**THE EUROPEAN
PHYSICAL JOURNAL C**

CrossMark

Generalisation of the identity method for determination of high-order moments of multiplicity distributions with a software implementation

Maja Maćkowiak-Pawłowska[1,a], **Piotr Przybyła**[2]

[1] Faculty of Physics, Warsaw University of Technology, Warsaw, Poland
[2] School of Computer Science, University of Manchester, Manchester, UK

Abstract The incomplete particle identification limits the experimentally-available phase space region for identified particle analysis. This problem affects ongoing fluctuation and correlation studies including the search for the critical point of strongly interacting matter performed on SPS and RHIC accelerators. In this paper we provide a procedure to obtain nth order moments of the multiplicity distribution using the identity method, generalising previously published solutions for $n = 2$ and $n = 3$. Moreover, we present an open source software implementation of this computation, called *Idhim*, that allows one to obtain the true moments of identified particle multiplicity distributions from the measured ones provided the response function of the detector is known.

1 Introduction

Search for the critical point of strongly interacting matter remains one of the most important goals of experimental searches in heavy ion physics [1,2]. Its basic property – the increase of the correlation length of the considered system – forces experimenters to shift their interests from inclusive spectra to higher-order moments and cumulants of the particle multiplicity distributions. A particular interest is paid towards net-proton fluctuations being the most sensitive to the searched phenomenon [3].

One of the most serious experimental issues, which largely limits the available phase-space coverage, and, possibly, affects the studied signal, is the incomplete particle identification caused by finite detector resolution. To overcome this problem an experimental technique, called *the identity method*, was proposed in Ref. [4], and extended in Refs. [5–7]. So far the identity method was described for the sec-

ond [5] and third [6] order moments. In Ref. [6] it was also used to reexamine the first moments of the identified particle distributions. The impact of particle losses due to detector inefficiencies on results from the identity method is discussed in Ref. [7]. The author shows that it remains applicable provided detection efficiencies can be determined with sufficient accuracy. With the ongoing development of theoretical studies concerning higher order moments it seems appropriate to extend experimental techniques and tools as well.

In the present study the identity method is extended in two ways. Firstly, a strict procedure to obtain nth order moments of the multiplicity distribution is shown. Secondly, a program, called *Idhim*, which performs such calculations for any given number of considered particle types, is presented. It also allows one to obtain moments up to any order provided the detector response function is known. The modification of first moments from Ref. [6], also included in *Idhim*, may address possible biases in other popular methods [8] (e.g. maximal likelihood method [9,10]).

The paper is organized as follows. In Sect. 2, basic quantities of the identity method are presented. The computation of nth moments of true multiplicity distribution is shown in Sect. 3. Modifications necessary to apply the general formulas in practice are addressed in Sect. 4. Description of the *Idhim* program which computes moments of the true multiplicity distribution is given in Sect. 5. Section 6 contains tests of the program with the detector response close to the ones measured in real experiments. Conclusion in Sect. 7 ends the paper.

2 Basic quantities

The identity method is developed under the assumption that particles are identified by measuring quantity x (e.g., a mass)

[a] e-mail: majam@if.pw.edu.pl

of observed particles. Due to the finite detector resolution one gets a continuous distribution for x, denoted by $\rho_j(x)$, where index j stands for one of k particle types. The density is expected to sum up to the mean of N_j, i.e., the multiplicity for this type:

$$\int dx \rho_j(x) = \langle N_j \rangle, \tag{1}$$

For a given particle observation, its conditional probability of being of a given type is expressed by a quantity called *identity*, defined as:

$$w_j(x) \equiv \frac{\rho_j(x)}{\sum_{l=1}^{k} \rho_l(x)}. \tag{2}$$

In the case of complete particle identification w_j is reduced to two extreme values: $w_j = 0$ for particles of types other than j and $w_j = 1$ for particles of type j.

In the same way, one can define an aggregated quantity for a given particle type:

$$W_j \equiv \sum_{i=1}^{N(\nu)} w_j(x_i), \tag{3}$$

where $N(\nu)$ is the total multiplicity (including all particle types) of the νth of considered N_{ev} events. From these events one obtains the distribution of different types of W with its moments defined as

$$\langle W_1^{n_1} \cdot W_2^{n_2} \cdot \ldots \cdot W_k^{n_k} \rangle = \frac{1}{N_{ev}} \sum_{i=1}^{N_{ev}} W_1^{n_1} \cdot W_2^{n_2} \cdot \ldots \cdot W_k^{n_k}, \tag{4}$$

where n_j denotes the order of the moment of the distribution of W_j.

