Multiplicities in Au-Au and Cu-Cu collisions at $\sqrt{s_{NN}} = 62.4$ and 200 GeV

Dariusz Prorok
Institute of Theoretical Physics, University of Wrocław, Pl. Maksa Borna 9, 50-204 Wrocław, Poland
(Received 3 August 2012; revised manuscript received 9 August 2013; published 30 September 2013)

Likelihood ratio tests are performed for the hypothesis that charged particle multiplicities measured in Au-Au and Cu-Cu collisions at $\sqrt{s_{NN}} = 62.4$ and 200 GeV are distributed according to the negative binomial form. Results suggest that the hypothesis should be rejected in all classes of collision systems and centralities of Pioneering High-Energy Nuclear Interaction Experiment Relativistic Heavy Ion Collider measurements. However, the application of the least-squares test statistic with systematic errors included shows that for the collision system Au-Au at $\sqrt{s_{NN}} = 62.4$ GeV the hypothesis could not be rejected in general.

DOI: 10.1103/PhysRevC.88.034910 PACS number(s): 25.75.Ag, 13.85.Hd, 25.75.Gz, 29.85.Fj

I. INTRODUCTION

The analysis of charged hadron multiplicities in Au-Au and Cu-Cu collisions at $\sqrt{s_{NN}} = 62.4$ and 200 GeV was done by the Pioneering High-Energy Nuclear Interaction Experiment (PHENIX) Collaboration in [1]. It was also claimed there that these multiplicities are distributed according to the negative binomial form. The UA5 Collaboration noticed for the first time that charged particle multiplicity distributions measured in high-energy proton-(anti)proton collisions in limited intervals of pseudorapidity have this form [2,3].

The negative binomial distribution (NBD) is defined as

$$P(n;p,k) = \frac{k(k+1)(k+2)\dots(k+n-1)}{n!}(1-p)^n p^k, \tag{1}$$

where $n = 0,1,2,\dots, 0 \leqslant p \leqslant 1$, and k is a positive real number. In the application to high-energy physics n has the meaning of the number of charged particles detected in an event. The expected value $\bar{n}$ and variance[1] $V(n)$ are expressed as

$$\bar{n} = \frac{k(1-p)}{p}, \quad V(n) = \frac{k(1-p)}{p^2}. \tag{2}$$

Multiplicity fluctuations are expressed in terms of the scaled variance:

$$\omega = \frac{\langle N_{\rm ch}^2\rangle - \langle N_{\rm ch}\rangle^2}{\langle N_{\rm ch}\rangle} = \frac{V(n)}{\bar{n}}, \tag{3}$$

where $N_{\rm ch}$ is the charged particle multiplicity and the last equality is valid only for the whole population (the set of all possible outcomes if the experiment is repeated infinitely many times), assuming that the hypothesis about the NBD is true.

In application to the high-energy physics, the parameters $k, \bar{n}$ instead of k, p are used usually and

$$\frac{1}{p} = 1 + \frac{\bar{n}}{k} = \omega, \tag{4}$$

which is the scaled variance, Eq. (3). But because the centrality bins have the nonzero width, fluctuations defined by Eq. (3) also include a nondynamical component. This component is the result of the fluctuations of the geometry of the collisions within a given centrality bin. The geometrical fluctuations were evaluated by the PHENIX Collaboration in [1]. It turned out that those fluctuations can be expressed by a correction factor, $f_{\rm geo}$, which is independent of centrality but varies with the collision type. Then the pure scaled variance now representing only dynamical fluctuations, i.e., after subtraction of the geometrical component, can be calculated from the following equation [1]:

$$\omega_{\rm dyn} - 1 = f_{\rm geo}(\omega - 1). \tag{5}$$

Also, parameter k changes to $k_{\rm dyn}$ accordingly:

$$k_{\rm dyn}^{-1} = f_{\rm geo} k^{-1}. \tag{6}$$

In this analysis the hypothesis that the charged particle multiplicities measured in ultrarelativistic heavy-ion collisions are distributed according to the NBD is verified with the use of the maximum likelihood method (ML) and the likelihood ratio test. More details of this approach can be found in Refs. [4–6].

There are two crucial reasons for this approach:

(i) The fitted quantity is a probability distribution function (PDF), so the most natural way is to use the ML method, where the likelihood function is constructed directly from the tested PDF. In fact, what is fitted are parameters of the distribution. The fitted values are the *estimators* of these parameters. It is well known in mathematical statistics that an ML estimator is consistent, asymptotically unbiased, and efficient [4,5,7], but even more important is that because of Wilks's theorem (see Appendix C) one can easily define a statistic, the distribution of which converges to a χ^2 distribution as the number of measurements goes to infinity. Thus for the large sample the goodness of fit can be expressed as a P value computed with the corresponding χ^2 distribution.

