Resampling Algorithms for High Energy Physics Simulations

Jimmy Olsson

KTH Royal Institute of Technology
Stockholm, Sweden

In collaboration with S. Plätzer and M. Sjödahl

PSR19, ESI, Vienna
12 June 2019

Aim

- Let
  - $\{(\mathsf{X}_n, \mathcal{X}_n)\}_{n \in \mathbb{N}}$ be *measurable spaces*;
  - $\{Q_n\}_{n \in \mathbb{N}}$ *transition kernels*;
  - γ_0 be a *measure* on $(\mathsf{X}_0, \mathcal{X}_0)$.
- For each $n \in \mathbb{N}$, define the measure

$$\gamma_{0:n}(dx_{0:n}) := \gamma_0(dx_0) \prod_{m=0}^{n-1} Q_m(x_m, dx_{m+1})$$

  on the *path space* $\mathsf{X}_{0:n} := \mathsf{X}_0 \times \cdots \times \mathsf{X}_n$.
- We aim at forming *weighted samples* $\{(\xi_{0:n}^i, \omega_n^i)\}_{i=1}^N$, $n \in \mathbb{N}$, targeting the sequence $\{\gamma_{0:n}\}_{n \in \mathbb{N}}$ in the sense that for all n and h,

$$\frac{1}{N} \sum_{i=1}^{N} \omega_n^i h(\xi_{0:n}^i) \simeq \gamma_{0:n} h := \int h(x_{0:n}) \, \gamma_{0:n}(dx_{0:n}).$$

Sequential importance sampling (SIS)

- Introduce
  - *Markov kernels* $\{P_n\}_{n \in \mathbb{N}}$ such that $Q_n \ll P_n$;
  - an initial distribution ρ on $(\mathsf{X}_0, \mathcal{X}_0)$.
- Then

$$\gamma_{0:n}(dx_{0:n}) = \frac{d\gamma_0}{d\rho}(x_0) \prod_{m=0}^{n-1} \frac{dQ_m(x_m, \cdot)}{dP_m(x_m, \cdot)}(x_{m+1}) \, \mathbb{P}_\rho(dx_{0:n}),$$

where $\mathbb{P}_\rho(dx_{0:n}) := \rho(dx_0) \prod_{m=0}^{n-1} P_m(x_m, dx_{m+1})$.

- Thus, naively, simulate each $\xi^i_{0:n}$ from $\mathbb{P}_\rho(dx_{0:n})$ and compute

$$\omega^i_n = \frac{d\gamma_0}{d\rho}(\xi^i_0) \prod_{m=0}^{n-1} \frac{dQ_m(\xi^i_m, \cdot)}{dP_m(\xi^i_m, \cdot)}(\xi^i_{m+1}).$$

SIS (cont.)

- Recursive updates $\{(\xi^i_{0:n}, \omega^i_n)\}_{i=1}^N \rightsquigarrow \{(\xi^i_{0:n+1}, \omega^i_{n+1})\}_{i=1}^N$:

for $i \leftarrow 1$ **to** N **do**

  draw $\xi^i_{n+1} \sim P_n(\xi^i_n, dx_{n+1})$;

  set $\xi^i_{0:n+1} \leftarrow (\xi^i_{0:n}, \xi^i_{n+1})$;

  set $\omega^i_{n+1} \leftarrow \omega^i_n \frac{dQ_n(\xi^i_n, \cdot)}{dP_n(\xi^i_n, \cdot)}(\xi^i_{n+1})$;

end

Example: parton shower simulation

- Given a starting scale Q, the scale q of the next emission and some additional splitting variables z have distribution

$$S_P(dq, dz|Q) := \Delta_P(\mu|Q)\delta_{(\mu,z_\mu)}(dq, dz) \\ + P(q, z)\Delta_P(q|Q)\mathbb{1}_{(\mu,Q)}(q)\, dq\, dz,$$

where

  - P is a splitting kernel;
  - μ is the infrared cutoff;
  - z_μ is a parameter point associated with the cutoff;
  - $\Delta_P(q|Q)$ is the *Sudakov form factor*

$$\Delta_P(q|Q) := \exp\left(-\int_q^Q dp \int dz\, P(p, z)\right).$$

- Our goal is to simulate the process

$$Q \overset{S_P}{\rightsquigarrow} (\tilde{q}_1, \tilde{z}_1) \overset{S_P}{\rightsquigarrow} (\tilde{q}_2, \tilde{z}_2) \overset{S_P}{\rightsquigarrow} (\tilde{q}_3, \tilde{z}_3) \overset{S_P}{\rightsquigarrow} \cdots$$

