A method for computing cumulants in relativistic heavy-ion collisions

Willian M Serenone[1], Frédérique Grassi[1], Renan Hirayama[2], Jean-Yves Ollitrault[3]

[1] Instituto de Fìsica, Universidade de São Paulo, Rua do Matão 1371, 05508-090 São Paulo-SP, Brazil
[2] Frankfurt Institute for Advanced Studies, Ruth-Moufang-Strasse 1, 60438 Frankfurt am Main, Germany
[3] Université Paris Saclay, CNRS, CEA, Institut de physique théorique, 91191 Gif-sur-Yvette, France

E-mail: wmatioli@if.usp.br

Abstract. To map QCD phase diagram is one of the greatest challenges for those who study QCD at finite temperature. A class of observables of particular interest in this study are the net-baryon cumulants. However, hydrodynamic-based hybrid models have trouble in computing higher order cumulants due to the high statistics needed. The usual oversampling procedure to increase statistics in such models does not help in this situation because it smears event-by-event fluctuations. We propose a way to compute the cumulants avoiding this issue and show its applications at LHC energies.

1. Introduction

Event-by-event fluctuations of conserved charges provide a direct probe for the behavior of the transition between the deconfined quark-gluon phase and the hadronic matter observed in heavy-ion collisions. In particular, susceptibilities will diverge if a critical point is present in the QCD phase diagram [1, 2]. Useful observables to quantify these fluctuations are the *cumulants* of the conserved charges, which are proportional to the susceptibilities. Experimentally, cumulants of net-protons are typically used as a proxy for the baryon density cumulants [3, 4] and were measured at both RHIC [3] and LHC [5].

Much like in experiments, obtaining precise values for these cumulants in heavy-ion collision simulations requires a very large sample of events, which poses a computational challenge for hydrodynamics-based simulations. On the other hand, particlization — the name given to the switching from the hydrodynamic prescription to the hadron gas one — is comparatively cheaper and stochastic in nature. It is common to use the random nature of the procedure to perform the particlization many times over a single hydrodynamic simulation (see, e.g. [6, 7]), a procedure usually called oversampling. In this way, one can achieve the same statistic level as experiments within a reasonable computational time. For works that are mainly interested in mean values this procedure has no drawbacks. However, it tends to underestimate fluctuations coming from particlization, severely compromising cumulants calculations.

In this work, we show a way to correctly account for the oversampling in the particlization procedure. This is the same two-step averaging procedure shown in [8]. Here, we also present a

toy model where we parameterize the probability of emitting particles during the particlization using an analytical function. The parameters of this function are dependent on the initial condition and can be numerically estimated for each event from the oversampling procedure. At the same time, an analytical value for the cumulants can also be computed. Therefore, we use it to benchmark the two-step averaging procedure. We show that one can obtain precise values for the cumulants with statistics which are easily available in a typical simulation. We also show the results applied in a boost-invariant simulation at LHC energies.

2. Cumulants and moments

As mentioned, the net-charge fluctuations are codified in the corresponding observable cumulants of the event-by-event probability distribution. They are derived analytically as the Maclaurin coefficients for the *Cumulant Generating Function* (CGF) $K^X(z)$. However, expanding it directly may lead to cumbersome expressions. Instead, one can use the *moments* μ_n, which can later be related to the cumulants. The momentum are defined from the *Moment Generating Function* (MGF)

$$M^X(z) \equiv \left\langle e^{zX} \right\rangle = 1 + \sum_{n=1}^{\infty} \mu_n^X \frac{z^n}{n!}. \tag{1}$$

Direct derivation of (1) shows that $\mu_n^X = \langle X^n \rangle$. The relation between cumulants and momentum are established by the CGF

$$K^X(z) = \ln M^X(z)\,. \tag{2}$$

The higher the order of the cumulants, the higher the necessary statistics are needed to accurately determine it. For hydrodynamic-based hybrid models this can be an issue, since hydrodynamic simulations are computationally expensive. As was also mentioned in Section 1, the typical solution of oversampling particles during the particlization leads for fluctuations being washed away. Naively applying the procedure outlined above would underestimate cumulants other than C_1.

2.1. Two-step averaging

The way we propose to account for the oversampling procedure is to split the calculation of the averages in Eq. (1) in two steps: the innermost (FO) runs over the different samples while the outermost (IC) contains loops over the different initial conditions. The event-by-event MGF becomes

$$M^X(z) = \left\langle e^{zX} \right\rangle = \left\langle \left\langle e^{zX} \right\rangle_{FO} \right\rangle_{IC} = \left\langle M_{FO}^X(z) \right\rangle_{IC}. \tag{3}$$

With this splitting, the moments μ_n in (1) can be computed as

$$\mu_n = \left.\frac{\mathrm{d}^n M}{\mathrm{d}z^n}\right|_{z=0} = \left\langle \left.\frac{\mathrm{d}^n M_{FO}}{\mathrm{d}z^n}\right|_{z=0} \right\rangle_{IC}. \tag{4}$$

Eq. (4) is the basis of the method developed in this work. Throughout this work we will refer to the procedure to correct cumulants and moments as two-step averaging.