3 Computing the nth moments of multiplicity distribution

We will now show how one can compute all the nth moments of the multiplicity distribution $\langle N_1^{n_1} \cdot N_2^{n_2} \cdot \ldots \cdot N_k^{n_k} \rangle$ with $n_1 + n_2 + \cdots + n_k = n$ using the moments of the measured identity variables. The procedure will be a generalisation of those published for $n = 2$ [5] and $n = 3$ [6]. First, we shall demonstrate how the value of a moment of identity variables $\langle W_1^{n_1} \cdot W_2^{n_2} \cdot \ldots \cdot W_k^{n_k} \rangle$, depends on the multiplicity distribution. We have the following:

$$\langle W_1^{n_1} \cdot W_2^{n_2} \cdot \ldots \cdot W_k^{n_k} \rangle = \sum_{N_1=0}^{\infty} \sum_{N_2=0}^{\infty} \cdots \sum_{N_k=0}^{\infty}$$

$$P(N_1, N_2, \ldots, N_k) \int dx_1^1 P_1(x_1^1) \ldots \int dx_{N_1}^1 P_1(x_{N_1}^1)$$

$$\int dx_1^2 P_2(x_1^2) \ldots \int dx_{N_2}^2 P_2(x_{N_2}^2) \cdots$$

$$\int dx_1^k P_k(x_1^k) \ldots \int dx_{N_k}^k P_k(x_{N_k}^k)$$

$$[w_1(x_1^1) + \cdots + w_1(x_{N_1}^1) + w_1(x_1^2) + \cdots + w_1(x_{N_2}^2)$$

$$+ \cdots + w_1(x_1^k) + \cdots + w_1(x_{N_k}^k)]^{n_1}$$

$$\times [w_2(x_1^1) + \cdots + w_2(x_{N_1}^1) + w_2(x_1^2) + \cdots + w_2(x_{N_2}^2)$$

$$+ \cdots + w_2(x_1^k) + \cdots + w_2(x_{N_k}^k)]^{n_2}$$

$$\times \cdots \times [w_k(x_1^1) + \ldots + w_k(x_{N_1}^1) + w_k(x_1^2) + \cdots + w_k(x_{N_2}^2)$$

$$+ \cdots + w_k(x_1^k) + \cdots + w_k(x_{N_k}^k)]^{n_k}, \tag{5}$$

where $P(N_1, N_2, \ldots, N_k)$ is the multiplicity distribution, i.e., the probability of observing N_1 particles of the first time, N_2 particles of the second type and so forth, and $P_j(x) = \frac{\rho_j(x)}{\langle N_j \rangle}$ is the probability distribution of the jth type.

Let us firstly focus on the innermost part of Eq. 5, denoted hereafter by ω:

$$\omega \equiv \prod_{l=1}^{k} \left[\sum_{j=1}^{k} \sum_{i=1}^{N_j} w_l \left(x_i^j \right) \right]^{n_l}. \tag{6}$$

We will now use the multinomial theorem to expand the n_lth power. Let us first define the following notation for brevity:

$$\sum_{a_1, a_2, \ldots, a_k}^{a} () \equiv \sum_{a_1 + a_2 + \cdots + a_k = a} \binom{a}{a_1, a_2, \ldots, a_k}. \tag{7}$$

In this notation the multinomial theorem is represented by

$$(x_1 + x_2 + \cdots + x_k)^a = \sum_{a_1, a_2, \ldots, a_k}^{a} () \prod_{i=1}^{k} x_i^{a_i}, \tag{8}$$

which allows us to express ω as

$$\omega = \prod_{l=1}^{k} \sum_{\eta_{(l)}^1, \ldots, \eta_{(l)}^k}^{n_l} () \prod_{j=1}^{k} \left[\sum_{i=1}^{N_j} w_l \left(x_i^j \right) \right]^{\eta_{(l)}^j}, \tag{9}$$

where the first summation is over all possible combinations of k nonnegative integers $\eta_{(l)}^1, \ldots, \eta_{(l)}^k$ that sum up to n_l. Let us now use the multinomial theorem again, this time to expand the $\eta_{(l)}^j$th power:

$$\omega = \prod_{l=1}^{k} \sum_{\eta_{(l)}^1, \ldots, \eta_{(l)}^k}^{n_l} () \prod_{j=1}^{k} \sum_{\eta_{1(l)}^j, \ldots, \eta_{N_j(l)}^j}^{\eta_{(l)}^j} () \prod_{i=1}^{N_j} \left[w_l(x_i^j) \right]^{\eta_{i(l)}^j}. \tag{10}$$

This formulation could be rearranged to give

$$
\omega = \sum_{\eta^1_{(1)},\ldots,\eta^k_{(1)}}^{n_1} () \sum_{\eta^1_{(2)},\ldots,\eta^k_{(2)}}^{n_2} () \cdots \sum_{\eta^1_{(k)},\ldots,\eta^k_{(k)}}^{n_k} ()
$$

$$
\prod_{j=1}^{k} \sum_{\eta^j_{1(1)},\ldots,\eta^j_{N_j(1)}}^{\eta^j_{(1)}} () \sum_{\eta^j_{1(2)},\ldots,\eta^j_{N_j(2)}}^{\eta^j_{(2)}} () \cdots \sum_{\eta^j_{1(k)},\ldots,\eta^j_{N_j(k)}}^{\eta^j_{(k)}} ()
$$

$$
\prod_{i=1}^{N_j} w_1(x_i^j)^{\eta^j_{i(1)}} \cdot w_2(x_i^j)^{\eta^j_{i(2)}} \cdot \ldots \cdot w_k(x_i^j)^{\eta^j_{i(k)}}. \tag{11}
$$

