Multiplicity-dependent and nonbinomial efficiency corrections for particle number cumulants

Adam Bzdak[*]
AGH University of Science and Technology, Faculty of Physics and Applied Computer Science, 30-059 Kraków, Poland

Romain Holzmann[†]
GSI Helmholtzzentrum für Schwerionenforschung GmbH, 64291 Darmstadt, Germany

Volker Koch[‡]
Nuclear Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA
(Received 4 April 2016; published 19 December 2016)

In this article we extend previous work on efficiency corrections for cumulant measurements [Bzdak and Koch, Phys. Rev. C **86**, 044904 (2012); **91**, 027901 (2015)]. We will discuss the limitations of the methods presented in these papers. Specifically we will consider multiplicity dependent efficiencies as well as nonbinomial efficiency distributions. We will discuss the most simple and straightforward methods to implement those corrections.

DOI: 10.1103/PhysRevC.94.064907

I. INTRODUCTION

Cumulants of conserved charges, such as the baryon number, are important observables in the search for a possible phase structure in the QCD phase diagram [1,2], and first measurements of the net-proton and net-charge cumulants up to fourth order have been carried out by the STAR collaboration [3–5]. As pointed out in [6] finite detection efficiencies give rise to fluctuations of the measured particle number and need to be corrected for. These corrections can be included in a straightforward manner if the efficiency follows a binomial distribution [6]. Appropriate formulas for a phase space dependent (binomial) efficiency have also been derived [7–9]. These corrections can be sizable as seen in the recent preliminary data by the STAR collaboration [10].

So far, however, all efficiency corrections have assumed that the efficiency (in a given phase space bin) is constant for a given centrality class, i.e., it does not depend on the multiplicity of particles under consideration. Furthermore, all the corrections have been carried out assuming that the detection efficiency follows a binomial distribution.

In reality, the efficiency does depend on the multiplicity of particles (see, e.g., Fig. 1 in [10]). Also it is not at all obvious if a binomial distribution correctly describes the detection probability. In the following we want to address both these issues and discuss methods how to improve the efficiency corrections. We demonstrate our points for proton cumulants, i.e., we neglect antiprotons, which is well justified at lower energies.

Let us start by defining more precisely what we mean by an "efficiency distribution". For simplicity, we will restrict ourselves to one kind of particles, such as protons, and ignore antiparticles. The extension to net-particles distributions is straightforward following the discussion in [6,7]. Let us denote the number distribution of the *produced* particles by $P(N)$ and that of the *observed* particles by $p(n)$.[1] Then the observed distribution is given by

$$p(n) = \sum_N B(n,N;\epsilon)P(N), \tag{1}$$

where $B(n,N;\epsilon)$ denotes the probability to observe n particles if N particles are produced. The probability $B(n,N;\epsilon)$ depends on the detection efficiency ϵ. It is this probability $B(n,N;\epsilon)$ that we call efficiency distribution. The detection efficiency, ϵ, is given by the ratio of mean number of *observed* particles, $\langle n\rangle$, over the mean number of produced particles, $\langle N\rangle$, $\epsilon = \langle n\rangle/\langle N\rangle$. Obviously, ϵ by itself, does not define the entire efficiency distribution. In practice however, B is typically *assumed* to be a binomial distribution

$$B(n,N;\epsilon) = \frac{N!}{n!(N-n)!}\epsilon^n(1-\epsilon)^{N-n}, \tag{2}$$

in which case the knowledge ϵ is sufficient to characterize the distribution. To which extent such an assumption is valid can only be verified by a detailed simulation of a given detector system.

Usually, the efficiency ϵ is assumed to be constant, i.e., independent of N, the number of particles under consideration. In this case, assuming a binomial efficiency distribution, factorial moments of the produced and observed particles are simply related by [6]

$$f_i = \epsilon^i F_i, \tag{3}$$

where the factorial moments are defined by

$$F_i = \sum_N P(N)\frac{N!}{(N-i)!}, \quad f_i = \sum_n p(n)\frac{n!}{(n-i)!}. \tag{4}$$

[*]bzdak@fis.agh.edu.pl
[†]r.holzmann@gsi.de
[‡]vkoch@lbl.gov

[1]Throughout this paper we will use lower case characters to refer to observed particles and upper case characters to refer to produced particles.

 ©2016 American Physical Society

Given the above relation for the factorial moments, efficiency corrections for the various cumulants are readily derived [6]. However, in reality the efficiency may depend on the multiplicity of particles under consideration. For example, in the case of the STAR measurement of net proton cumulants, the centrality is determined by the number of charged particles other than protons and antiprotons. The efficiency, on the other hand depends on *all* charged particles, including protons and antiprotons. Therefore, even for the tightest centrality cuts, the number of charged particles, and thus the efficiency, fluctuates from event to event in a given centrality class, simply because the number of produced protons and antiprotons fluctuate. In this case, as already pointed out in [11], the above relation between the factorial moments, Eq. (3), is not valid anymore. In addition, as already mentioned, the efficiency distribution may not be exactly binomial. In these cases, the above simple formula (3) will not hold and, as we will show in this paper, may lead to wrong conclusions.

Recent preliminary results by the STAR collaboration [10] at the lowest available BNL Relativistic Heavy Ion Collider energies show that the efficiency corrections play a crucial role in a proper interpretation of data. Therefore, it is essential that the correct unfolding procedure is applied. It is the purpose of this article to discuss various corrections and modifications to the unfolding procedure, and we should point out that we will only discuss the most simple and straightforward unfolding methods and apply them to cumulants. There are other, more refined, methods used to correct multiplicity distributions (see, e.g., [12–16]). While in Ref. [16] the Bayes unfolding procedure of [12] has been successfully tested in a model calculation for net-charge cumulants, we are not aware that these methods so far have been applied to actual data, and we hope that this article may motivate some work in applying these other methods to cumulant measurements.