Weighted Sudakov veto (WSV) [Bel+16]

Data: $(q', z', \varepsilon'), \omega'$
Result: $(q, z, \varepsilon), \omega$
draw $(q, z) \sim S_R(dq, dz|q')$;
if $(q, z) = (\mu, z_\mu)$ **then**
  $\varepsilon \leftarrow 0$;
  $\omega \leftarrow \omega'$;
end
else
  draw $\varepsilon \sim \mathrm{Be}(\pi(q, z|q', z'))$;
  if $\varepsilon = 1$ **then**
    set $\omega \leftarrow \omega' \frac{1}{\pi(q,z|q',z')} \frac{P(q,z)}{R(q,z)}$;
  end
  else
    set $\omega \leftarrow \omega' \frac{1}{1-\pi(q,z|q',z')} \left(1 - \frac{P(q,z)}{R(q,z)}\right)$
  end
end

WSV as SIS

- Iterating the previous algorithm yields a Markovian dynamical process $\{(q_m, z_m, \varepsilon_m, \omega_m)\}_{m\in\mathbb{N}^*}$ with $q_0 = Q$ and $\omega_0 = 1$. We denote its law by $\mathbb{P}_Q$.
- Set $n_0 = 0$ and define recursively

$$n_\ell := \min\{m > n_{\ell-1} : \varepsilon_m = 1 \text{ or } q_m = \mu\}, \quad \ell > 1.$$

- By [Bel+16, Section 2.3] and the strong Markov property, with $\tilde{q}_0 = Q$,

$$\mathbb{E}_Q\left(h_p(q_{n_1}, z_{n_1}, q_{n_2}, z_{n_2}, \ldots, q_{n_p}, z_{n_p})\,\omega_{n_p}\right) \tag{1}$$
$$= \int \cdots \int h_p(\tilde{q}_1, \tilde{z}_1, \ldots, \tilde{q}_p, \tilde{z}_p) \prod_{\ell=0}^{p-1} S_P(d\tilde{q}_{\ell+1}, d\tilde{z}_{\ell+1}|\tilde{q}_\ell),$$

- Consequently, WSV can be expressed as SIS.

WSV output for the Durham y_{45} observable

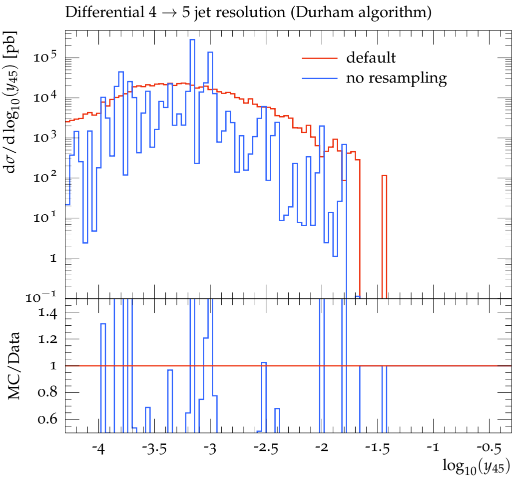


Figure: The Durham y_{45} observable as implemented in [Höc15] (default). The same observable when sampled with WSV (no resampling).