2.2. Poisson parametrization of emitted particles

If the particlization distribution of particles is parametrized or known analytically, the computation of two-step averaging in a simulation is reduced to an averaging over the parameters specific to each IC, by directly computing $M_{FO}(z)$. We will assume the simplest case, which is an independent emission of particles resulting in a Poisson distribution. In this scenario,

$$X \sim \mathrm{Pois}(\lambda): \qquad M_{FO}^X(z) = \exp\left(\lambda e^z - \lambda\right), \tag{5}$$

where the parameter λ can be estimated as $\lambda = \langle X \rangle_{FO}$. A useful property of a MGF is that $M^{X\pm Y}(z) = M^X(z)M^Y(\pm z)$, so for the net-particle $X - \bar{X}$, with $\bar{X} \sim \text{Pois}(\bar{\lambda})$, it becomes

$$M_{FO}^{X-\bar{X}}(z) = \exp\left[\lambda(e^z - 1) + \bar{\lambda}(e^{-z} - 1)\right]. \tag{6}$$

The two-step averaging are straightforwardly computed applying (6) into (4). We show, as an example, the first four moment calculated explicitly

$$\mu_1^{X-\bar{X}} = \langle q \rangle_{IC} \,, \tag{7}$$
$$\mu_2^{X-\bar{X}} = \left\langle n + q^2 \right\rangle_{IC} \,, \tag{8}$$
$$\mu_3^{X-\bar{X}} = \left\langle q + 3nq + q^3 \right\rangle_{IC} \,, \tag{9}$$
$$\mu_4^{X-\bar{X}} = \left\langle n + q^2 + 6n^2 + 6nq^2 + q^4 \right\rangle_{IC} \,. \tag{10}$$

where we introduced the shorthand $q = \lambda - \bar{\lambda}$ and $n = \lambda + \bar{\lambda}$. The event-by-event cumulants can then be determined by inserting the two-step averaging (7)-(10) into the usual relations between moments and cumulants.

3. Results

We will apply the calculations proposed above in the simple scenario of a toy model. In this scenario the probability density function is completely known and thus no hypotheses is necessary for it. More importantly, it allow us to generate a high amount of statistics without the need to run the computationally intensive task of hydrodynamic simulations and compare the results with a semi-analytical approach as well as the situation where no oversampling is performed. It also serves as a way to validate our analysis tools.

We then proceed to show the results for a real-life simulation where we run boost-invariant viscous hydrodynamic for 100 thousand events, in which we repeat the Cooper-Frye procedure a thousand times. We repeat the same kind of analysis as in the toy model above, i.e. compare the results of cumulants computed using the two-step averaging versus the cumulants computed using a single particlization sample.

3.1. Toy Model

As stated above, the Toy Model does not contain any physical information, but serves rather as a way to easily generate large amounts of data, allowing us to study statistical effects in the two calculations methods. Nevertheless, we desire for the data generated to resemble real data.

We start by setting the probability density function of selecting an impact parameter b as

$$P(b) = \frac{2b}{B^2} \,, \tag{11}$$

where B is the biggest impact parameter we will allow[1]. This naturally leads to the usual expression relating centralities bin edges $\mathcal{C}$ and impact parameter

$$b = B\sqrt{\mathcal{C}} \,. \tag{12}$$

Since we desire a simple model that allow us to keep track of the density probability functions, we will assume the colliding nuclei are circular with radius $R = B/2$. Then, the area where the

[1] This choice comes from the argument that the differential cross section can be written as $d\sigma = 2\pi b\, db$, i.e., it will be proportional to the flux of nuclei crossing a ring of radius b

two circles intersect will be given by

$$A(b) = \frac{B^2}{2}\left[\arccos\left(\frac{b}{B}\right) - \frac{b}{B}\sqrt{1-\left(\frac{b}{B}\right)^2}\right] . \tag{13}$$

Lastly, we say that the entropy produced in a collision will be proportional to the area in Eq. (13) and thus that the number of particles produced in such event will be proportional to this area as well. In practice, this is done by normalizing Eq. (13) to unity and multiplying for the expected multiplicity $\tilde{M}(0)$ at $b = 0\,\text{fm}$, i.e.

$$\tilde{M}(b) = \frac{4\tilde{M}(0)}{\pi B^2} A(b) . \tag{14}$$

Note also that Eq. (14) can be rewritten in terms of the centrality dependency, hence this model does not incorporate any system-size effects. We are also neglecting IC fluctuations due to different number of participants and/or difference on their positions. But we stress that incorporating these effects is not our main objective in this section. We simply want to obtain a probability density function which functional form can be kept track of and that has a minimal resemblance with experiments.