This paper is organized as follows. In the next section we will show how the dependence of the efficiency on the number of particles changes the results. After that we will discuss the effect of non-binomial efficiency distributions by studying a few alternative distributions. Then we will discuss the simplest unfolding procedure. We will finish with a few comments and conclusions.

II. MULTIPLICITY DEPENDENT EFFICIENCY

In most experiments, the efficiency depends on the number of particles in the detector. This is also the case for the STAR experiment, where the efficiency does depend on the total number of charged particles, and thus may also depend on the number of particles under consideration, N, such as protons. While this does not preclude the distribution B from having the binomial form, Eq. (2), we now have an efficiency $\epsilon(N)$ that depends on the number of produced particles, N. To illustrate this point, consider the case of protons at midrapidity. As already eluded to in the Introduction, the centrality selection, which determines the mean efficiency, involves cuts on charged particles other than protons at midrapidity. Otherwise, fluctuation measurements would be biased by the centrality selection and autocorrelation effects. Therefore, in events with more protons than the average (in a given centrality class), $N > \langle N\rangle$, the multiplicity of all charged particles including protons is larger than the average, and, consequently the efficiency for these events is different from the mean efficiency. Obviously this effect is largest at low energies, where the protons represent a significant fraction of all charge particles. Furthermore, this correction cannot be removed by ever tighter centrality cuts as they do not affect the fluctuations of the proton number.[2] Given the dependence of the efficiency on N, $\epsilon(N)$ the relation between the factorial moments,

$$f_i = \sum\nolimits_N \epsilon^i(N)P(N)\frac{N!}{(N-i)!}, \tag{5}$$

is not as simple as in Eq. (3) [11]. Furthermore, the unfolding derived in [6] and used by the STAR collaboration [5] will not be possible anymore, even for a binomial efficiency distribution B.

In order to estimate the effect of a multiplicity dependent efficiency, let us consider a simple example based on a Poisson distribution for the produced particles and assume that the efficiency depends linearly on the number of produced particles N,

$$\begin{aligned} P(N) &= \frac{\langle N\rangle^N}{N!}e^{-\langle N\rangle}, \\ \epsilon(N) &= \epsilon_0 + \epsilon'(N-\langle N\rangle), \end{aligned} \tag{6}$$

where ϵ_0 is the average efficiency $\epsilon_0 = \sum_N P(N)\epsilon(N)$. In this case, the true cumulants ratios $K_{4,5,6}/K_2$ equal 1. Using Eq. (5) the factorial moments of the observed distribution are then

$$\begin{aligned} f_1 &= \langle N\rangle(\epsilon_0+\epsilon'), \\ f_2 &= \langle N\rangle^2[(\epsilon_0+2\epsilon')^2 + \langle N\rangle(\epsilon')^2], \\ f_3 &= \langle N\rangle^3[(\epsilon_0+3\epsilon')^3 + \langle N\rangle(\epsilon')^2(3\epsilon_0+10\epsilon')], \\ f_4 &= \langle N\rangle^4\big[(\epsilon_0+4\epsilon')^4 + \langle N\rangle(\epsilon')^2\big(6\epsilon_0^2+52\epsilon_0\epsilon'+113(\epsilon')^2 \\ &\quad + 3\langle N\rangle(\epsilon')^2\big)\big], \end{aligned} \tag{7}$$

and more complicated formulas for f_5 and f_6. Now we can correct using constant efficiency $F_i = f_i/\epsilon_0^i$, as described in Ref. [6] and calculate all cumulants. The obtained results are presented in Fig. 1, where we show K_4/K_2, K_5/K_2, and K_6/K_2 as a function of ϵ'. Obviously for $\epsilon' = 0$ our procedure is exact, $\epsilon(N) = \epsilon_0$, and we obtain $K_{4,5,6}/K_2 = 1$. Interestingly even a very small ϵ' leads to substantial deviation from unity.

To put things in perspective, for the STAR measurement at $\sqrt{s} = 7.7$ GeV, $\epsilon' \simeq -0.1/250 \simeq -4\times10^{-4}$ so that the correction for the ratio of K_4/K_2 is about 30% and much larger for K_6/K_2.[3] In this paper we are mostly interested in

[2]Moreover, selecting a very tight phase-space bins can still lead to sizable fluctuations of the multiplicity in the involved detector parts. The detector lives in geometric space $(\theta-\phi)$ meaning that tracks belonging to very different phase-space $(y-p_t)$ bins can still cross the same detector segment. How severe this is depends on a given detector.

[3]We note that better results are obtained when we use an effective constant efficiency $F_i = f_i/(\epsilon_0+i\epsilon')^i$, as can be seen from Eq. (7).

