number of samples that contribute effectively to the approximation. It takes on its maximal value N when all the weights are equal and its minimal value 1 when all weights are zero except for a single one. Thus, a possible adaptive resampling scheme could be obtained by executing resampling only when ESS falls below, say, 50%.³ We refer to [32] for a discussion and

³ We remark that the usefulness of the concept effective sample size in the context of resampling, lies solely in the ability to trigger sampling. Calculating the ESS on a resampled set is meaningless, since all event weights are put equal during the resampling procedure.

a theoretical analysis of these measures in the context of sequential Monte Carlo sampling.

With respect to negative contributions to the radiation probability, showing up as negative weighted events, we note that events may well keep a negative weight. In this case, resampling is just carried through in proportion to $|w_i|$, whereas events with $w_i < 0$ contribute negatively when added to histograms showing observables. Negative contributions, for example from subleading color or NLO, do thus not pose an additional problem from the resampling perspective.

3 A view to variance estimation

Since the resampling operation induces complex statistical dependencies between the generated events, statistical analysis of the final sample is non-trivial. Still, as we will argue in the following, such an analysis is possible in the light of recent results on variance estimation in sequential Monte Carlo methods [33,34]. As explained in Sect. 2, subjecting continuously evolving events to resampling induces genealogies of the same. In [33] the authors propose estimating the variance of sequential Monte Carlo methods by stratifying the final samples with respect to the distinct time-zero ancestors traced by these genealogies. As the families of events formed in this way will be only weakly dependent (via resampling), this approach allows the Monte Carlo variance to be estimated on the basis of a statistic that is reminiscent of the standard sample variance for independent data. More precisely, let $\{(\xi_i, w_i)\}_{i=1}^N$ be a sample of events with associated weights generated by means of n resampling operations and let ε_n^i denote the index of the time-zero ancestor of the event ξ_i ;⁴ then the estimator of an observable h may be expressed as

$$\frac{1}{N} \sum_{i=1}^N w_i h(\xi_i) = \frac{1}{N} \sum_{i=1}^N v_i \stackrel{\text{not.}}{=} \bar{v}, \tag{5}$$

where $v_i = \sum_{\ell: \varepsilon_n^\ell = i} w_\ell h(\xi_\ell)$ is the contribution of the family of events originating from the time-zero ancestor i . Note that the time-zero ancestors can be easily kept track of as the sample propagates; indeed, given the time-zero ancestor indices $\{\varepsilon_n^i\}_{i=1}^N$ after n resampling operations, the indices $\{\varepsilon_{n+1}^i\}_{i=1}^N$ after $n + 1$ operations are, for all i , given by

$$\varepsilon_{n+1}^i = \varepsilon_n^i,$$

where $\{\ell_i\}_{i=1}^N$ are the ancestor indices. Since the variables $\{v_i\}_{i=1}^N$ are derived from disjoint sets of event paths, there

⁴ In [34], these indices are referred to as *Eve indices*. Note that we have added n to the notation of these indices in order to stress that the genealogical lineages, and hence the time-zero ancestors, typically change every time resampling is applied.

is only a weak statistical dependence between the same. Recall that in the case of truly independent and identically distributed draws $\{y_i\}_{i=1}^N$, the classical estimator

$$s^2 = \frac{1}{N(N-1)} \sum_{i=1}^N (y_i - \bar{y})^2 = \bar{y}^2 \left(1 - \frac{N}{N-1}\right) + \left(\frac{N}{N-1}\right) \frac{1}{N^2} \sum_{i=1}^N y_i^2$$

of the variance of the sample mean \bar{y} is unbiased for all N . For sequential importance sampling with resampling, the authors of [34] propose (as a refinement of the estimator suggested by [33]) the generalisation

$$\bar{v}^2 \left[1 - \left(\frac{N}{N-1}\right)^{n+1}\right] + \left(\frac{N}{N-1}\right)^{n+1} \frac{1}{N^2} \sum_{i=1}^N v_i^2 \tag{6}$$