Weight degeneracy in WSV

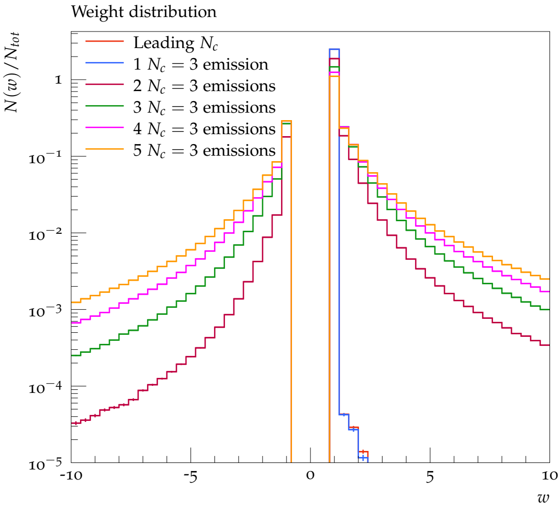
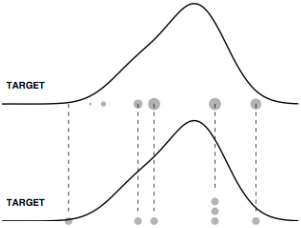


Figure: WSV weight distribution using $1, \ldots, 5$ color-corrected emissions in the `Herwig` dipole shower.

Multinomial selection [Rub87]



- Weight degeneracy is a *universal problem* of the SIS algorithm.
- Remedy: transform, at each SIS iteration, $\{(\xi^i_{0:n}, \omega^i_n)\}_{i=1}^N$ into a *uniformly weighted* sample $\{(\tilde{\xi}^i_{0:n}, \tilde{\omega}^i_n)\}_{i=1}^N$ by selecting $\{\tilde{\xi}^i_{0:n}\}_{i=1}^N$ among $\{\xi^i_{0:n}\}_{i=1}^N$ with replacement and in proportion to $\{\omega^i_n\}_{i=1}^N$ and letting

$$\tilde{\omega}^1_n = \tilde{\omega}^2_n = \cdots = \tilde{\omega}^N_n = \frac{\Omega_n}{N},$$

with $\Omega_n = \sum_{i=1}^N \omega^i_n$.

SIS with resampling (SISR) [GSS93]

Data: $\{(\xi^i_{0:n|n}, \omega^i_n)\}_{i=1}^N$
Result: $\{(\xi^i_{0:n+1|n}, \omega^i_{n+1})\}_{i=1}^N$
for $i \leftarrow 1$ **to** N **do**
  $\iota^i \sim \sum_{i=1}^N \frac{\omega^i_n}{\Omega_n} \delta_i$;
  `/* selection */`
  draw $\xi^i_{n+1|n+1} \sim P_n(\xi^{\iota^i}_{n|n}, dx_{n+1})$;
  set $\xi^i_{0:n+1|n+1} \leftarrow (\xi^{\iota^i}_{0:n|n}, \xi^i_{n+1|n+1})$;
end
for $i \leftarrow 1$ **to** N **do**
  set $\omega^i_{n+1} \leftarrow \frac{dQ_n(\xi^i_{n|n+1}, \cdot)}{dP_n(\xi^i_{n|n+1}, \cdot)}(\xi^i_{n+1|n+1}) \frac{\Omega_n}{N}$;
end

Example: WSV with resampling (WSVR)

- To cope with the weight degeneracy of WSV, furnish the same with interleaved resampling operations $\Rightarrow$ WSVR.
- Different possible resampling schedules:
  - Systematically after each WSV iteration.
  - Only at emissions (embedded chain).
  - Only when indicated by some weight skewness criterion, such as the *coefficient of variation*

$$\mathrm{CV}^2 = N \sum_{i=1}^{N} \left(\frac{\omega_n^i}{\Omega_n} - \frac{1}{N} \right)^2$$

or the *efficient sample size*

$$\mathrm{ESS} = \frac{N}{1 + \mathrm{CV}^2}.$$

WSVR output for the Durham y_{45} observable

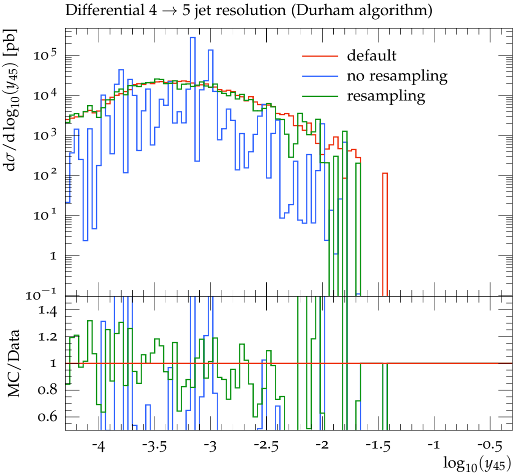


Figure: The Durham y_{45} observable as implemented in [Höc15] (default). The same observable when sampled with WSV without resampling (no resampling) and when sampled with WSVR with systematic resampling (resampling). The plot is based on 10,000 events.