(ii) The most commonly used method, the least-squares (LS) method (called also the χ^2 minimization), has the disadvantage of providing only the qualitative measure of the significance of the fit, in general. Only if observables are represented by Gaussian random variables with known variances, the conclusion about the goodness of fit equivalent to that mentioned in point (i) can be derived (see Appendix B).

[1]Here, these quantities are distinguished from the experimentally measured average charged particle multiplicity $\langle N_{\rm ch}\rangle$ and the variance σ^2.

TABLE I. Results of fitting multiplicity distributions measured by the PHENIX Collaboration in Au-Au collisions at $\sqrt{s_{NN}} = 200$ GeV, $f_{geo} = 0.37 \pm 0.027$ [1]. Fitting ranges are limited to the bins with $n_i > 5$, where n_i is the number of events in the ith bin.

Centrality (%)	N	$\hat{k}$	$\hat{n}$	$1/\hat{k}_{dyn}$	ω_{dyn}	χ^2_λ/n_d χ^2_λ (n_d)	P value (%)	χ^2_{PHEN}/n_d χ^2_{PHEN}	P value (%)
0–5	653 145	270.0 ±2.5	61.85 ±0.01	1.37×10^{-3} $\pm 0.10 \times 10^{-3}$	1.08 ±0.01	23.73 1756.0 (74)	0	0.98 72.36	0
5–10	657 944	163.4 ±1.2	53.91 ±0.01	2.26×10^{-3} $\pm 0.17 \times 10^{-3}$	1.12 ±0.01	9.12 592.7 (65)	0	0.69 44.95	0
10–15	658 739	112.5 ±0.7	46.50 ±0.01	3.29×10^{-3} $\pm 0.24 \times 10^{-3}$	1.15 ±0.01	11.5 795.5 (69)	0	0.66 45.43	0
15–20	659 607	85.1 ±0.5	39.72 ±0.01	4.35×10^{-3} $\pm 0.32 \times 10^{-3}$	1.17 ±0.01	8.9 585.8 (66)	0	0.52 34.20	0
20–25	658 785	67.6 ±0.4	33.56 ±0.01	5.48×10^{-3} $\pm 0.40 \times 10^{-3}$	1.18 ±0.01	13.5 848.8 (63)	0	0.46 29.01	0
25–30	659 632	56.7 ±0.3	28.01 ±0.01	6.52×10^{-3} $\pm 0.48 \times 10^{-3}$	1.18 ±0.01	10.9 640.6 (59)	0	0.37 22.10	0
30–35	659 303	47.4 ±0.3	23.02 ±0.01	7.81×10^{-3} $\pm 0.57 \times 10^{-3}$	1.18 ±0.01	7.9 429.9 (54)	0	0.31 16.72	0
35–40	661 174	40.5 ±0.2	18.64 ±0.01	9.13×10^{-3} $\pm 0.67 \times 10^{-3}$	1.17 ±0.01	8.5 389.7 (46)	0	0.37 17.21	0
40–45	661 599	34.0 ±0.2	14.84 ±0.01	1.09×10^{-2} $\pm 0.80 \times 10^{-3}$	1.16 ±0.01	7.3 301.0 (41)	0	0.35 14.34	0
45–50	661 765	27.3 ±0.2	11.57 ±0.005	1.35×10^{-2} $\pm 0.99 \times 10^{-3}$	1.16 ±0.01	10.5 390.2 (37)	0	0.92 34.19	0
50–55	662 114	21.3 ±0.1	8.82 ±0.004	1.74×10^{-2} $\pm 0.13 \times 10^{-2}$	1.15 ±0.01	38.8 1436.4 (37)	0	12.06 446.2	0

It is worth noting that the ML method with binned data and Poisson fluctuations within a bin was already applied to fitting multiplicity distributions to the NBD but at much lower energies (E-802 Collaboration [8]).