If we now put ω expressed in such way back to Eq. 5, we can notice that integration over x_i^j can be applied to the product $w_1(x_i^j)^{\eta^j_{i(1)}} \cdot \ldots \cdot w_k(x_i^j)^{\eta^j_{i(k)}}$ to give

$$
\langle W_1^{n_1} \cdot \ldots \cdot W_k^{n_k} \rangle
$$

$$
= \sum_{N_1=0}^{\infty} \cdots \sum_{N_k=0}^{\infty} P(N_1,\ldots,N_k)
$$

$$
\sum_{\eta^1_{(1)},\ldots,\eta^k_{(1)}}^{n_1} () \sum_{\eta^1_{(2)},\ldots,\eta^k_{(2)}}^{n_2} () \cdots \sum_{\eta^1_{(k)},\ldots,\eta^k_{(k)}}^{n_k} ()
$$

$$
\prod_{j=1}^{k} \sum_{\eta^j_{1(1)},\ldots,\eta^j_{N_j(1)}}^{\eta^j_{(1)}} () \sum_{\eta^j_{1(2)},\ldots,\eta^j_{N_j(2)}}^{\eta^j_{(2)}} () \cdots \sum_{\eta^j_{1(k)},\ldots,\eta^j_{N_j(k)}}^{\eta^j_{(k)}} ()
$$

$$
\prod_{i=1}^{N_j} u_j(\eta^j_{i(1)},\eta^j_{i(2)},\ldots,\eta^j_{i(k)}), \tag{12}
$$

where function u_j is defined as

$$
u_j(\eta^j_{i(1)},\eta^j_{i(2)},\ldots,\eta^j_{i(k)}) \equiv \frac{1}{\langle N_j \rangle} \int w_1(x)^{\eta^j_{i(1)}} \cdot w_2(x)^{\eta^j_{i(2)}}
$$

$$
\cdot \ldots \cdot w_k(x)^{\eta^j_{i(k)}} \rho_j(x) \mathrm{d}x. \tag{13}
$$

Let us now focus on the part of Eq. 12 depending on j, which will be denoted by λ^j,

$$
\lambda^j \equiv \sum_{\eta^j_{1(1)},\ldots,\eta^j_{N_j(1)}}^{\eta^j_{(1)}} () \sum_{\eta^j_{1(2)},\ldots,\eta^j_{N_j(2)}}^{\eta^j_{(2)}} () \cdots \sum_{\eta^j_{1(k)},\ldots,\eta^j_{N_j(k)}}^{\eta^j_{(k)}} ()
$$

$$
\prod_{i=1}^{N_j} u_j(\eta^j_{i(1)},\eta^j_{i(2)},\ldots,\eta^j_{i(k)}). \tag{14}
$$

Since each u_j depends on a tuple of values of length k, it is convenient to introduce a notation for such tuples:

$$
\boldsymbol{\eta}^j_i \equiv (\eta^j_{i(1)},\eta^j_{i(2)},\ldots,\eta^j_{i(k)})
$$

$$
\boldsymbol{\eta}^j \equiv (\eta^j_{(1)},\eta^j_{(2)},\ldots,\eta^j_{(k)})
$$

$$
\boldsymbol{\eta}^j_p + \boldsymbol{\eta}^j_q \equiv (\eta^j_{p(1)} + \eta^j_{q(1)}, \eta^j_{p(2)} + \eta^j_{q(2)}, \ldots, \eta^j_{p(k)}
$$

$$
+ \eta^j_{q(k)}). \tag{15}
$$

This allows us to express λ^j as

$$
\lambda^i = \sum_{\boldsymbol{\eta}^j_1 + \cdots + \boldsymbol{\eta}^j_{N_j} = \boldsymbol{\eta}^j}
$$

$$
\left[\prod_{l=1}^{k} \binom{\eta^j_{(l)}}{\eta^j_{1(l)},\ldots,\eta^j_{N_j(l)}} \cdot \prod_{i=1}^{N_j} u_j(\boldsymbol{\eta}^j_i) \right], \tag{16}
$$

where the first summation is over all possible combinations of N_j tuples $\boldsymbol{\eta}^j_1,\ldots,\boldsymbol{\eta}^j_{N_j}$ (each containing k nonnegative integers) that sum up to $\boldsymbol{\eta}^j$. We can notice that zero tuples, i.e., $\boldsymbol{\eta}^j_i = (0,0,\ldots,0)$, do not contribute to λ^j since $u_j(0,0,\ldots,0) = 1$. Let us then express the sequence $\boldsymbol{\eta}^j_1, \boldsymbol{\eta}^j_2, \ldots, \boldsymbol{\eta}^j_{N_j}$ as a combination of several non-zero tuples.