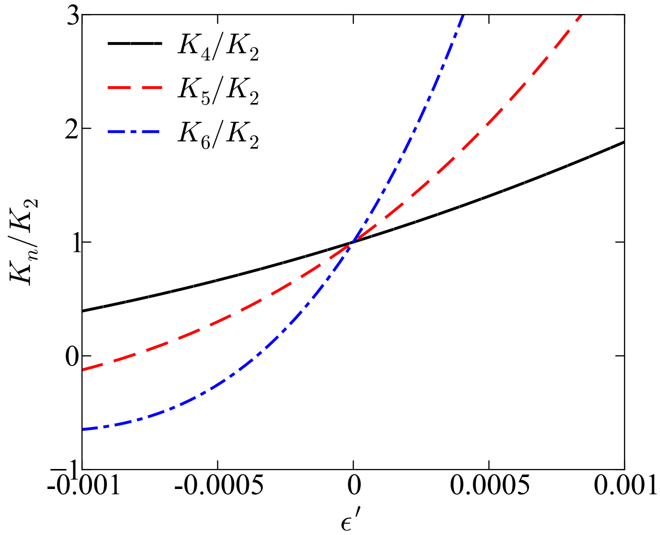

FIG. 1. $K_n/K_2, n = 4, 5, 6$, as a function of ϵ' when corrected using average, multiplicity independent, efficiency, ϵ_0, where in reality efficiency depends on the number of produced protons, $\epsilon(N) = \epsilon_0 + \epsilon'(N - \langle N \rangle)$. In this calculation $\epsilon_0 = 0.65$ and $\langle N \rangle = 40$ which are roughly the numbers for the STAR measurement at low energies.

lower energies where the net-proton number is dominated by protons and corrections are relatively easy to estimate. For higher energies things are more involved since a change in the net-proton not necessarily changes the number of charged particles or tracks. The same applies to net-charge cumulants as measured by PHENIX [17] and STAR [5].

While the 30% correction for K_4/K_2 may not seem like much, we should keep in mind that we have used a simple Poisson distribution to illustrate things. In reality, especially if we are close to a critical point, the true distribution will be far from Poisson and we need to be able to unfold in a reliable way. As already mentioned, the analytic methods described in [6] cannot be applied the moment we have a multiplicity dependent efficiency. In Sec. III, we will explore other means of unfolding the probability distribution.

A. Nonbinomial distributions

Next let us explore what happens if the efficiency does not follow a binomial distribution. To this end we will calculate the factorial moments and subsequent cumulants for other nonbinomial distributions. Here we chose the hypergeometric, the beta-binomial, and the Gaussian distributions. The first two have limits corresponding to the binomial distribution, allowing us to study the deviations from binomial systematically. Our choice of nonbinomial distributions is by no means motivated by any possible detector design or effect, but simply to estimate the effect of a possible nonbinomial distribution. Our strategy is to calculate the factorial moments f_i using Eq. (1) with these nonbinomial distributions $B(n,N)$ and unfold them using the formula for the binomial distribution with constant efficiency, $F_i = f_i/\epsilon^i$. As "input" distribution for the *produced* particles we chose again a Poisson distribution, with $\langle N \rangle = 40$. Therefore, the fact that we unfold a nonbinomial distribution using formulas based on the binomial distribution will be reflected by deviations from $K_n/K_2 = 1$. In order to isolate nonbinomial effects from issues related to multiplicity dependant efficiency, which we have discussed previously, we ensure that for the nonbinomial $B(n,N)$ distributions the effective efficiencies given by

$$\epsilon(N) = \frac{\langle n \rangle_N}{N} = \frac{1}{N} \sum_n n B(n,N) = \text{constant} \tag{8}$$

do not depend on N.

1. Hypergeometric distribution

As the first example we consider the hypergeometric distribution. Suppose we have an urn with N_w white balls and N_b black balls. For each produced particle we sample a ball, if it is white we accept a particle and if it is black we reject. In the case of the binomial distribution we return balls to the urn and for the hypergeometric distribution balls are not returned. In this case once we accept a particle (a white ball is removed from the urn) the probability to accept the next one is a bit smaller. The initial probability to accept a particle is given by $N_w/(N_w + N_b)$. The probability to accept n particles at a given produced N is given by $(N \leqslant N_w + N_b)$

$$B(n,N) = \frac{1}{\binom{N_w+N_b}{N}} \binom{N_w}{n} \binom{N_b}{N-n}, \tag{9}$$

where we chose

$$N_w = 2\alpha N, \quad N_b = \alpha N. \tag{10}$$

In this case

$$\frac{\langle n \rangle_N}{N} = \frac{N_w}{N_w + N_b} = \frac{2}{3} \tag{11}$$

for each value of N, which corresponds to $\epsilon = 2/3$ for the binomial distribution. We note that in the limit of $\alpha \to \infty$ the hypergeometric distribution approaches a binomial.[4] In Fig. 2 we show several curves for different values of α and fixed $N = 40$. As seen the hypergeometric distribution results in a narrower distribution than binomial.[5]

Finally we compute $p(n)$ using Eq. (1) and calculate the factorial moments f_i. Next we correct them $F_i = f_i/\epsilon^i$ and obtain the values presented in Table I.

2. Beta-binomial distribution

The beta-binomial distribution is obtained from the binomial one when the binomial success probability is random and follows the beta distribution. Another interpretation (for positive integer α and β being the numbers of white and black balls, respectively) is similar to the hypergeometric distribution however in this case once a white ball is drawn two white balls are returned to the urn (and similar for black

[4]For $N_w/N \to \infty$ and $N_b/N \to \infty$ the fact that balls are not returned to the urn is irrelevant.

[5]By expressing $B(n,N)$, Eq. (9), in terms of Γ functions, one is not restricted to integer values for N_b and N_w allowing to consider rather narrow distribution such as the example of $\alpha = 0.6$ discussed here.