(where, again, n is the number of resampling operations) of s^2 . Interestingly, in the case where the full Monte Carlo sample is involved in every resampling operation, it is possible to show that (6) is indeed an unbiased estimator of the variance of \bar{v} for all N ; see [34, Theorem 4]. Even though we proceed somewhat differently and exclude non-evolving events from the resampling, we expect the same approach to apply to our estimator. More specifically, since resampling among non-evolving events can only deteriorate statistical convergence, the variance estimate in (6) from the all-evolving case, must serve as an upper estimate of our partially resampled sets. However, a full study of the theoretical and empirical properties of an estimator of type (6) in our context is the topic of ongoing research and hence beyond the scope of the present paper.

4 Benchmark examples

In this section we illustrate the impact of the resampling procedure both in a simplified setup and within a full parton shower algorithm. As a toy model we consider a prototype shower algorithm with infrared cutoff Q_0 and splitting kernels

$$dP(q, z, x) = a \frac{dq}{q} \frac{(1+z^2)dz}{1-z} \theta(1 - Q_0/q - z) \theta(z - x), \tag{7}$$

with $a > 0$ being some parameter (similar to a coupling constant) and x an external parameter in the form of an initial ‘‘momentum fraction’’ (first arbitrarily set to 0.1 and later changing in the shower with each splitting).

Starting from a scale Q (initially we make the arbitrary choice $Q = 1$, and pick a cut-off value of $Q_0 = 0.01$) we obtain a new lower scale q and a momentum fraction z (of

the previous momentum), sampled according to the Sudakov-type density $dS_P(q|Q, z, x)$ associated with the splitting kernel above.

The momentum fraction parameter x will be increased to x/z after the emission, such that for an emission characterized by (q_i, z_i, x_i) the variables for the next emission will be drawn according to $dS_P(q_{i+1}|q_i, z_{i+1}, x_i/z_i = x_{i+1})$. This algorithm resembles ‘‘backward evolution’’ to some extent, but in this case we rather want to define a model which features both the infrared cutoff (upper bound, $1 - \frac{Q_0}{Q}$, on z and lower bound on evolution scale $q > Q_0$) and the dynamically evolving phase space for later emissions (lower bound on z), while still being a realistic example of a parton shower, with double and single logarithmic enhancement in Q/q .

Proposals for the dynamic variables q and z are generated using a splitting kernel of $R(q, z, x) = \frac{a}{q} \frac{z}{1-z}$ in place of $P(q, z, x) = \frac{a}{q} \frac{1+z^2}{1-z}$, and with simplified phase space boundaries $0 < z < 1 - Q_0/q$. We consider the distributions of the intermediate variables q and z , as well as the generated overall momentum fraction x , for up to eight emissions.

To consider a realistic scenario, similar to what is encountered in actual parton shower algorithms, where many emission channels typically compete, we generate benchmark distributions by splitting the parameter a into a sum of different values, (arbitrarily [0.01, 0.002, 0.003, 0.001, 0.01, 0.03, 0.002, 0.002, 0.02, 0.02]) corresponding to competing channels. The usage of weighted shower algorithms amplified by competition is a major source of weight degeneration, as the weights from the individually competing channels need to be multiplied together.

We perform resampling after the weighted veto algorithm (choosing a constant $\epsilon = 0.5$ in this example) and the competition algorithm have proposed a new transition, as schematically indicated in the left panel in Fig. 1. In Fig. 2 we illustrate the improvement of the resampling algorithm.