Let Γ^j denote a set of all such combinations possible for $\boldsymbol{\eta}^j$, so that each $\gamma \in \Gamma^j$ consists of $|\gamma|$ different non-zero tuples: $\boldsymbol{\mu}^{\gamma}_1, \boldsymbol{\mu}^{\gamma}_2, \ldots, \boldsymbol{\mu}^{\gamma}_{|\gamma|}$, each occurring $m^{\gamma}_1, m^{\gamma}_2, \ldots, m^{\gamma}_{|\gamma|}$ times, respectively. There are also $N_j - \sum_{p=1}^{|\gamma|} m^{\gamma}_p$ tuples equal to zero in the original sequence $\boldsymbol{\eta}^j_1, \boldsymbol{\eta}^j_2, \ldots, \boldsymbol{\eta}^j_{N_j}$. We therefore have

$$
\forall_{\gamma \in \Gamma^j} \sum_{p=1}^{|\gamma|} m^{\gamma}_p \cdot \boldsymbol{\mu}^{\gamma}_p = \boldsymbol{\eta}^j. \tag{17}
$$

If we use this to compute λ^j, we get

$$
\lambda^j = \sum_{\gamma \in \Gamma^j} \binom{N_j}{m^{\gamma}_1, m^{\gamma}_2, \ldots, m^{\gamma}_{|\gamma|}, N_j - \sum_{p=1}^{|\gamma|} m^{\gamma}_p}
$$

$$
\times \left[\prod_{l=1}^{k} \binom{\eta^j_{(l)}}{m^{\gamma}_1 * \mu^{\gamma}_{1(l)}, m^{\gamma}_2 * \mu^{\gamma}_{2(l)}, \ldots, m^{\gamma}_{|\gamma|} * \mu^{\gamma}_{|\gamma|(l)}} \right.
$$

$$
\left. \times \prod_{p=1}^{|\gamma|} [u_j(\boldsymbol{\mu}^j_p)]^{m^{\gamma}_p} \right] + \mathbb{1}[\boldsymbol{\eta}^j = 0], \tag{18}
$$

where $a * b$ means that the value b appears a times in the multinomial symbol. The indicator variable is necessary so that in the case of $\boldsymbol{\eta}^j = 0$, we have $\lambda^j = 1$, as in Eq. 16.

If we now focus on the multinomial symbol involving N_j, it can be expanded as

$$
\binom{N_j}{m^{\gamma}_1, \ldots, m^{\gamma}_{|\gamma|}, N_j - \sum_{p=1}^{|\gamma|} m^{\gamma}_p}
$$

$$
= \frac{N_j!}{m^{\gamma}_1! \cdot m^{\gamma}_2! \cdot \ldots \cdot m^{\gamma}_{|\gamma|}! \cdot \left(N_j - \sum_{p=1}^{|\gamma|} m^{\gamma}_p\right)!}
$$

$$
= \frac{N_j \cdot (N_j - 1) \cdot \ldots \cdot \left(N_j - \left(\sum_{p=1}^{|\gamma|} m^{\gamma}_p - 1\right)\right)}{m^{\gamma}_1! \cdot m^{\gamma}_2! \cdot \ldots \cdot m^{\gamma}_{|\gamma|}!}. \tag{19}
$$

We can see it is a polynomial of N_j of degree $\sum_{p=1}^{|\gamma|} m_p^\gamma$, which is at least 1 and at most $\eta_\Sigma^j \equiv \sum_{l=1}^k \eta_{(l)}^j$.

Since λ^j is a weighted sum of such polynomials (and indicator variable), it is a polynomial of N_j of at most the same degree and can therefore be expressed as

$$\lambda^j = \lambda_0^j + \lambda_1^j N_j + \lambda_2^j N_j^2 + \cdots + \lambda_{\eta_\Sigma^j}^j N_j^{\eta_\Sigma^j}. \tag{20}$$

Further coefficients, i.e., λ_p^j for $p > \eta_\Sigma^j$, equal zero.

Let us put this formulation of λ^j back to Eq. 12. This gives us

$$\langle W_1^{n_1} \cdot \ldots \cdot W_k^{n_k} \rangle$$
$$= \sum_{N_1=0}^\infty \cdots \sum_{N_k=0}^\infty P(N_1, \ldots, N_k)$$
$$\sum_{\eta_{(1)}^1, \ldots, \eta_{(1)}^k}^{n_1} () \sum_{\eta_{(2)}^1, \ldots, \eta_{(2)}^k}^{n_2} () \cdots \sum_{\eta_{(k)}^1, \ldots, \eta_{(k)}^k}^{n_k} ()$$
$$\prod_{j=1}^k \lambda_0^j + \lambda_1^j N_j + \lambda_2^j N_j^2 + \cdots + \lambda_{\eta_\Sigma^j}^j N_j^{\eta_\Sigma^j}. \tag{21}$$