II. LIKELIHOOD RATIO TEST

The number of charged particles N_{ch} is assumed to be a random variable with the PDF given by Eq. (1). Each event is treated as an independent observation of N_{ch} and a set of a given class of events is a sample. For N events in the class there are N measurements of N_{ch}, for example, $\mathbf{X} = \{X_1, X_2, \dots, X_N\}$. Some of these measurements can be equal; i.e., $X_i = X_j$ for $i \neq j$ can happen. The whole population consists of all possible events with the measurements of 0, 1, 2, ... charged particles and by definition is infinite.[2]

Let us divide the sample into m bins characterized by Y_i—the number of measured charged particles[3]—and n_i—the number of entries in the ith bin, $N = \sum_{i=1}^{m} n_i$ (details of the theoretical framework of this section can be found in Refs. [4–6]). Then the expectation value of the number of events in the ith bin can be written as

$$\nu_i(\nu_{tot}, p, k) = \nu_{tot} P(Y_i; p, k), \tag{7}$$

[2]Precisely, because of the energy conservation the number of produced charged particles is limited but the number of collisions is not.

[3]Now $Y_i \neq Y_j$ for $i \neq j$ and $i, j = 1, 2, \dots, m$.

where ν_{tot} is the expected number of all events in the sample, $\nu_{\text{tot}} = \sum_{i=1}^{m} \nu_i$. This is because one can treat the number of events in the sample N also as a random variable with its own distribution—a Poisson one. Generally, the whole histogram can be treated as one measurement of m-dimensional random vector $\mathbf{n} = (n_1, \ldots, n_m)$ which has a multinomial distribution, so the joint PDF for the measurement of N and $\mathbf{n}$ can be converted to the form [4,6]

$$f(\mathbf{n}; \nu_1, \ldots, \nu_m) = \prod_{i=1}^{m} \frac{\nu_i^{n_i}}{n_i!} \exp(-\nu_i). \tag{8}$$

Since now $f(\mathbf{n}; \nu_1, \ldots, \nu_m)$ is the PDF for one measurement, f is also the likelihood function:

$$L(\mathbf{n}|\nu_1, \ldots, \nu_m) = f(\mathbf{n}; \nu_1, \ldots, \nu_m). \tag{9}$$

With the use of Eq. (7) the corresponding likelihood function can be written as

$$L(\mathbf{n}|\nu_{\text{tot}}, p, k) = L(\mathbf{n}|\nu_1(\nu_{\text{tot}}, p, k), \ldots, \nu_m(\nu_{\text{tot}}, p, k)). \tag{10}$$

Then the likelihood ratio is defined as

$$\lambda = \frac{L(\mathbf{n}|\hat{\nu}_{\text{tot}}, \hat{p}, \hat{k})}{L(\mathbf{n}|\check{\nu}_1, \ldots, \check{\nu}_m)} = \frac{L(\mathbf{n}|\hat{\nu}_{\text{tot}}, \hat{p}, \hat{k})}{L(\mathbf{n}|n_1, \ldots, n_m)}, \tag{11}$$

where $\hat{\nu}_{\text{tot}}$, $\hat{p}$, and $\hat{k}$ are the ML estimates of ν_{tot}, p, and k, respectively, with the likelihood function given by Eq. (10); and $\check{\nu}_i = n_i$, $i = 1, 2, \ldots m$ are the ML estimates of ν_i treated as free parameters. Note that since the denominator in Eq. (11) does not depend on parameters the log ratio

TABLE II. Results of fitting multiplicity distributions measured by the PHENIX Collaboration in Au-Au collisions at $\sqrt{s_{NN}} = 200$ GeV, $f_{\text{geo}} = 0.37 \pm 0.027$ [1]. Fitting ranges are limited to the bins with $n_i > 60$, where n_i is the number of events in the ith bin.