We also try the resampling strategy out in a more realistic parton shower setting, using the Python dipole parton shower from [23]. In this case, we run 10,000 events in parallel, and each event is first assigned a new emission scale (or the termination scale Q_0) using the standard overestimate Sudakov veto algorithm combined with the competition algorithm. For each shower separately, we then use one veto step as in Algorithm 2 such that some emissions are rejected. In the ratio P/R , P contains the actual splitting kernel and coupling constant, whereas R contains an overestimate with the singular part of the splitting kernel and an overestimate of the strong coupling constant, see [23] for details. The constant ϵ in Algorithm 2 is put to 0.5. This results in weighted shower events, and the resampling step, where 10,000 showers among the 10,000 are randomly kept in proportion to their weight, is performed, again resulting in unweighted showers. This procedure is illustrated in the right panel in Fig. 1, and

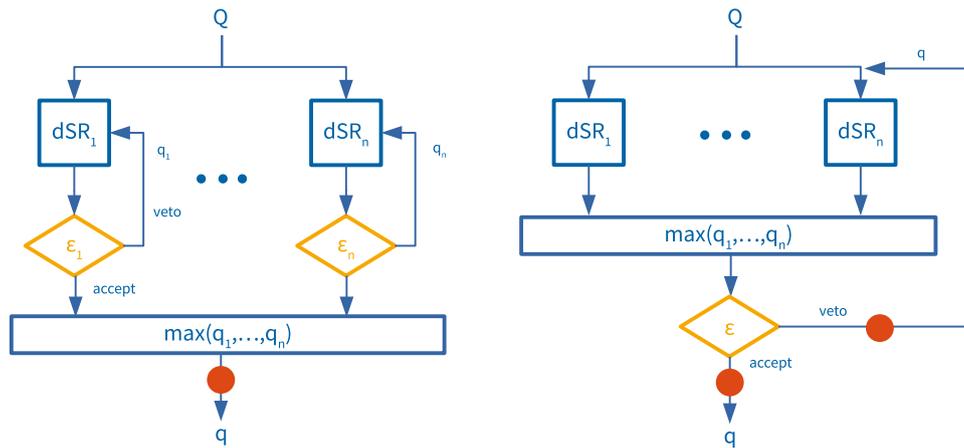


Fig. 1 Illustration of the different algorithms we consider as benchmark algorithms; we depict the algorithm as executed in each shower instance. Each dSR_i block corresponds to a proposal density for one of n competing channels, ϵ diamonds depict the acceptance/rejection step with the weighting procedure of the Sudakov veto algorithm, and the selection of the highest scale within the competition algorithm. Red dots indicate when the algorithm in each shower instance is interrupted and resampling in between the different showers is performed. Depending on the resampling strategy, events which reached the cutoff can be

included in the resampling, or put aside as is done in our benchmark studies. The left flow diagram corresponds to the benchmark example discussed, while the right flow diagram corresponds to the implementation in the Python dipole shower. Re-entering after a veto step (the ‘veto’ branches), will start the proposal from the scale just vetoed. Note that in the first case (left), resampling is performed only after emission (or shower termination), whereas in the second case, it is performed after each (sometimes rejected) emission attempt

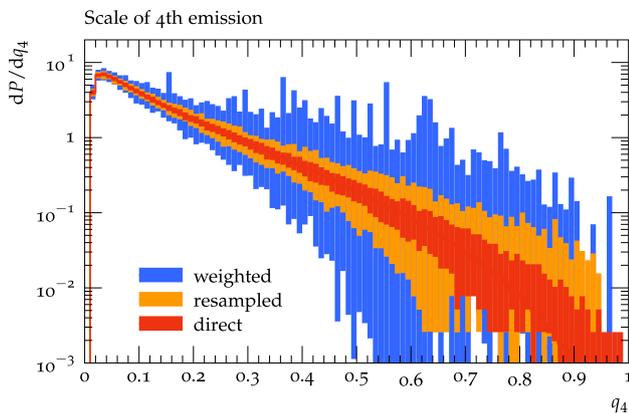


Fig. 2 Distribution of the scale of the 4th emission, for sampling the distribution with the competition algorithm splitting a into a sum across 10 different competing channels, with red being the direct algorithm, blue the weighted Sudakov algorithm and orange the resampled, weighted Sudakov algorithm. The bands in the distribution reflect the minimum and maximum estimate encountered in 300 runs with different random seeds and thus give a rough estimate of the width of the distribution of estimates

we remark that the resampling step can be combined with the shower in many different ways, for example, as in the first example above (to the left in the figure) after each actual emission (or reaching of Q_0), or, as in the second example (to the right) also after rejected emission attempts.