To emulate the Cooper-Frye procedure, we will assign the multiplicity obtained above as the mean of a Poisson distribution. The actual number of observed particles is then sampled from this distribution. The number of times we sample this distribution corresponds to the number of times we "oversample" the event. With these ingredients we may write the probability density function of generating an event with multiplicity m as

$$P_M(m,b) = \frac{e^{-\tilde{M}(b)}\left[\tilde{M}(b)\right]^m}{m!}\frac{2b}{B^2} . \tag{15}$$

With the probability distribution above, one can compute μ_n in the centrality window $\mathcal{C}_1$-$\mathcal{C}_2$ as

$$\mu_n = \int_{B\sqrt{\mathcal{C}_1}}^{B\sqrt{\mathcal{C}_2}} \sum_{m=0}^{m=\infty} m^n \frac{P_M(m,b)}{\mathcal{C}_2 - \mathcal{C}_1}\, db . \tag{16}$$

In our case, we choose a centrality window in the 50-60% range. A typical simulation generates data of the order of 10^3-10^4 events. Meanwhile, experiments can easily go above 10^6-10^7 events. We will mock-up a situation "simulation vs. experiment" by applying this toy model to a scenario where a 1000 initial conditions are generated, each oversampled 1000 times. Then we compare with a second scenario where we generate 1 million events, each sampled only once.

In Fig. 1 we show the cumulants computed using the two-step averaging procedure for the first scenario as red points. Results for the second scenario are shown as blue points. Notice that for the second scenario, since no oversample is emulated, one must not employ two-step averaging but compute moments with the usual expression $\mu_n = \langle X^n \rangle$. We also show the calculation of C_n using only one sample per initial condition from scenario one. Since we are using only a single sample, moments are computed in the usual way, as in scenario two. But we do have many samples from the same simulation available. This allow us to repeat the calculation for each one of them. The distribution of these values are shown as a violin plot (light red) in Fig. 1.A possible alternative would be to take the mean of the cumulants computed in the way just described. However, this leads to larger deviations from the expected values (black points) — which are obtained by evaluating Eq. (16). Also, the results tends to be skewed to lower values, specially for the higher-order cumulants. Meanwhile, comparison between the two-step averaging method and the semi-analytical calculation shows no such bias within statistical errors.

Another point to be made is that we obtain $C_1 \cong C_2$. This is a simple consequence that on each bin, $b \approx$ constant and thus Poisson fluctuations dominates.

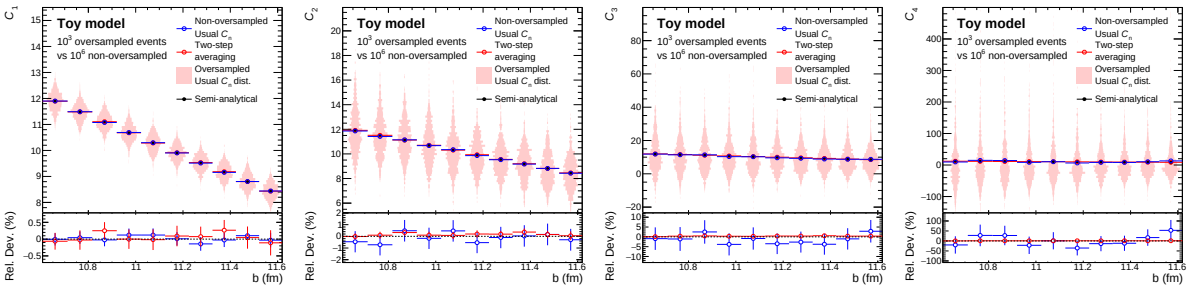

Figure 1. Comparison between cumulants computed via two-step averaging on top of 1000 toy model events sampled 1000 times each (red points), calculations performed on top of one million toy model events without oversampling (blue points) and semi-analytical cumulants (black points) from evaluating Eq. (16). The violin distribution comes from evaluating the cumulants taking only one sample from each event of the toy model events which were oversampled. The relative deviation is computed using the semi-analytical result as reference.

3.2. MUSIC Simulation results

Here we show a real world simulation where we used data from events evolved with relativistic hydrodynamics. We use a simulation chain similar to [9], but without pre-equilibrium dynamics and without simulation of the hadronic phase. In short, we use Trento as [10] as initial condition generator, evolve these initial conditions with second-order, boost-invariant, hydrodynamics[11, 12] on top of which we apply the Cooper-Frye procedure as implemented by iSS [6]. We also do not consider any sort of decays. As in [9], we restrict our study to Pb-Pb collisions at $\sqrt{s_{NN}}$. To make simulations computationally cheaper, we restricted the initial condition generation to the range $b \in [10.48, 11.73]$ fm. The choice of peripheral collisions with a boost-invariant simulation allowed us to push the statistics to 10^5 events, each one sampled a thousand times. In contrast to what was done in the previous section, instead of looking into the cumulants of protons, we look at the cumulants of net baryonic number, that is, the cumulants of net protons, which we use as a proxy for net-baryons.