We can rearrange it as

$$\langle W_1^{n_1} \cdot \ldots \cdot W_k^{n_k} \rangle$$
$$= \sum_{N_1=0}^\infty \cdots \sum_{N_k=0}^\infty P(N_1, \ldots, N_k)$$
$$\sum_{\eta_{(1)}^1, \ldots, \eta_{(1)}^k}^{n_1} () \cdots \sum_{\eta_{(k)}^1, \ldots, \eta_{(k)}^k}^{n_k} () \sum_{q_1=0}^{\eta_\Sigma^1} \sum_{q_2=0}^{\eta_\Sigma^2} \cdots \sum_{q_k=0}^{\eta_\Sigma^k}$$
$$\lambda_{q_1}^1 \lambda_{q_2}^2 \ldots \lambda_{q_k}^k \times N_1^{q_1} N_2^{q_2} \ldots N_k^{q_k}, \tag{22}$$

which finally gives us

$$\langle W_1^{n_1} \cdot \ldots \cdot W_k^{n_k} \rangle = \sum_{\eta_{(1)}^1, \ldots, \eta_{(1)}^k}^{n_1} () \cdots \sum_{\eta_{(k)}^1, \ldots, \eta_{(k)}^k}^{n_k} ()$$
$$\sum_{q_1=0}^{\eta_\Sigma^1} \sum_{q_2=0}^{\eta_\Sigma^2} \cdots \sum_{q_k=0}^{\eta_\Sigma^k} \lambda_{q_1}^1 \lambda_{q_2}^2 \ldots \lambda_{q_k}^k \langle N_1^{q_1} \cdot N_2^{q_2} \cdot \ldots \cdot N_k^{q_k} \rangle. \tag{23}$$

Since $\sum_{j=1}^k \eta_\Sigma^j = n_1 + n_2 + \cdots + n_k = n$, the order of the moments at the right hand side is at most equal n. We have therefore just shown how to express any moment of W distributions with order n as a sum of moments of N distributions with order $\leq n$. Since this dependency is linear, we can define the whole problem as a set of linear equations.

It will have the coefficients

$$a_{n_1, n_2, \ldots, n_k}^{q_1, q_2, \ldots, q_k} \equiv \sum_{\eta_{(1)}^1, \ldots, \eta_{(1)}^k}^{n_1} () \cdots \sum_{\eta_{(k)}^1, \ldots, \eta_{(k)}^k}^{n_k} () \lambda_{q_1}^1 \lambda_{q_2}^2 \ldots \lambda_{q_k}^k, \tag{24}$$

where $n_1 + n_2 + \cdots + n_k = n$ and $q_1 + q_2 + \cdots + q_k = n$. We also need to define elements which will take into account the contribution of moments of N with orders lower than n:

$$b_{n_1, n_2, \ldots, n_k} \equiv \langle W_1^{n_1} \cdot \ldots \cdot W_k^{n_k} \rangle$$
$$- \sum_{\eta_{(1)}^1, \ldots, \eta_{(1)}^k}^{n_1} () \cdots \sum_{\eta_{(k)}^1, \ldots, \eta_{(k)}^k}^{n_k} ()$$
$$\sum_{q_1=0}^{\eta_\Sigma^1} \sum_{q_2=0}^{\eta_\Sigma^2} \cdots \sum_{q_k=0}^{\eta_\Sigma^k} \mathbb{1}[q_1 + q_2 + \cdots + q_k < n]$$
$$\lambda_{q_1}^1 \lambda_{q_2}^2 \ldots \lambda_{q_k}^k \langle N_1^{q_1} \cdot N_2^{q_2} \cdot \ldots \cdot N_k^{q_k} \rangle. \tag{25}$$

To arrange the moments in a linear order, let us now choose any one-to-one function f from sequences of length k summing up to n to numbers $1, 2, \ldots, \binom{n+k-1}{k-1}$. We can use it to construct a matrix $\mathbf{A}$ having elements $A_{\xi, \zeta} = a_{n_1, n_2, \ldots, n_k}^{q_1, q_2, \ldots, q_k}$ and vector $\mathbf{B}$ with elements $B_\xi = b_{n_1, n_2, \ldots, n_k}$ for $\xi = f(n_1, n_2, \ldots, n_k)$ and $\zeta = f(q_1, q_2, \ldots, q_k)$. We can also arrange unknown moments in a vector $\mathbf{N}$ such that $N_\zeta = \langle N_1^{q_1} \cdot N_2^{q_2} \cdot \ldots \cdot N_k^{q_k} \rangle$. This allows us to express equation 23 as

$$\sum_\zeta A_{\xi, \zeta} N_\zeta = B_\xi, \tag{26}$$

or in matrix notation:

$$\mathbf{AN} = \mathbf{B}. \tag{27}$$

If $\det \mathbf{A} \neq 0$, the moment we are looking for can be computed as

$$\mathbf{N} = \mathbf{A}^{-1} \mathbf{B}. \tag{28}$$

4 Modifications

In the previous section we have shown the procedure to compute the nth moments of the multiplicity distribution as a generalisation of the computations for $n = 2$ and $n = 3$, but to apply it in practice we needed to make three modifications.