The result for the Durham y_{45} jet observable [35] is shown in Fig. 3. The smooth curve in red gives the unaltered result for the Durham y_{45} jet observable [35]; the jagged blue curve

represents the same parton shower with equally many events, but where the weighted Sudakov algorithm, Algorithm 2, has been used. The resulting distribution is so jagged and distorted that it appears incompatible at first sight. Only when adding interleaved resampling, in-between the emissions steps (in orange) is it clear that the curves actually represent the same statistical distribution. This beautifully illustrates the power of the resampling method. To avoid oversampling, the resampling is turned off for events for which the shower has terminated.

We stress that, while the usage of the weighted Sudakov algorithm in the above examples is completely unmotivated from a statistical perspective, since the default starting curve already has optimal statistics, the resampling procedure can equally well be applied in scenario with real weight problems, coming for example from applying the weighted Sudakov algorithm in a sub-leading N_c parton shower.

5 Conclusion and outlook

In this note we advocate to use the resampling method for drastically improving the statistical convergence for parton shower algorithms where weighted algorithms are used. We also stress that the resampling procedure can always be applied separately to only those weights which have been generated by the weighted shower algorithm, leaving weights from the hard process aside. Inclusive quantities are there-

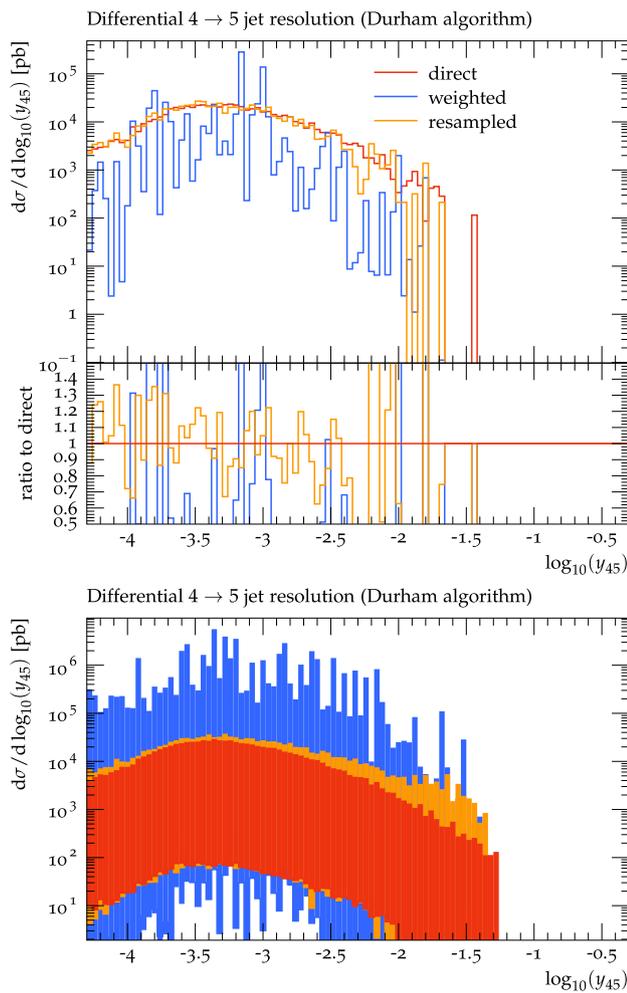


Fig. 3 The Durham y_{45} observable as implemented in [23] (default). The same observable when sampled with Algorithm 2 without resampling (no resampling), and when sampled with interleaved resampling after each emission or rejection step in Algorithm 2 (resampling). In the lower panel we show the range of variation as in Fig. 2

fore expected to be reproduced just as it would be the case for unweighted shower evolution.