Variability estimates

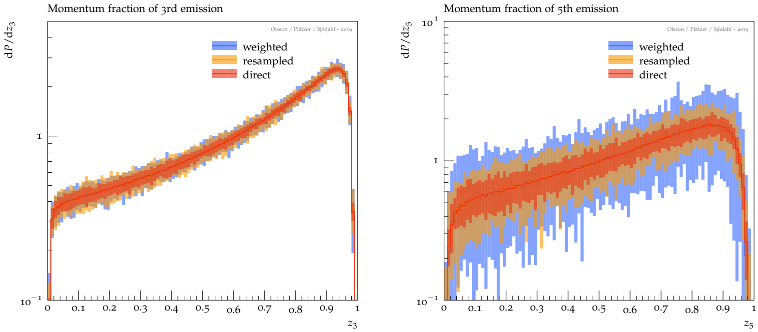


Figure: The left and right panels display momentum fractions after 2 and 5 emissions, respectively. Blue and yellow boxes correspond to WSV and WSVR, respectively.

Some theoretical properties of SISR

- The following can be established; see e.g. [Del04; DMG99].
  - For all n and h,

$$\mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N}\omega_n^i h(\xi_{0:n|n}^i)\right) = \gamma_{0:n}h.$$

  - For all n and $h \in \mathsf{L}_2(\gamma_{0:n})$ there exists $\sigma_n^2(h)$ such that, as $N \to \infty$,

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}\omega_n^i h(\xi_{0:n|n}^i) - \gamma_{0:n}h\right) \stackrel{\mathcal{L}}{\longrightarrow} N(0, \sigma_n^2(h)).$$

- For objective functions h depending on x_n only, the sequence $\{\sigma_n^2(h)\}_{n\in\mathbb{N}}$ can, under mild ergodicity assumptions, be bounded *uniformly* in n [DMO14]. For SIS it increases exponentially.

Conclusion and future work

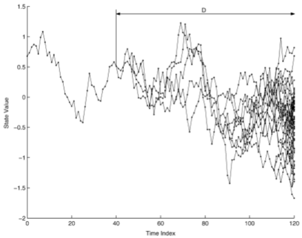


Figure: The particle path degeneracy phenomenon.

- "Change of philosophy": independent events vs. computationally efficient tool for reproducing observables.
- Particle path degeneracy phenomenon when n is large compared to N. Apply *backward-sampling* techniques [GDW04].
- Towards parallel implementations using *particle island models* [Ver+15].

References I

[Bel+16] J. Bellm et al. “Reweighting Parton Showers”. In: *Phys. Rev.* D94.3 (2016), p. 034028.

[Del04] P. Del Moral. *Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications*. Springer, 2004.

[DMG99] P. Del Moral and A. Guionnet. “Central limit theorem for nonlinear filtering and interacting particle systems”. In: *Ann. Appl. Probab.* 9.2 (1999), pp. 275–297.

[DMO14] R. Douc, E. Moulines, and J. Olsson. “Long-term stability of sequential Monte Carlo methods under verifiable conditions”. In: *Ann. Appl. Probab.* 24.5 (2014), pp. 1767–1802.

[GDW04] S. J. Godsill, A. Doucet, and M. West. “Monte Carlo smoothing for non-linear time series”. In: *J. Am. Statist. Assoc.* 50 (2004), pp. 438–449.

References II

[GSS93] N. Gordon, D. Salmond, and A. F. Smith. "Novel approach to nonlinear/non-Gaussian Bayesian state estimation". In: *IEE Proc. F, Radar Signal Process.* 140 (1993), pp. 107–113.