Firstly, to compute the first moments as proposed in Ref. [6], we need to replace Eq. 1 with the following:

$$\int \mathrm{d}x \rho_j(x) = \langle A_j \rangle. \tag{29}$$

Now the distribution of a measured x for a given particle type j is normalised to arbitrary value A_j, which does not have to equal N_j. As a result, we also need modify Eq. 13, which now becomes

$$u_j(\eta^j_{i(1)}, \eta^j_{i(2)}, \ldots, \eta^j_{i(k)}) \equiv \frac{1}{\langle A_j \rangle}$$
$$\times \int w_1(x)^{\eta^j_{i(1)}} \cdot w_2(x)^{\eta^j_{i(2)}} \cdot \ldots \cdot w_k(x)^{\eta^j_{i(k)}} \rho_j(x) \mathrm{d}x. \tag{30}$$

The rest of the procedure holds, and corrected $\langle N_j \rangle$ could be computed by applying it for $n = 1$.

Secondly, the measured x is traditionally associated with the particle mass, but it can be any measured quantity, not necessarily a single scalar value. In general, it could be a multi-dimensional vector $\mathbf{x}$, e.g. mean energy loss and time-of-flight, as long as integration in function u is performed accordingly.

Thirdly, measurement of x could be performed in several phase space bins, corresponding to different detector configurations. In such cases Eq. 1 takes form

$$\sum_{\theta \in \Theta} \int \mathrm{d}x \rho_j(x, \theta) = \langle N_j \rangle, \tag{31}$$

where θ denotes a configuration from a configuration space Θ. Analogously, the definition of w (Eq. 2) has to take into account θ as well:

$$w_j(x, \theta) \equiv \frac{\rho_j(x, \theta)}{\sum_{l=1}^k \rho_l(x, \theta)}, \tag{32}$$

where $w_j(x, \theta)$ denotes value of the jth identity variable for a measurement x registered in configuration θ. Finally, the computation of u (Eq. 13) has to take into account measurements in all configurations, so

$$u_j(\eta^j_{i(1)}, \eta^j_{i(2)}, \ldots, \eta^j_{i(k)}) \equiv \frac{1}{\langle N_j \rangle} \sum_{\theta \in \Theta}$$
$$\int w_1(x, \theta)^{\eta^j_{i(1)}} \cdot \ldots \cdot w_k(x, \theta)^{\eta^j_{i(k)}} \rho_j(x, \theta) \mathrm{d}x, \tag{33}$$

All three modifications have been described here separately for simplicity, but could be combined if necessary.

5 Implementation

The *Idhim* program was designed to provide an easy way to obtain moments of the true multiplicity distribution of identified particles provided the detector resolution is know.

The implementation in *Java*, using *EJML*[1] library for linear algebra operations, is available as open source.[2] The required input to the program includes:

(i) a list of particle types in a text file, with each line providing a particle type name,
(ii) $\langle W_1^{n_1} \cdot \ldots \cdot W_k^{n_k} \rangle$ moments in a tsv (i.e., tab-separated values) file, with each line describing one moment as a list if $n_1, \ldots, n_k$ indices, followed by the moment value,
(iii) a list of phase space bins, where a detector response is known, as a tsv file (if there is more than one kinematic variables which define such bins, multiple tab-separated indices may be provided),
(iv) a directory containing files with a detector response functions in each bin.

An exemplary set of all needed files is provided with the program.

The input format allows for applicability to a wide range of different experiments. Firstly, a number of considered particle types is arbitrary. In a typical case of particle identification it depends on a collision energy and available statistics, e.g., at low interaction energies one does not need to consider deuterons and/or Helium-3, whereas at high energies or with large available statistics they must be taken into account. Secondly, only $\langle W_1^{n_1} \cdot \ldots \cdot W_k^{n_k} \rangle$ moments, not the full distributions, need to be provided. Finally, in a typical experiment, a particle identification is performed by a set of detectors with an overlapping momentum coverage. Thus, a full momentum coverage of an experiment consists of regions with ρ being 1D function, e.g., when particles are identified only by dE/dx or time-of-flight (ToF), or 2D function, e.g., when particles are identified by combined measurements of dE/dx and ToF. An example of such a non-uniform detector acceptance is shown in Fig. 1. Bins with any number of dimensions, which reflect changing detector configuration or particle yields, can be defined as long as density function for all particle types is given in the same points of the space.