The implementation tests, performed in the most simplistic way, illustrate beautifully the power of the basic resampling method, and prove its usefulness for parton showers. Nevertheless, we argue that further improvement could be achievable by various extensions of the algorithm to mitigate the ancestor depletion problem. One possible approach is to furnish the proposed resampling-based algorithm with an additional so-called *backward simulation pass* generating randomly new, non-collapsed ancestral paths on the stochastic grid formed by the events generated by the algorithm [36,37]. Another direction of improvement goes in the direction of parallelisation of the algorithm, which is essential in scenarios with large sample sizes. Parallelization of sequential Monte Carlo methods is non-trivial due to the “global” sample interaction imposed by the resampling operation.

Nevertheless, a possibility is to divide the full sample into a number of batches, or *islands*, handled by different processors, and to subject—“locally”—each island to sequential weighing and resampling.

In such a distributed system, statistical convergence may, following [38,39], may be improved further by interleaving the evolution of the distinct islands with *island-level resampling* operations, in which the islands are resampled according to their average weights.

At the more conceptual level, let us remark that the method we suggest represents a first step in a change of paradigm, where parton showers are viewed as tools for generating correct statistical distributions, rather than tools for simulating independent events. We believe that the most advanced and precise high energy physics simulations, if they should run in an efficient way, will have no chance to avoid weighted algorithms, and as such heavily need to rely on methods like the resampling method outlined in this note. We will further investigate the method, also including the possibility of making cancellations in higher order corrections explicit already in intermediate steps of such a Monte Carlo algorithm rather than by adding large weights at the very end. We will also address the structural challenges in implementing these methods in existing event generators which we hope will help to make design decisions, keeping in mind the necessity of methods like the one presented here.

Acknowledgements We thank Stefan Prestel for constructive feedback on the manuscript. SP is grateful for the kind hospitality of Mainz Institute for Theoretical Physics (MITP) of the DFG Cluster of Excellence PRISMA⁺ (project ID 39083149), where some of this work has been carried out and finalised. JO and MS thank the Erwin Schrödinger Institute for the kind hospitality during the PSR workshop. JO gratefully acknowledges support by the Swedish Research Council, Grant 2018-05230. MS was supported by the Swedish Research Council (contract number 2012-02744) as well as the European Union’s Horizon 2020 research and innovation programme (Grant agreement no. 668679). This work has also been supported in part by the European Union’s Horizon 2020 research and innovation programme as part of the Marie Skłodowska-Curie Innovative Training Network MCnetITN3 (Grant agreement no. 722104). SP acknowledges partial support by the COST actions CA16201 “PARTICLEFACE” and CA16108 “VBSCAN”.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors’ comment: Aside from the simulated data contained in the plots, this project generated no data.]

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>

[ons.org/licenses/by/4.0/](https://creativecommons.org/licenses/by/4.0/).
 Funded by SCOAP³.

Appendix A: Mathematical details on unbiasedness

In this section we show mathematically that the resampling operation does not introduce statistical bias. We will assume that all random variables are defined on a common probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and let \mathbb{E} denote expectations under \mathbb{P} . Now assume that $\{(\xi_i, w_i)\}_{i=1}^N$ is a sample of events with associated weights providing the estimator $\hat{h}_N = \sum_{i=1}^N w_i h(\xi_i)/N$ of the integral of h with respect to the dynamics governed by Eq. (1). In order to subject the sample to resampling in accordance with Sect. 2, let J be the set of indices $j \in \{1, \dots, N\}$ for which the corresponding event paths ξ_j have still not reached the cut-off scale. In other words, denoting by E the path-space region containing events evolving above the cut-off scale, it holds that $j \in J$ if and only if $\xi_j \in E$. At resampling, new events $\{\bar{\xi}_j\}_{j \in J}$ are drawn from the set $\{\bar{\xi}_j\}_{j \in J}$ of still evolving events with replacement and in proportion to the corresponding weights $\{w_j\}_{j \in J}$. Afterwards, the resampled events $\{\bar{\xi}_j\}_{j \in J}$ are assigned the common weight