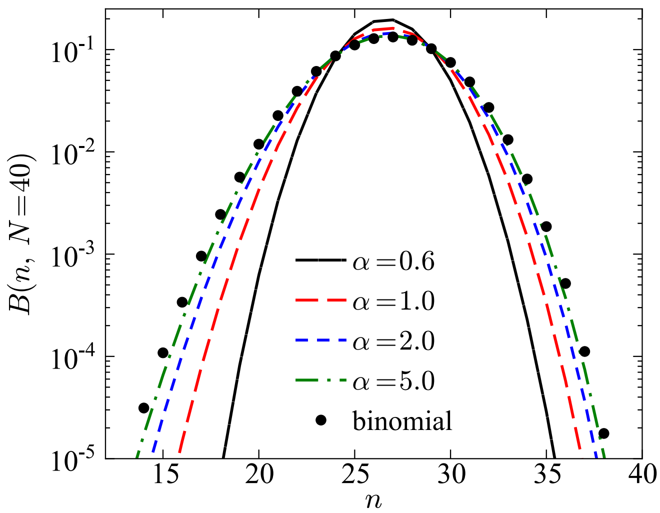

FIG. 2. The hypergeometric distribution for different values of α compared with the binomial distribution (black points). Here $N = 40$ and $\epsilon = 2/3$.

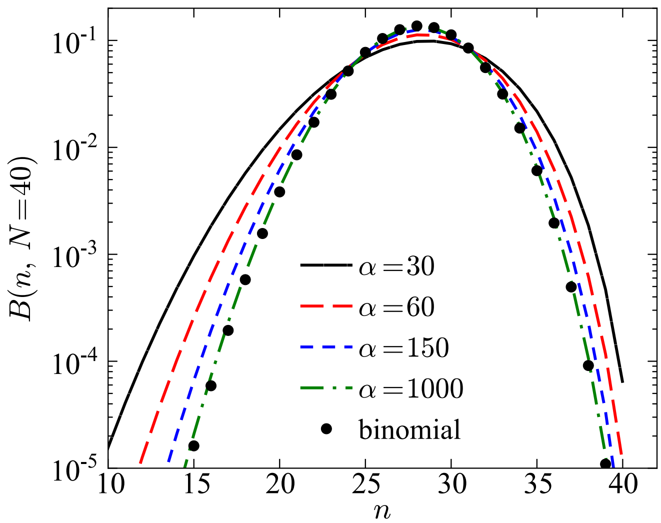

FIG. 3. The beta-binomial distribution for different values of α compared with the binomial distribution (black points). Here $N = 40$ and $\epsilon = 0.7$.

balls). The resulting distribution of n at a given N is broader than binomial and is given by

$$B(n,N) = \binom{N}{n} \frac{B_{\text{eta}}(n+\alpha, N-n+\beta)}{B_{\text{eta}}(\alpha,\beta)}, \tag{12}$$

where

$$B_{\text{eta}}(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \tag{13}$$

is the beta function or Euler integral of the first kind. Taking

$$\beta = \alpha\frac{1-\epsilon}{\epsilon}, \tag{14}$$

we obtain

$$\frac{\langle n\rangle_N}{N} = \epsilon, \tag{15}$$

that is the efficiency does not depend on N. When $\alpha \to +\infty$ the beta-binomial distribution goes into binomial. In Fig. 3 we present four curves for $N = 40$, $\epsilon = 0.7$ and different values of α.

Assuming the beta-binomial distribution we compute $p(n)$ using Eq. (1) with $P(N)$ given by Poisson and calculate the factorial moments f_i. Using $F_i = f_i/\epsilon^i$ we obtain the values presented in Table II.

3. Gaussian distribution

As the last example we consider the Gaussian distribution

$$B(n,N) = \mathcal{N}(N,\epsilon,\alpha)\exp\left(-\alpha\frac{(n-N\epsilon)^2}{2N\epsilon(1-\epsilon)}\right)\Theta(N-n), \tag{16}$$

where $\mathcal{N}$ is a normalization factor ensuring $\sum_{n=0}^{N} B(n,N) = 1$ and we enforce $B(n,N) = 0$ for $n > N$. For this distribution we approximately have (provided α is not too small)

$$\frac{\langle n\rangle_N}{N} \simeq \epsilon, \tag{17}$$

except for small values of N (which is of no interest since we consider $\langle N\rangle = 40$). In Fig. 4 we present four curves for different values of α and in Table III we show the corresponding values of cumulant ratios.

To summarize nonbinomial distributions result in K_n/K_2 which are different from one, as expected. However it is somewhat surprising and of course encouraging that for small deviations from binomial the effect on K_4/K_2 is rather weak, especially if distributions are narrower than binomial. We note that we also checked distributions for the produced particles other than Poisson and found qualitatively similar effects.

TABLE I. The obtained values of K_n/K_2 for the hypergeometric distribution, using $F_i = f_i/\epsilon^i$ with $\epsilon = 2/3$, for different values of α as presented in Fig. 2.

Hypergeometric	$\alpha = 0.6$	$\alpha = 1.0$	$\alpha = 2.0$	$\alpha = 5.0$
K_3/K_2	1.16	1.12	1.07	1.03
K_4/K_2	0.66	0.88	0.98	1.00
K_5/K_2	2.19	1.68	1.23	1.05
K_6/K_2	−3.99	−1.38	0.31	0.89

TABLE II. The obtained values of K_n/K_2 for the beta-binomial distribution, using $F_i = f_i/\epsilon^i$ with $\epsilon = 0.7$, for different values of α as presented in Fig. 3.

beta binomial	$\alpha = 30$	$\alpha = 60$	$\alpha = 150$	$\alpha = 1000$
K_3/K_2	1.28	1.24	1.13	1.02
K_4/K_2	0.82	1.45	1.35	1.07
K_5/K_2	−1.11	1.15	1.63	1.16
K_6/K_2	5.71	−0.44	1.80	1.32