The next section includes example demonstrating the usefulness of the program features described above.
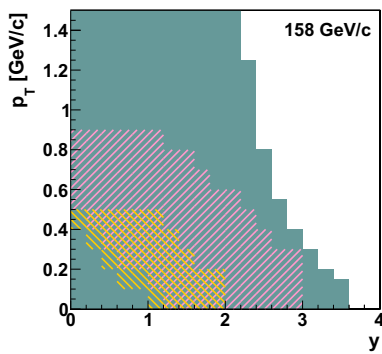
Fig. 1 NA61/SHINE detector acceptance (solid area) with indicated region where particles are identified via their energy loss (magenta stripes) and their time-of-flight (yellow stripes) in p+p interactions at $\sqrt{s_{NN}} = 17.3$ GeV
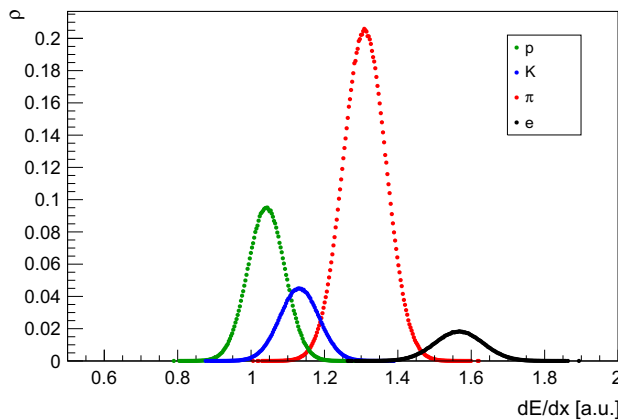

Fig. 2 Generated dE/dx distribution in a single bin. For details see text

6 Test on simulated data

The computation of all moments of multiplicity distributions up to the fourth order was tested on two models. The first one is a Monte Carlo model (so-called fast generator), where the number of particles of a given type produced in a single event was generated from Poisson distributions with a different free parameter λ for each considered particle type. Test included four most popular particle types, namely electrons, pions, kaons and protons, with their respective λ as 1, 10, 2, 4. The number of events is set to 1,000,000.

Particles are generated according to the Poisson distribution and are uncorrelated (except the detector response), so the true values of generated moments are

$$\langle N_j \rangle = \lambda, \tag{34}$$

$$\langle N_j^2 \rangle = \lambda(1 + \lambda), \tag{35}$$

$$\langle N_j^3 \rangle = \lambda(1 + 3\lambda + \lambda^2), \tag{36}$$

$$\langle N_j^4 \rangle = \lambda(1 + 7\lambda + 6\lambda^2 + \lambda^3). \tag{37}$$

The generated cross-moments are defined as the multiplication of the pure ones.

A simulated detector response consists of mean energy loss measurements in the Time Projection Chamber. For each particle, its mean energy loss was generated from a Gaussian distribution with parameters based on experimental data from Refs. [9,11] in two bins simulating the momentum dependence of the detector response. Testing several different momentum dependencies showed that particle distribution between bins does not affect the final results. An exemplary simulated dE/dx distribution in a single bin is shown in Fig. 2.
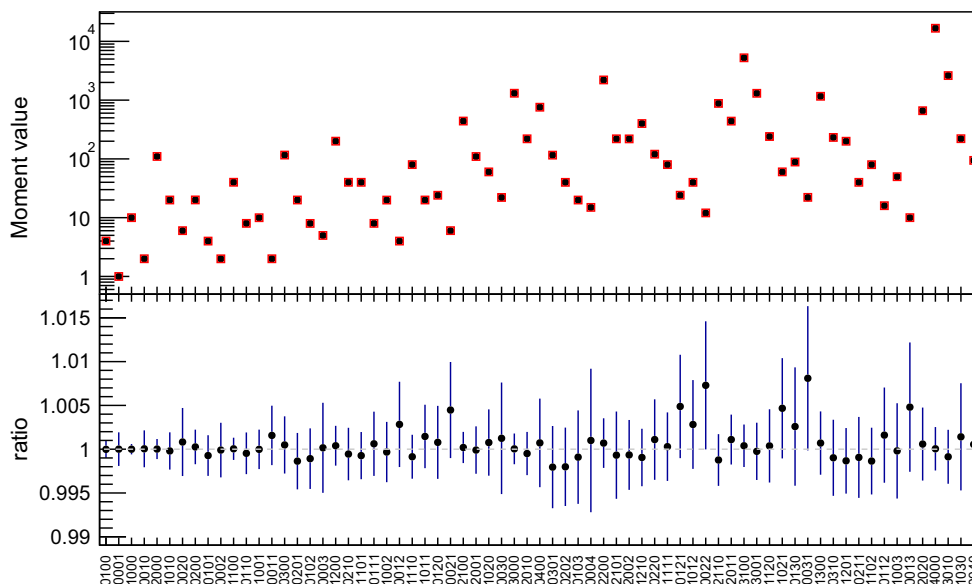

Fig. 3 Reconstructed (black circles) and generated (red squares) moments and their ratio

Fig. 4 Generated dE/dx and m^2 distribution in p + p interactions at 158 GeV/c beam momentum in EPOS

The *Idhim* program is used to obtain reconstructed moments of the considered particle types up to the fourth order. The statistical uncertainty of the reconstructed moments results from uncertainty of the fitted distributions $\rho_j(x)$ as well as from the $\langle W_1^{n_1} \cdot \ldots \cdot W_k^{n_k} \rangle$ moment values. Both sources are correlated, so the standard error propagation is complicated and inconvenient. Instead, the statistical uncertainty is obtained using the bootstrap method [12].