$$\bar{w}_j = \frac{1}{M} \sum_{\ell \in J} w_\ell,$$

where $M = \text{card}(J)$, and \hat{h}_N is replaced by the estimator

$$\bar{h}_N = \frac{1}{N} \left(\sum_{i \in J^c} w_i h(\xi_i) + \sum_{i \in J} \bar{w}_i h(\bar{\xi}_i) \right). \tag{8}$$

Note that in Eq. (8), the weights of the non-evolving events $\{\xi_i\}_{i \in J^c}$ are unchanged. It now holds that

$$\mathbb{E}[\hat{h}_N] = \mathbb{E}[\bar{h}_N], \tag{9}$$

implying that resampling does not, as claimed in Sect. 2, bias the estimator. In order to show Eq. (9), first note that

$$\sum_{i \in J^c} w_i h(\xi_i) = \sum_{i=1}^N w_i h(\xi_i) \mathbb{1}_{E^c}(\xi_i). \tag{10}$$

Moreover, let $\mathcal{F} \subset \mathcal{A}$ be the σ -field generated by $\{(\xi_i, w_i)\}_{i=1}^N$; then by the tower property of conditional expectations, since $\{j \in J\} \in \mathcal{F}$ for all j and M is \mathcal{F} -measurable,

$$\mathbb{E} \left[\sum_{i \in J} \bar{w}_i h(\bar{\xi}_i) \right] = \mathbb{E} \left[\frac{1}{M} \sum_{\ell \in J} w_\ell \sum_{j \in J} \mathbb{E} [h(\bar{\xi}_j) \mid \mathcal{F}] \right].$$

Now, since $\{\bar{\xi}_j\}_{j \in J}$ are produced by means of resampling,

$$\mathbb{E} [h(\bar{\xi}_j) \mid \mathcal{F}] = \sum_{i \in J} \frac{w_i}{\sum_{\ell \in J} w_\ell} h(\xi_i),$$

where the right-hand side does not depend on j . Thus, combining the previous two identities yields

$$\begin{aligned} \mathbb{E} \left[\sum_{i \in J} \bar{w}_i h(\bar{\xi}_i) \right] &= \mathbb{E} \left[\frac{1}{M} \sum_{\ell \in J} w_\ell \times M \sum_{j \in J} \frac{w_j}{\sum_{\ell \in J} w_\ell} h(\bar{\xi}_j) \right] \\ &= \mathbb{E} \left[\sum_{j \in J} w_j h(\bar{\xi}_j) \right] = \mathbb{E} \left[\sum_{i=1}^N w_i h(\xi_i) \mathbb{1}_E(\xi_i) \right]. \end{aligned} \tag{11}$$

Now, Eqs. (10, 11) allow us to straightforwardly establish Eq. (9) according to

$$\begin{aligned} \mathbb{E}[\bar{h}_N] &= \frac{1}{N} \left(\mathbb{E} \left[\sum_{i=1}^N w_i h(\xi_i) \mathbb{1}_{E^c}(\xi_i) \right] + \mathbb{E} \left[\sum_{i=1}^N w_i h(\xi_i) \mathbb{1}_E(\xi_i) \right] \right) \\ &= \mathbb{E} \left[\frac{1}{N} \sum_{i=1}^N w_i h(\xi_i) \{ \mathbb{1}_{E^c}(\xi_i) + \mathbb{1}_E(\xi_i) \} \right] = \mathbb{E}[\hat{h}_N]. \end{aligned}$$

Finally, we remark that the previous calculation holds generally true for arbitrary subsets E of the path space, and not only in the case where E contains events evolving above the cut-off scale.