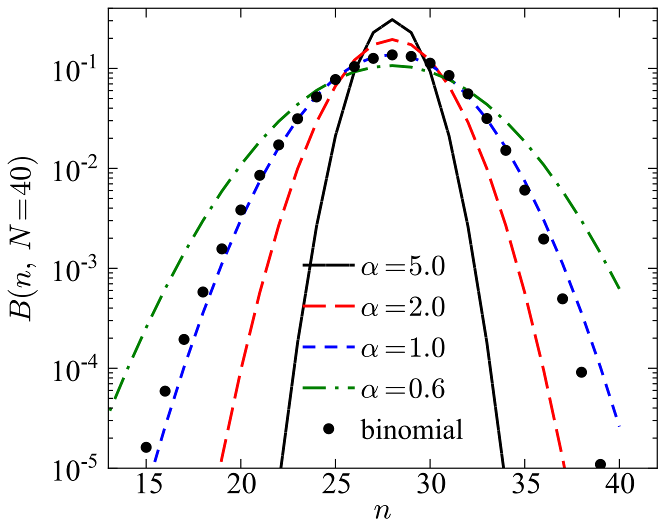

FIG. 4. The Gaussian distribution for different values of α compared with the binomial distribution (black points). Here $N = 40$ and $\epsilon = 0.7$.

III. UNFOLDING METHODS

Obviously the analytic formulas of Ref. [6] are rather limited and for the more general case of either multiplicity dependent efficiency or nonbinomial efficiency distributions, we need different methods to unfold the measured distributions and/or cumulants. In this section we demonstrate that the simplest unfolding method, based on solving triangular equations, result in correct cumulants even though the obtained multiplicity distribution, $P(N)$, is usually unphysical.

Let us first see what happens when we try to unfold the entire multiplicity distribution for the binomial efficiency distribution.

A. Multiplicity distribution

Our starting relation is

$$p(n) = \sum_{N=n}^{\infty} P(N)\frac{N!}{n!(N-n)!}\epsilon^n(1-\epsilon)^{N-n}, \tag{18}$$

which can also be cast in matrix form

$$p(n) = B(n,N)P(N), \tag{19}$$

where elements of B are given by Eq. (2). p is the measured distribution and P is the true one. To make analytical calculations we assume that ϵ does not depend on N. Later on we show numerical calculations with ϵ depending on N. So the problem of unfolding the multiplicity distribution is equivalent to inverting the above equation. We note, that although we will assume here that $B(n,N)$ is given by binomial our discussion is valid for other choices as well, as long as $B(n,N)$ is not a singular matrix.

Suppose that in our experiment we measure n from $n = 0$ to $n = M$, where M is sufficiently large so that $P(N) \simeq 0$ for all $N > M$. In this case the matrix gets finite and for example if $M = 4$ we have

$$\begin{pmatrix} p(0)\\ p(1)\\ p(2)\\ p(3)\\ p(4)\end{pmatrix} = \begin{pmatrix} 1 & 1-\epsilon & (1-\epsilon)^2 & (1-\epsilon)^3 & (1-\epsilon)^4\\ 0 & \epsilon & 2\epsilon(1-\epsilon) & 3\epsilon(1-\epsilon)^2 & 4\epsilon(1-\epsilon)^3\\ 0 & 0 & \epsilon^2 & 3\epsilon^2(1-\epsilon) & 6\epsilon^2(1-\epsilon)^2\\ 0 & 0 & 0 & \epsilon^3 & 4\epsilon^3(1-\epsilon)\\ 0 & 0 & 0 & 0 & \epsilon^4 \end{pmatrix} \times \begin{pmatrix} P(0)\\ P(1)\\ P(2)\\ P(3)\\ P(4)\end{pmatrix}. \tag{20}$$

Our goal is to solve Eq. (19) and obtain $P(N)$. One immediate problem is that the matrix B is practically singular in realistic situations. Indeed, the determinant of the triangular matrix B is given by a product of its diagonal elements. We obtain

$$\det(B) = \prod_{i=0}^{i=M} B(i,i) = \prod_{i=0}^{i=M} \epsilon^i = \epsilon^{0+1+\cdots+M-1+M} = \epsilon^{M(M+1)/2}. \tag{21}$$

For example for $\epsilon = 0.7$ and $M = 100$ we obtain $\det(B) \sim 10^{-782}$, which is zero for all practical purposes. Consequently solving Eq. (19) usually leads to unphysical $P(N)$. However we will show later that even though $P(N)$ is usually unphysical the obtained cumulants are correct.

In the case when B is given by a binomial distribution the inverse relation can be given analytically,

$$P(N) = \sum_{n=N}^{\infty} p(n)\frac{n!}{N!(n-N)!}\frac{1}{\epsilon^n}(-1+\epsilon)^{n-N}, \tag{22}$$

or in other words, the inverse of the binomial matrix is given by

$$B^{-1}(N,n) = \frac{n!}{N!(n-N)!}\frac{1}{\epsilon^n}(-1+\epsilon)^{n-N}, \tag{23}$$

TABLE III. The obtained values of K_n/K_2 for the Gaussian distribution, using $F_i = f_i/\epsilon^i$ with $\epsilon = 0.7$, for different values of α as presented in Fig. 4.