The results we obtain are shown in Fig. 2. As before, the violin plot shows the distribution of C_n taking one sample individually from each event. The open point is the mean of this distribution, while the filled point is the cumulant computed by the two-step averaging procedure. The same effect as in the toy model where the two-step averaging is always above the mean of the distribution of C_n's is present for $n \geq 2$, highlighting the importance of correcting for the oversampling procedure.

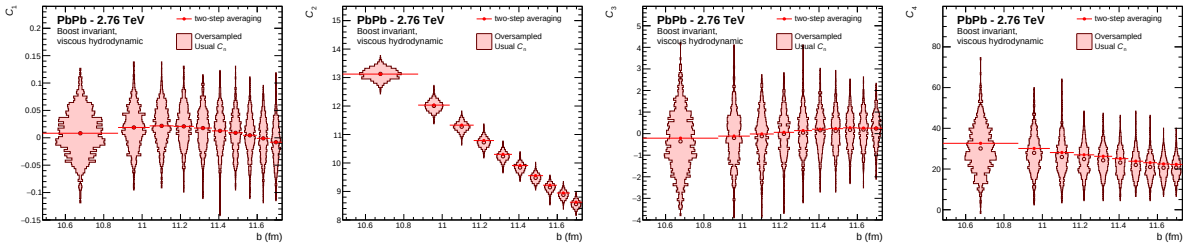

Figure 2. Cumulants of net protons for boost-invariant hydrodynamic simulations (closed points). The light red violin plot distribution comes from evaluating the cumulants taking only one sample from each event (in the same manner as the violin plot from Fig. 1 (see the text for further explanation). The open circle corresponds to the mean of this distribution.

4. Summary

In this work we have shown that an event-by-event simulation of an hydro-based hybrid model cannot use the naive expressions for the net-proton cumulants. Instead, one needs to account for the oversampling procedure by changing the usual cumulant expressions. We showed that even in a simulation with very large statistics, these effects are still present and must be taken into consideration.

A main limitation of what is shown here is the use of the Poisson distribution. This does not reproduce well data since it does not accounts for cuts in the phase space, commonly employed when comparing these calculations with experimental data. Results which take these into considerations were computed and are available in the pre-print in Ref. [8].

Acknowledgments

This work is supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) through grants 2018/24720-6, 2018/14479-0, 2021/01670-6, and by the Helmholtz Forschungsakademie Hessen für FAIR (HFHF).

References

[1] Stephanov M A 2009 *Phys. Rev. Lett.* **102** 032301 (*Preprint* 0809.3450)
[2] Stephanov M A 2011 *Phys. Rev. Lett.* **107** 052301 (*Preprint* 1104.1627)
[3] Abdallah M *et al.* (STAR) 2021 *Phys. Rev. C* **104** 024902 (*Preprint* 2101.12413)
[4] Hatta Y and Stephanov M A 2003 *Phys. Rev. Lett.* **91** 102003 [Erratum: Phys.Rev.Lett. 91, 129901 (2003)] (*Preprint* hep-ph/0302002)
[5] Rustamov A (ALICE) 2017 *Nucl. Phys. A* **967** 453–456 (*Preprint* 1704.05329)
[6] Shen C, Qiu Z, Song H, Bernhard J, Bass S and Heinz U 2016 *Comput. Phys. Commun.* **199** 61–85 (*Preprint* 1409.8164)
[7] Bernhard J E 2018 *Bayesian parameter estimation for relativistic heavy-ion collisions* Ph.D. thesis Duke U. (*Preprint* 1804.06469)
[8] Hirayama R, Grassi F, Serenone W M and Ollitrault J Y 2022 (*Preprint* 2204.07181)
[9] Nunes da Silva T, Chinellato D, Hippert M, Serenone W, Takahashi J, Denicol G S, Luzum M and Noronha J 2021 *Phys. Rev. C* **103** 054906 (*Preprint* 2006.02324)
[10] Moreland J S, Bernhard J E and Bass S A 2015 *Phys. Rev. C* **92** 011901 (*Preprint* 1412.4708)
[11] Schenke B, Jeon S and Gale C 2010 *Phys. Rev. C* **82** 014903 (*Preprint* 1004.1408)
[12] Schenke B, Jeon S and Gale C 2012 *Phys. Rev. C* **85** 024901 (*Preprint* 1109.6289)