References

1. M.H. Seymour, Matrix element corrections to parton shower algorithms. *Comput. Phys. Commun.* **90**, 95–101 (1995). [arXiv:hep-ph/9410414](https://arxiv.org/abs/hep-ph/9410414)
2. S. Plätzer, M. Sjö Dahl, The Sudakov veto algorithm reloaded. *Eur. Phys. J. Plus* **127**, 26 (2012). [arXiv:1108.6180](https://arxiv.org/abs/1108.6180)
3. A. Buckley et al., General-purpose event generators for LHC physics. *Phys. Rep.* **504**, 145–233 (2011). [arXiv:1101.2599](https://arxiv.org/abs/1101.2599)
4. S. Plätzer, ExSample: a library for sampling Sudakov-type distributions. *Eur. Phys. J. C* **72**, 1929 (2012). [arXiv:1108.6182](https://arxiv.org/abs/1108.6182)
5. L. Lönnblad, Fooling around with the Sudakov veto algorithm. *Eur. Phys. J. C* **73**(3), 2350 (2013). [arXiv:1211.7204](https://arxiv.org/abs/1211.7204)
6. S. Hoeche, F. Krauss, M. Schonherr, F. Siegert, A critical appraisal of NLO + PS matching methods. *JHEP* **09**, 049 (2012). [arXiv:1111.1220](https://arxiv.org/abs/1111.1220)
7. J. Bellm, S. Plätzer, P. Richardson, A. Siódmok, S. Webster, Reweighting parton showers. *Phys. Rev. D* **94**(3), 034028 (2016). [arXiv:1605.0825](https://arxiv.org/abs/1605.0825)
8. S. Mrenna, P. Skands, Automated parton-shower variations in Pythia 8. *Phys. Rev. D* **94**(7), 074005 (2016). [arXiv:1605.0835](https://arxiv.org/abs/1605.0835)
9. E. Bothmann, M. Schön herr, S. Schumann, Reweighting QCD matrix-element and parton-shower calculations. *Eur. Phys. J. C* **76**(11), 590 (2016). [arXiv:1606.0875](https://arxiv.org/abs/1606.0875)
10. S. Plätzer, M. Sjö Dahl, J. Thorén, Color matrix element corrections for parton showers. *JHEP* **11**, 009 (2018). [arXiv:1808.0033](https://arxiv.org/abs/1808.0033)
11. H.T. Li, P. Skands, A framework for second-order parton showers. *Phys. Lett. B* **771**, 59–66 (2017). [arXiv:1611.0001](https://arxiv.org/abs/1611.0001)
12. S. Höche, S. Prestel, Triple collinear emissions in parton showers. *Phys. Rev. D* **96**(7), 074017 (2017). [arXiv:1705.0074](https://arxiv.org/abs/1705.0074)
13. F. Dulat, S. Höche, S. Prestel, Leading-color fully differential two-loop soft corrections to QCD dipole showers. *Phys. Rev. D* **98**(7), 074013 (2018). [arXiv:1805.0375](https://arxiv.org/abs/1805.0375)