Gaussian	$\alpha = 5.0$	$\alpha = 2.0$	$\alpha = 1.0$	$\alpha = 0.6$
K_3/K_2	1.00	1.12	1.24	1.33
K_4/K_2	0.54	0.93	1.58	2.22
K_5/K_2	1.40	1.77	2.30	0.57
K_6/K_2	− 1.97	− 0.46	3.31	− 14.3

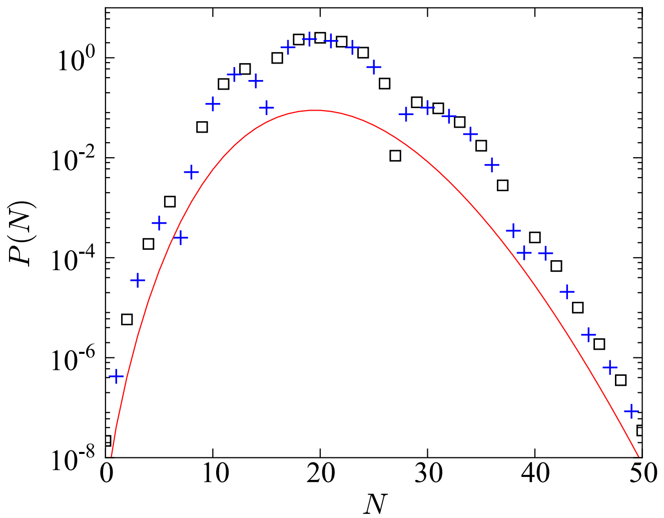

FIG. 5. The calculated $P(N)$ using the exact Eq. (22) and the exact $p(n)$ given by Eq. (18) multiplied by $(1+\delta_n)$, where δ_n is a random number very close to zero. The solid red line represents the Poisson distribution, which is our input. The blue crosses represent negative $P(N)$ and are shown as $-P(N)$. Positive $P(N)$ are shown as black open squares. As discussed in the main text this unphysical $P(N)$ results, with a very good accuracy, in correct cumulants.

so that[6]

$$P(N) = B^{-1}(N,n)p(n), \tag{24}$$

or, explicitly, the first few terms,

$$\begin{pmatrix} P(0) \\ P(1) \\ P(2) \\ P(3) \\ P(4) \end{pmatrix} = \begin{pmatrix} 1 & \frac{\epsilon-1}{\epsilon} & \frac{(\epsilon-1)^2}{\epsilon^2} & \frac{(\epsilon-1)^3}{\epsilon^3} & \frac{(\epsilon-1)^4}{\epsilon^4} \\ 0 & \frac{1}{\epsilon} & \frac{2(\epsilon-1)}{\epsilon^2} & \frac{3(\epsilon-1)^2}{\epsilon^3} & \frac{4(\epsilon-1)^3}{\epsilon^4} \\ 0 & 0 & \frac{1}{\epsilon^2} & \frac{3(\epsilon-1)}{\epsilon^3} & \frac{6(\epsilon-1)^2}{\epsilon^4} \\ 0 & 0 & 0 & \frac{1}{\epsilon^3} & \frac{4(\epsilon-1)}{\epsilon^4} \\ 0 & 0 & 0 & 0 & \frac{1}{\epsilon^4} \end{pmatrix} \begin{pmatrix} p(0) \\ p(1) \\ p(2) \\ p(3) \\ p(4) \end{pmatrix}. \tag{25}$$

As seen from Eq. (22), $P(N)$ is prone to large errors since we add many terms of alternating sign. This is the main reason why $P(N)$ is usually unphysical. This is especially problematic since $p(n)$ will only be known within statistical uncertainties. However, as we will argue below the resulting cumulants are usually correct (within statistical errors) even if $P(N)$ is unphysical.

B. Cumulants

As already pointed out, the sum in Eq. (22) has subsequent positive and negative terms and there is a delicate cancellation between them. As a consequence, even for negligible "noise" on $p(n)$ we get an incorrect $P(N)$, unless ϵ is very close to unity, $\epsilon \simeq 1$. This is demonstrated in Fig. 5, where we have started out with a Poisson distribution with $\langle N \rangle = 20$ for the produced particles. We then calculate the distribution of observed particles, $p(n)$ using Eq. (18) for $\epsilon = 0.6$. We introduce the "noise" by replacing $p(n)$ with $p(n)(1+\delta_n)$, where δ_n is sampled from the Gaussian distribution, $e^{-\delta_n^2/2\sigma^2}$ with $\sigma = 10^{-5}$. Finally we use Eq. (22) to obtain $P(N)$. Our result is presented in Fig. 5. The solid red line represents Poisson, which is our input. The blue crosses represent negative $P(N)$ and are shown as $-P(N)$. Positive $P(N)$ are shown as black open squares.

However this problem vanishes once we sum over N. Indeed the factorial moment F_i is given by

$$\begin{aligned} F_i &\equiv \sum_{N=i}^{\infty} P(N) \frac{N!}{(N-i)!} = \sum_{N=i}^{\infty} \sum_{x=N}^{\infty} p(x)(1+\delta_x) \\ &\quad \times \frac{x!}{N!(x-N)!} \frac{1}{\epsilon^x} (-1+\epsilon)^{x-N} \frac{N!}{(N-i)!} \\ &= \sum_{x=i}^{\infty} p(x)(1+\delta_x) \frac{1}{\epsilon^i} \frac{x!}{(x-i)!} \simeq \frac{1}{\epsilon^i} f_i, \end{aligned} \tag{26}$$

and the noise δ_n does not affect the result in a significant way since we add only positive numbers, as $|\delta_n| \ll 1$.

Consequently, while the extracted multiplicity distribution $P(N)$ is rather erratic and clearly unphysical, the calculated cumulant ratio, K_4/K_2 equals 1 with good accuracy.[7]

C. Cumulants with multiplicity dependent efficiency

In this section we test the previously discussed method by using the multiplicity dependent efficiency given by Eq. (6). We sample the produced particles from a Poisson distribution $P(N)$ and we parametrize the efficiency by $\epsilon(N) = \epsilon_0 + \epsilon'(N - \langle N \rangle)$. We use the following parameters for our calculation: $\langle N \rangle = 40$, $\epsilon_0 = 0.7$, $\epsilon' = -0.0005$ and $\langle N \rangle = 20$, $\epsilon_0 = 0.6$, $\epsilon' = -0.001$.