Reconstructed and generated moments as well as their ratio are shown in Fig. 3. The ratio is 1 within the statistical uncertainty for all considered values.

Another test was performed using 3 million p+p interactions at $\sqrt{s_{NN}} = 17.3$ GeV generated using the EPOS [13, 14] model with the detector acceptance containing two types of acceptance regions — dE/dx only and combined ToF and dE/dx. An example of such a two dimensional distribution is shown in Fig. 4. The shape of the 2D distribution and its parameters were based on a real data analysis in Ref. [15].

Again, in order to mimic the momentum dependence of the detector response, it was divided into several bins.

Reconstructed and generated moments as well as their ratio are shown in Fig. 5. Again, the ratio is 1 within the statistical uncertainty for all considered values.

Both the procedure and its implementation are functioning as expected. The difference between generated and reconstructed first moments of N and W is negligible but in case of the higher orders, the differences can reach 70%. In order to accommodate for different possible shapes of the ρ functions, they are delivered in a binned form. Thus, a proper binning is important to describe the functions' shapes. The identity method does not address other detector biases or its efficiency. Other possible biases should be addressed by the appropriate experimental tools (for examples and details see Refs. [16–18]).

7 Conclusion

In this paper we extend the identity method in two ways. Firstly, a new strict procedure to obtain nth order moments of multiplicity distribution of an arbitrary number of particles is discussed. Secondly, a software implementation of this procedure is presented. Provided a detector response is known, it computes any moments, including the first ones. It is equally precise both for low and high mean multiplicities. Two tests were performed in order to validate the program. The first test, based on simple fast generator check, showed that program works well in case of lack of correlations between particles. The difference between the recon-

Fig. 5 Reconstructed (black circles) and generated (red squares) moments and their ratio. The ratio uncertainty values for moments 0003 and 0004, omitted from the plot for clarity, equal 1.00 ± 0.07 and 1.00 ± 0.55, respectively

structed and generated moments is at the level of statistical uncertainty or below. The second test was performed on p+p interactions simulated in the EPOS model. The second test confirmed that correlations between particles do not affect the program's efficiency. It also showed that *Idhim* can be easily used in the case of a non-uniform detector acceptance which contains different detector types.

As a last comment we would like to stress that the successful analysis of moments of identified particle distributions depends on an understanding of a detector response. Possible flaws in description of the ρ functions will propagate to the identity method and the final results. Moreover, the identity method does not compensate for a limited detector efficiency. Thus, ρ distributions and mean $\langle W \rangle$'s have to be corrected for a limited and often momentum-dependent detector efficiency by other known methods.

References

1. N. Antoniou, et al., Study of hadron production in hadron nucleus and nucleus nucleus collisions at the CERN SPS CERN-SPSC-2006-034 (2006)
2. The STAR Coll. Beam energy scan white paper: studying the phase diagram of QCD matter at RHIC (2014)
3. Y. Hatta, M.A. Stephanov, Phys. Rev. Lett. **91**, 102003 (2003) (**Erratum: Phys. Rev. Lett.91,129901(2003)**))
4. M. Gazdzicki, K. Grebieszkow, M. Mackowiak, S. Mrowczynski, Phys. Rev. **C83**, 054907 (2011)
5. M. Gorenstein, Phys. Rev. **C84**, 024902 (2011)
6. A. Rustamov, M. Gorenstein, Phys. Rev. **C86**, 044906 (2012)
7. C.A. Pruneau, Phys. Rev. C **96**(5), 054902 (2017)
8. R.R. Prado, in *Proceedings of Science the 35th International Cosmic Ray Conference 2017*, Busan, South Korea (2017)
9. M. van Leeuwen, A practical guide to de/dx analysis in na49, note, CERN EDMS (2008)
10. M. van Leeuwen, in *Proceedings, 38th Rencontres de Moriond on QCD and High-energy Hadronic Interactions: Les Arcs, France, March 22–29, 2003* (2003)
11. A. Aduszkiewicz et al., Eur. Phys. J. **C77**(10), 671 (2017)
12. B. Efron, SIAM Rev. **21**(4), 460–480 (1979)
13. K. Werner, in *Proceedings of 14th International Symposium on Very High Energy Cosmic Ray Interactions (ISVHECRI 2006)*, Nucl. Phys. Proc. Suppl. 175–176, 81–87 (2008)
14. T. Pierog, K. Werner", Nucl. Phys. Proc. Suppl. **196**, 102–105 (2009)
15. M. Kuich, in *Proceedings of Science of The European Physical Society Conference on High Energy Physics*, Venice, Italy (2017)
16. A. Aduszkiewicz et al., [NA61/SHINE Collaboration], Eur. Phys. J. C **76**(11), 635 (2016)
17. A. Bzdak, R. Holzmann, V. Koch, Phys. Rev. C **94**(6), 064907 (2016)
18. M. Lorenz, Hadron production at SIS energies: an update from HADES, slides at Critical Point and Onset of Deconfinement, Wroclaw, Poland (2016)