Centrality (%)	N	$\hat{k}$	$\hat{\bar{n}}$	$1/\hat{k}_{\text{dyn}}$	ω_{dyn}	χ^2_λ/n_d χ^2_λ (n_d)	P value (%)	χ^2_{PHEN}/n_d χ^2_{PHEN}	P value (%)
0–5	652 579	289.0 ±2.9	61.86 ±0.01	1.28×10^{-3} $\pm 0.94 \times 10^{-4}$	1.08 ±0.01	20.0 1160.2 (58)	0	0.57 32.86	0
5–10	657 571	168.1 ±1.2	53.91 ±0.01	2.20×10^{-3} $\pm 0.16 \times 10^{-3}$	1.12 ±0.01	20.56 1151.6 (56)	0	0.61 34.41	0
10–15	658 258	116.4 ±0.7	46.50 ±0.01	3.18×10^{-3} $\pm 0.23 \times 10^{-3}$	1.15 ±0.01	18.4 991.7 (54)	0	0.53 28.81	0
15–20	659 302	86.9 ±0.5	39.72 ±0.01	4.26×10^{-3} $\pm 0.31 \times 10^{-3}$	1.17 ±0.01	12.6 667.5 (53)	0	0.43 22.97	0
20–25	658 461	69.1 ±0.4	33.56 ±0.01	5.36×10^{-3} $\pm 0.39 \times 10^{-3}$	1.18 ±0.01	12.3 604.7 (49)	0	0.34 16.46	0
25–30	659 337	57.9 ±0.3	28.0 ±0.01	6.39×10^{-3} $\pm 0.47 \times 10^{-3}$	1.18 ±0.01	10.4 469.1 (45)	0	0.28 12.80	6.7×10^{-8}
30–35	659 021	48.3 ±0.3	23.02 ±0.01	7.66×10^{-3} $\pm 0.56 \times 10^{-3}$	1.18 ±0.01	8.6 351.02 (41)	0	0.16 6.62	0.76
35–40	660 937	41.3 ±0.2	18.64 ±0.01	8.96×10^{-3} $\pm 0.66 \times 10^{-3}$	1.17 ±0.01	7.6 280.3 (37)	0	0.19 6.85	0.12
40–45	661 422	34.6 ±0.2	14.84 ±0.01	1.07×10^{-2} $\pm 0.78 \times 10^{-3}$	1.16 ±0.01	7.9 260.3 (33)	0	0.21 7.06	0.015
45–50	661 577	27.9 ±0.2	11.56 ±0.005	1.33×10^{-2} $\pm 0.97 \times 10^{-3}$	1.15 ±0.01	10.0 279.9 (28)	0	0.23 6.44	0.011
50–55	661 877	21.9 ±0.1	8.81 ±0.004	1.69×10^{-2} $\pm 0.12 \times 10^{-2}$	1.15 ±0.01	40.0 959.2 (24)	0	0.30 7.29	7.8×10^{-5}

defined as

$$\ln\lambda(\nu_{\rm tot}, p, k) = \ln\frac{L(\mathbf{n}|\nu_{\rm tot}, p, k)}{L(\mathbf{n}|n_1,\ldots,n_m)} = -\sum_{i=1}^{m}\left(n_i \ln\frac{n_i}{\nu_i} + \nu_i - n_i\right) = -\nu_{\rm tot} + N - \sum_{i=1}^{m} n_i \ln\frac{n_i}{\nu_i}, \tag{12}$$

where ν_i are expressed by Eq. (7), can be used to find the ML estimates of $\nu_{\rm tot}$, p, and k. The values $\hat{\nu}_{\rm tot}$, $\hat{p}$, and $\hat{k}$ for which $\lambda(\nu_{\rm tot}, p, k)$ has its maximum are the maximum likelihood estimates of parameters $\nu_{\rm tot}$, p, and k. Then one can define the test statistic called "likelihood χ^2" [6]:

$$\chi^2_\lambda = -2\ln\lambda(\nu_{\rm tot}, p, k) = 2\sum_{i=1}^{m}\left(\nu_i - n_i + n_i \ln\frac{n_i}{\nu_i}\right). \tag{13}$$

Note that the maximum of $\ln\lambda$ is the minimum of χ^2_λ, so the estimates from the condition of the minimum of χ^2_λ are the ML estimates. Further, the statistic given by

$$\chi^2_{\lambda,\rm min} = -2\ln\lambda(\hat{\nu}_{\rm tot}, \hat{p}, \hat{k}) = 2\sum_{i=1}^{m}\left(n_i \ln\frac{n_i}{\hat{\nu}_i} + \hat{\nu}_i - n_i\right) \tag{14}$$

approaches a χ^2 distribution asymptotically, i.e., as the number of measurements, here the number of events N, goes to infinity (the consequence of the Wilks's theorem, see Appendix C). The values $\hat{\nu}_i$ are the estimates of ν_i given by

$$\hat{\nu}_i = \hat{\nu}_{\rm tot} P(Y_i; \hat{p}, \hat{k}), \tag{15}$$

and if one assumes that $\nu_{\rm tot}$ does not depend on p and k then $\hat{\nu}_{\rm tot} = N$. For such a case

$$\sum_{i=1}^{m}\hat{\nu}_i = \sum_{i=1}^{m} n_i, \tag{16}$$

TABLE III. Results of fitting multiplicity distributions measured by the PHENIX Collaboration in Au-Au collisions at $\sqrt{s_{NN}} = 62.4$ GeV, $f_{\rm geo} = 0.33 \pm 0.031$ [1]. Fitting ranges are limited to the bins with $n_i > 5$, where n_i is the number of events in the ith bin.