We first sample N from the Poisson distribution. Next for each of these N particles we decide whether it is detected or not with binomial probability $\epsilon(N) = \epsilon_0 + \epsilon'(N - \langle N \rangle)$. We run 10^7 events which allows to calculate the measured $p(n)$. Our efficiency matrix is given by

$$B(n,N) = \frac{N!}{n!(N-n)!} \epsilon(N)^n [1 - \epsilon(N)]^{N-n}. \tag{27}$$

Next we solve Eq. (19) for the true distribution $P(N)$. Here we take advantage of the fact that the matrix B is triangular, which allows for straightforward backward substitution. We note there is no need to invert B, which is not advised for pseudosingular matrices.[8] Finally we calculate the true K_4/K_2. We repeat this exercise a few thousand times and

[6] We note that in practical applications inverting a pseudo-singular matrix B is not advised. Instead, equations should be solved directly taking advantage of the fact that B is triangular (by definition $N \geqslant n$).

[7] This argument applies only to the situation where an unphysical $P(N)$ originates from the noise on $p(n)$. However, even with $\delta_x = 0$, $P(N)$ in Eq. (22) can get unphysical due to numerical errors, and one has to ensure that the numerical accuracy is able to handle large terms of alternating signs (especially for small ϵ).

[8] We checked that inverting B and using Eq. (24) leads to unphysical cumulants.

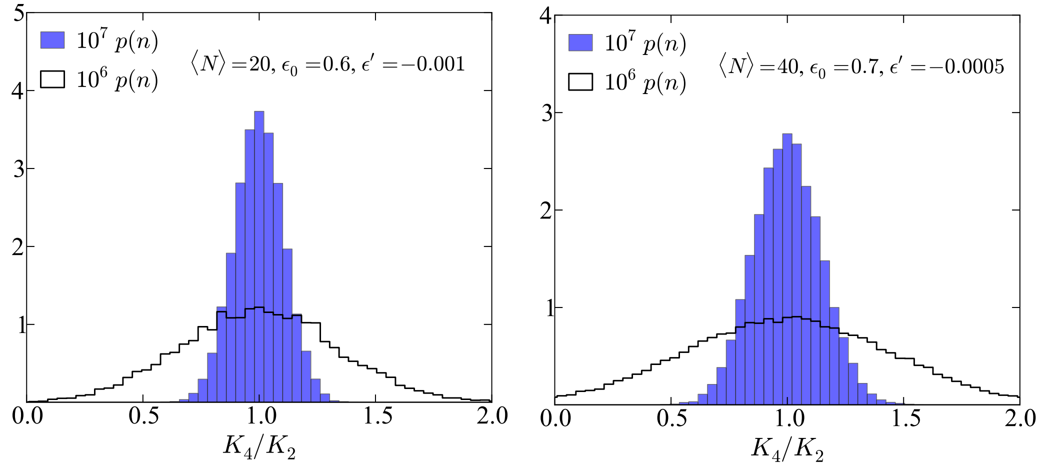

FIG. 6. Histogram (normalized to unity) of K_4/K_2 in the case of sampled $p(n)$ (10^7 events) and analytical matrix given by Eq. (27) with $\epsilon(N) = \epsilon_0 + \epsilon'(N - \langle N\rangle)$. For comparison we also show results with 10^6 events.

plot histogram of the resulting cumulant ratios K_4/K_2. This is presented in Fig. 6

As seen from Fig. 6 the cumulant ratios are centered around 1 with rather small statistical spread (for 10^7 events in each experiment). For comparison we also show calculations with 10^6 events, which obviously results in larger statistical error.

One could worry that this method introduces larger errors than the method with constant efficiency proposed in Ref. [6]. We checked this explicitly and this is not the case. We took the sampled $p(n)$, calculated f_i and corrected them using a constant average efficiency, ϵ_0, so that $F_i = f_i/\epsilon_0^i$. The resulting cumulant ratios K_4/K_2 are shown in Fig. 7 as a black solid line.

As discussed in Sec. I, K_4/K_2 comes out too small (in agreement with Fig. 1). However the statistical error is comparable to the method with solving triangular Eq. (19).

The obvious advantage of solving Eq. (19) is that we now obtain the correct cumulant ratio K_4/K_2. Therefore, as long as we know the efficiency matrix and if it is not singular, we can determine the cumulants of the original distribution of produced particles, $P(N)$ within purely statistical uncertainties.

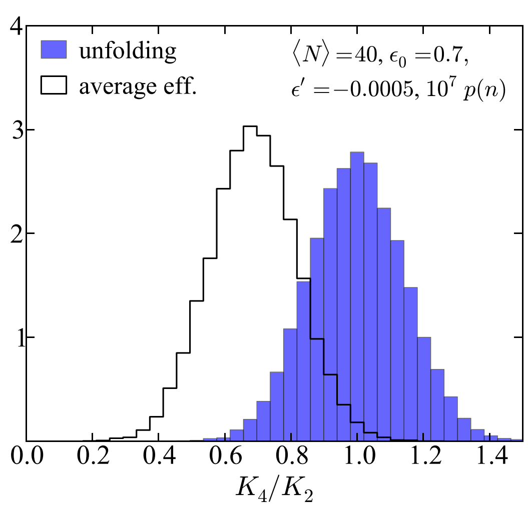

FIG. 7. Histogram (normalized to unity) of K_4/K_2 calculated using the sampled $p(n)$ and (i) by solving Eq. (19) and (ii) by correcting factorial moments using average constant efficiency. Both methods result in comparable statistical errors.