[Höc15] S. Höche. *Tutorial on Parton Showers, CTEQ Summer School 2015*. Unpublished. 2015.

[Rub87] 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)". In: *J. Am. Statist. Assoc.* 82 (1987), pp. 543–546.

[Ver+15] C. Vergé et al. "On parallel implementation of sequential Monte Carlo methods: the island particle model". In: *Statistics and Computing* 25.2 (2015), pp. 243–260.

Proof of (1)

- We assume that the function π is such that for all m, $\mathbb{P}_Q$-a.s., $n_m < \infty$.
- By the functional monotone class theorem it is enough to establish (1) for $h_p(\tilde{q}_{0:p}, \tilde{z}_{0:p}) = \prod_{\ell=0}^{p} g_\ell(\tilde{q}_\ell, \tilde{z}_\ell)$.
- Proceed by induction and assume that (1) holds true for $p-1$.
- Define the weight function

$$\Psi(q, z, \varepsilon | q', z') := \mathbb{1}_{\{1\}}(\varepsilon) \frac{1}{\pi(q, z | q', z')} \frac{P(q, z)}{R(q, z)} + \mathbb{1}_{\{0\} \times \{\mu\}^{\complement}}(\varepsilon, q) \frac{1}{1 - \pi(q, z | q', z')} \left(1 - \frac{P(q, z)}{R(q, z)}\right) + \mathbb{1}_{\{\mu\}}(q).$$

Proof of (1) (cont.)

- By the tower property followed by the strong Markov property,

$$
\begin{aligned}
&\mathbb{E}_Q\left(h_p(q_{n_1}, z_{n_1}, \ldots, q_{n_p}, z_{n_p})\,\omega_{n_p}\right) \\
&= \mathbb{E}_Q\Bigg(h_{p-1}(q_{n_1}, z_{n_1}, \ldots, q_{n_{p-1}}, z_{n_{p-1}})\,\omega_{n_{p-1}} \\
&\times \mathbb{E}_Q\Bigg(g(q_{n_p}, z_{n_p}) \prod_{m=n_{p-1}+1}^{n_p} \Psi(q_m, z_m, \varepsilon_m | q_{m-1}, z_{m-1}) \mid \mathcal{F}_{n_{p-1}}\Bigg)\Bigg) \\
&= \mathbb{E}_Q\Bigg(h_{p-1}(q_{n_1}, z_{n_1}, \ldots, q_{n_{p-1}}, z_{n_{p-1}})\,\omega_{n_{p-1}} \qquad (2) \\
&\times \mathbb{E}_{q_{n_{p-1}}}\Bigg(g(q_{n_1}, z_{n_1}) \prod_{m=1}^{n_1} \Psi(q_m, z_m, \varepsilon_m | q_{m-1}, z_{m-1})\Bigg)\Bigg).
\end{aligned}
$$

Proof of (1) (cont.)

- Now, using the main result of [Bel+16, Section 2.3],

$$\mathbb{E}_{q_{n_{p-1}}}\left(g(q_{n_1}, z_{n_1}) \prod_{m=1}^{n_1} \Psi(q_m, z_m, \varepsilon_m | q_{m-1}, z_{m-1})\right) \\ = \int g(\tilde{q}_p, \tilde{z}_p)\, S_P(d\tilde{q}_p, d\tilde{z}_p | q_{n_{p-1}}).$$

- Plugging the previous identity into (2) and applying the induction hypothesis yields

$$\mathbb{E}_Q\left(h_p(q_{n_1}, z_{n_1}, \ldots, q_{n_p}, z_{n_p})\, \omega_{n_p}\right) \\ = \int \cdots \int h_p(\tilde{q}_1, \tilde{z}_1, \ldots, \tilde{q}_p, \tilde{z}_p) \prod_{\ell=0}^{p-1} S_P(d\tilde{q}_{\ell+1}, d\tilde{z}_{\ell+1} | \tilde{q}_\ell)$$

(with $\tilde{q}_0 = Q$), which was to be established.