Dear Editor,

I am sorry to report that I am not convinced by the answers of the authors. Out of my four main points, two have not been answered convincingly, and another one has been fully ignored. It seems that the authors did not fully understand the points. Consequently, my referee report is conceptually the same as the first one and my recommendation of asking another revision as well. In this second report, I tried to be more specific to make sure that the authors do not miss the point.

1. Like in my previous report, I do agree with the authors that their algorithm holds for any numerical value of w_{max}. I further agree with them that too small values will lead to a poor performance. And I also agree that for a well-trained surrogate, this should be manageable.

   However, the authors claim that their statement concerning the effective sample size holds. I do not agree with that statement. For example, take the following case:

   $$\max_i(w_i) < w_{\text{max}} < \max_i(s_i), \text{ and } x_{\text{max}} > \max_i(x_i). \tag{1}$$

   In that case, standard one step un-weighting (using the same value of w_{max}) would not produce any over-weight. The two stage un-weighting would have some over-weight events (only from the first stage), but Eq. (6) claims that

   $$N_{\text{eff}} = N, \qquad \alpha = 1, \tag{2}$$

   as if all events were correctly un-weighted. Eq. (6) misses to include the over-weight related to the first un-weighting. The author should update that metric to include both source of overweight. Following their approach, I would suggest to use

   $$\alpha = \frac{1}{N}\frac{(\sum \tilde{w})^2}{\sum \tilde{w}^2}, \tag{3}$$

   which is the same as

   $$\alpha = \frac{1}{N}\frac{\left(\sum \max(1, s_i/w_{\text{max}}) \max(1, x_i/x_{\text{max}})\right)^2}{\sum \left(\max(1, s_i/w_{\text{max}}) \max(1, x_i/x_{\text{max}})\right)^2} \tag{4}$$

   or (if they want to compare to the standard over-weight allowed by AMEGIC algorithm):

   $$\alpha = \frac{\left(\sum \max(1, s_i/w_{\text{max}}) \max(1, x_i/x_{\text{max}})\right)^2}{\sum \left(\max(1, s_i/w_{\text{max}}) \max(1, x_i/x_{\text{max}})\right)^2} \frac{\sum \max(1, w_i/w_{\text{max}})^2}{\left(\sum \max(1, w_i/w_{\text{max}})\right)^2}. \tag{5}$$

   To be clear, those are suggestion of formula, I am open to any metric that does include both sources of overweight.

   As pointed by the authors in their answer, the additional Figure (3b) and (6b) indicates that $\max(s_i) < \max(w_i)$. Therefore, one can guess that the change of definition of α will not (or barely) impact the rest of the paper (if at all). Obviously, what is true in those examples does not mean it will be true for all cases, and the algorithm section should define a metric that takes into account all the sources of overweight (or at least all new sources of over-weights compared to the previous method).

   In their answer, the author claims that using w_{max} rather than s_{max}, (which I understand as using the distribution of weight w_i rather than s_i for the determination of median/pm value) they gain in performance at a cost of some additional over-weight. From the plots/tables of the paper, it is difficult to see why. If the authors have additional elements to support that statement, I would be interested to see them, but this is certainly not mandatory and not needed to be included in the paper.

2. I do believe that the authors fully miss this second point. So let me rephrase it. My remark here was why the author use Equation (5):

$$\tilde{w} = \max\left(1, \frac{x}{x_{\text{max}}}\right) \cdot \max\left(1, \frac{s}{w_{\text{max}}}\right) \tag{6}$$

and not the formula used in the literature for multiple stage un-weighting (for example see alpgen paper: hep-ph/0206293) –formula adapted to use authors convention–.

$$\tilde{w} = \max\left(1, \frac{x}{x_{\text{max}}} \max\left(1, \frac{s}{w_{\text{max}}}\right)\right) \tag{7}$$

As said in my previous report, it would be appropriate that the authors compare the two formula or comment why they do not think that the formula used in other code/context is relevant.

3. Here I failed to fully understand the answer of the authors, so let me be more specific in my issue/question.

If I look at their Eq. (12), the authors do include the Jacobian within the function learned by the Neural-Network (as they should do). However, that Jacobian depends not only on the four-momenta but also of the channel of integration used to generate that particular four-momenta. So I would expect that one need to train one surrogate for each channel of integration.

I would suggest that the author stress explicitly, if they train a single surrogate (my understanding of the paper) or one per channel of integration. If they use a single surrogate for all channels of integration, then it will be important to include a comment on the non-unicity of the Jacobian/weight for a given four-momenta and on the impact/limitation of that choice (or, in case, how that issue is avoided).

4. Sorry for the miss-understanding here. This fully answers my concern.

minor points

5. If you look at this paper: hep-ph/0006269, you will see that it presents the VEGAS algorithm as learning a surrogate function. In that paper, it is also presented that importance sampling is nothing else than a change of variable from a function on which you are able to generate events according to its density.

You can indeed see/present your algorithm as an extension of the hit-or-miss algorithm, but it also is an importance sampling method where the surrogate is learned by a Neural-Network and the density is obtained by hit or miss.

I do believe that mentioning, in the paper, that your method is an importance sampling method and that your method extends what VEGAS does, makes sense. But, as said, in the previous report, this is a minor point and I will not insist more on it any further.

6. Thanks for the modification.

7. I would suggest the authors to be more precise in their footnote. As they certainly know, a lot of non-expert are simply multiplying the weight by the generated cross-section, which is not correct in presence of over-weight. If the authors can provide the exact formula/method to use in presence of over-weights, this might help to limit the problem.

8. That's good to know.

9. Thanks to provide this reference. I would suggest to the authors to add that reference into the paper.