In principle one could simulate the efficiency matrix, $B(n,N)$, by careful analysis of a given detector, for example using GEANT. In this case we would run many simulations and for each true N we determine the measured probability to observe n particles resulting in the efficiency matrix $B(n,N)$, which should reflect all detector effects (to the extent that they are properly simulated in GEANT). Unfortunately this method most likely results in the matrix B being mathematically singular. The problem is that with growing N it is getting very unlikely to observe $n = N$ (unless ϵ is very close to 1) and we get zeros for some diagonal elements. For a triangular matrix it means that its determinant vanishes and that it is therefore singular. One could cut the matrix so that it is nonsingular, however this results in too small matrix.[9]

Let us briefly discuss two ways to overcome this difficulty. First, one could try to fit the simulated matrix $B(n,N)$ with some function, for example a binomial distribution with ϵ depending on N. Using the analytical form for B we obtain the full matrix and we can successfully extract true cumulants, as discussed in the previous section. Of course this method is model-dependent and relies on a correct extrapolation of B to higher values of N.

Another way is to keep the simulated (incomplete) $B(n,N)$ as it is and solve the matrix equation using the singular value decomposition (Moore-Penrose pseudoinverse). This method is designed to obtain a solution from an underdetermined set of equations (less equations than unknowns, which is our case). We checked empirically that this method reproduces a distribution of cumulant ratios which is also centered at $K_4/K_2 = 1$, however, with very long tails, so that the width is not only due to finite statistics.

[9]Unless we can simulate our detector with very large, currently impossible, number of events. For example if B is binomial with $\epsilon = 0.7$, the diagonal elements are $B(n = N,N) = \epsilon^N$. For instance it gives roughly 10^{-10} for $N = 65$.

Obviously one may consider alternative methods, such as the one already employed to extract the charged particle multiplicity distribution, see, e.g., [13], namely a Bayesian unfolding method [12], or other refined unfolding methods [14]. However, we are not aware of any work where this method has been applied to the determination of higher order cumulants and it would be worthwhile to assess its suitability.

IV. CONCLUSIONS

(i) We stress that it is very unlikely that a binomial distribution with constant efficiency is a correct model for the efficiency distribution. To the very least one likely has to take into account a multiplicity dependent efficiency. Given the substantial uncertainties demonstrated in this article, it appears mandatory that each experiment wishing to measure higher order cumulants needs to extract the full efficiency matrix, specific to this experiment. To which extent this can be done reliably, remains to be seen. Maybe the only solution for a credible measurement of higher order cumulants is to design a dedicated experiment which has an efficiency very close to 1.

(ii) As already stated in the Introduction, in this article we studied the most straightforward method for the determination of the cumulants of the true distribution. However, we are aware that there are more sophisticated unfolding methods, e.g., those applied in high-energy physics to various problems (see, e.g., [12–15]). To our knowledge these have not yet been applied to the determination of higher order cumulants, however, and it would be worthwhile to study their suitability to do so.

ACKNOWLEDGMENTS

We thank A. Kalweit and J. Thaeder for useful discussions. We also thank the HIC for FAIR and ExtreMe Matter Institute (EMMI) for support to attend two workshops where this work was initiated. A.B. was supported by the Ministry of Science and Higher Education (MNiSW), by funding from the Foundation for Polish Science, and by the National Science Centre (Narodowe Centrum Nauki), Grant No. DEC-2014/15/B/ST2/00175, and in part by DEC-2013/09/B/ST2/00497. V.K. was supported by the Office of Nuclear Physics in the US Department of Energy's Office of Science under Contract No. DE-AC02-05CH11231.

[1] M. A. Stephanov, Phys. Rev. Lett. **102**, 032301 (2009).
[2] V. Koch, in *Relativistic Heavy Ion Physics*, edited by R. Stock, Landolt-Boernstein New Series I, v. 23 (Springer, Heidelberg, 2010), pp. 626–652.
[3] N. Xu (STAR Collaboration), Nucl. Phys. A **931**, 1 (2014).
[4] L. Adamczyk *et al.* (STAR Collaboration), Phys. Rev. Lett. **113**, 092301 (2014).
[5] L. Adamczyk *et al.* (STAR Collaboration), Phys. Rev. Lett. **112**, 032302 (2014).
[6] A. Bzdak and V. Koch, Phys. Rev. C **86**, 044904 (2012).
[7] A. Bzdak and V. Koch, Phys. Rev. C **91**, 027901 (2015).
[8] X. Luo, Phys. Rev. C **91**, 034907 (2015).
[9] M. Kitazawa, Phys. Rev. C **93**, 044911 (2016).
[10] X. Luo (STAR Collaboration), PoS CPOD **2014**, 019 (2015).
[11] A. H. Tang and G. Wang, Phys. Rev. C **88**, 024905 (2013).
[12] G. D'Agostini, Nucl. Instrum. Methods Phys. Res. A **362**, 487 (1995).
[13] V. Khachatryan *et al.* (CMS Collaboration), JHEP 01 (2011) 079.
[14] S. Schmitt, JINST **7**, T10003 (2012).
[15] J. F. Grosse-Oetringhaus and K. Reygers, J. Phys. G **37**, 083001 (2010).
[16] P. Garg, D. K. Mishra, P. K. Netrakanti, A. K. Mohanty, and B. Mohanty, J. Phys. G **40**, 055103 (2013).
[17] A. Adare *et al.* (PHENIX Collaboration), Phys. Rev. C **93**, 011901(R) (2016).