14. S. Plätzer, M. Sjö Dahl, Subleading N_c improved parton showers. *JHEP* **1207**, 042 (2012). [arXiv:1201.0260](#)
15. Z. Nagy, D.E. Soper, Parton shower evolution with subleading color. *JHEP* **06**, 044 (2012). [arXiv:1202.4496](#)
16. Z. Nagy, D.E. Soper, Effects of subleading color in a parton shower. *JHEP* **07**, 119 (2015). [arXiv:1501.0077](#)
17. S. Plätzer, Summing large- N towers in colour flow evolution. *Eur. Phys. J. C* **74**(6), 2907 (2014). [arXiv:1312.2448](#)
18. J. Isaacson, S. Prestel, On stochastically sampling color configurations. [arXiv:1806.1010](#)
19. R. Á. Martínez, M. De Angelis, J.R. Forshaw, S. Plätzer, M.H. Seymour, Soft gluon evolution and non-global logarithms. [arXiv:1802.0853](#)
20. J.R. Forshaw, J. Holguin, S. Plätzer, Parton branching at amplitude level. *JHEP* **08**, 145 (2019). [arXiv:1905.0868](#)
21. D.B. Rubin, A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when the fraction of missing information is modest: the SIR algorithm (discussion of Tanner and Wong). *J. Am. Stat. Assoc.* **82**, 543–546 (1987)
22. N. Gordon, D. Salmond, A.F. Smith, Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc. F Radar Signal Process.* **140**, 107–113 (1993)
23. S. Höche, Tutorial on Parton Showers, CTEQ Summer School 2015, unpublished. <http://www.slac.stanford.edu/~shoeche/cteq15/ws/PS.pdf>. Accessed 26 Nov 2018
24. K. Kroeninger, S. Schumann, B. Willenberg, 3—A multi-channel Markov chain Monte Carlo algorithm for phase-space sampling. *Comput. Phys. Commun.* **186**, 1–10 (2015). [arXiv:1404.4328](#)
25. D.B. Rubin, Using the SIR algorithm to simulate posterior distribution, in *Bayesian Statistics 3*, ed. by J.M. Bernardo, M. DeGroot, D. Lindley, A. Smith (Clarendon Press, New York, 1988), pp. 395–402
26. J. Handschin, D. Mayne, Monte Carlo techniques to estimate the conditional expectation in multi-stage non-linear filtering. *Int. J. Control* **9**, 547–559 (1969)
27. P. Del Moral, A. Guionnet, On the stability of interacting processes with applications to filtering and genetic algorithms. *Annales de l'Institut Henri Poincaré* **37**, 155–194 (2001)
28. A. Doucet, N. De Freitas, N. Gordon (eds.), *Sequential Monte Carlo Methods in Practice* (Springer, New York, 2001)
29. L. Devroye, *Non-uniform Random Variate Generation* (Springer, Berlin, 1986)
30. P.E. Jacob, L.M. Murray, S. Rubenthaler, Path storage in the particle filter. *Stat. Comput.* **25**(2), 487–496 (2015)
31. A. Kong, J.S. Liu, W. Wong, Sequential imputation and Bayesian missing data problems. *J. Am. Stat. Assoc.* **89**, 590–599 (1994)
32. J. Cornebise, E. Moulines, J. Olsson, Adaptive methods for sequential importance sampling with application to state space models. *Stat. Comput.* **18**(4), 461–480 (2008)
33. H.P. Chan, T.L. Lai, A general theory of particle filters in hidden Markov models and some applications. *Ann. Stat.* **41**(6), 2877–2904 (2013)
34. A. Lee, N. Whiteley, Variance estimation in particle filters. *Biometrika* **105**(3), 609–625 (2018)
35. S. Catani, Y.L. Dokshitzer, M. Olsson, G. Turnock, B.R. Webber, New clustering algorithm for multi-jet cross-sections in e^+e^- annihilation. *Phys. Lett. B* **269**, 432–438 (1991)
36. S.J. Godsill, A. Doucet, M. West, Monte Carlo smoothing for nonlinear time series. *J. Am. Stat. Assoc.* **50**, 438–449 (2004)
37. R. Douc, A. Garivier, E. Moulines, J. Olsson, Sequential Monte Carlo smoothing for general state space hidden Markov models. *Ann. Appl. Probab.* **21**(6), 1201–1245 (2011)
38. C. Vergé, C. Dubarry, P. Del Moral, E. Moulines, On parallel implementation of sequential Monte Carlo methods: the island particle model. *Stat. Comput.* **25**(2), 243–260 (2015)
39. P. Del Moral, E. Moulines, J. Olsson, C. Vergé, Convergence properties of weighted particle islands with application to the double bootstrap algorithm. *Stoch. Syst.* **2**(6), 367–419 (2016)