Centrality (%)	N	$\hat{k}$	$\hat{n}$	$1/\hat{k}_{\rm dyn}$	$\omega_{\rm dyn}$	χ^2_λ/n_d χ^2_λ (n_d)	P value (%)	$\chi^2_{\rm PHEN}/n_d$ $\chi^2_{\rm PHEN}$	P value (%)
0–5	607 155	225.2 ±2.5	44.67 ±0.01	1.47×10^{-3} $\pm 0.14 \times 10^{-3}$	1.07 ±0.01	2.37 139.6 (59)	1.7×10^{-8}	0.18 10.65	0.015
5–10	752 392	142.3 ±1.1	37.96 ±0.01	2.32×10^{-3} $\pm 0.22 \times 10^{-3}$	1.09 ±0.01	2.44 131.9 (54)	1.9×10^{-8}	0.11 5.91	29.3
10–15	752 837	115.2 ±0.9	31.53 ±0.01	2.87×10^{-3} $\pm 0.27 \times 10^{-3}$	1.09 ±0.01	2.06 107.1 (52)	1.1×10^{-5}	0.13 6.88	6.0
15–20	752 553	88.0 ±0.6	26.07 ±0.01	3.75×10^{-3} $\pm 0.35 \times 10^{-3}$	1.10 ±0.01	1.86 87.3 (47)	3.2×10^{-4}	0.13 5.98	9.9
20–25	752 296	68.5 ±0.5	21.35 ±0.01	4.82×10^{-3} $\pm 0.45 \times 10^{-3}$	1.10 ±0.01	2.63 113.2 (43)	3.1×10^{-8}	0.21 9.10	2.7×10^{-3}
25–30	752 183	53.2 ±0.4	17.30 ±0.01	6.21×10^{-3} $\pm 0.59 \times 10^{-3}$	1.11 ±0.01	2.75 107.3 (39)	2.7×10^{-8}	0.23 8.81	1.2×10^{-3}
30–35	751 375	40.1 ±0.3	13.84 ±0.005	8.22×10^{-3} $\pm 0.77 \times 10^{-3}$	1.11 ±0.01	2.97 103.9 (35)	9.6×10^{-9}	0.25 8.65	3.0×10^{-4}
35–40	751 661	31.7 ±0.2	10.89 ±0.004	1.04×10^{-2} $\pm 0.98 \times 10^{-3}$	1.11 ±0.01	6.72 194.9 (29)	0	0.16 4.54	2.7
40–45	750 884	25.1 ±0.2	8.42 ±0.004	1.31×10^{-2} $\pm 0.12 \times 10^{-2}$	1.11 ±0.01	37.5 937.4 (25)	0	40.36 1009.1	0
45–50	751 421	21.8 ±0.2	6.41 ±0.003	1.51×10^{-2} $\pm 0.14 \times 10^{-2}$	1.10 ±0.01	209.0 4806.8 (23)	0	285.9 6576.7	0

and Eq. (14) becomes

$$\chi^2_{\lambda,\min}(\hat{p},\hat{k}) = 2\sum_{i=1}^{m} n_i \ln\frac{n_i}{\hat{\nu}_i}. \tag{17}$$

Also, then one can write $\nu_{\rm tot} = N$ and Eq. (12) can be rewritten as

$$\begin{aligned}\ln\lambda(p,k) &= N\ln N - \sum_{i=1}^{m} n_i \ln n_i + \sum_{i=1}^{m} n_i \ln P(Y_i;p,k)\\ &= -\sum_{i=1}^{m} n_i \ln\frac{n_i}{N} + N\sum_{i=1}^{m}\frac{n_i}{N}\ln P(Y_i;p,k)\\ &= -N\sum_{i=1}^{m} P_i^{\rm ex}\ln P_i^{\rm ex} + N\sum_{i=1}^{m} P_i^{\rm ex}\ln P(Y_i;p,k),\end{aligned} \tag{18}$$

where $P_i^{\rm ex} = n_i/N$. Thus, with the help of Eqs. (14) and (18) one arrives at

$$\chi^2_{\lambda,\min} = 2\,N\sum_{i=1}^{m} P_i^{\rm ex}\ln\frac{P_i^{\rm ex}}{P(Y_i;\hat{p},\hat{k})}. \tag{19}$$

It can be proven that one of the necessary conditions for the existence of the maximum is (see Appendix A for details)

$$\bar{n} = \langle N_{\rm ch}\rangle, \tag{20}$$

i.e., the distribution average has to be equal to the experimental average. This is very good because $\langle N_{\rm ch}\rangle$ is what is called in statistics *a sample mean*. The sample mean is an estimator for the expectation value of the random variable, which is consistent and unbiased [4]. In other words the ML estimator of $\bar{n}$ is $\langle N_{\rm ch}\rangle$ ($\hat{\bar{n}} = \langle N_{\rm ch}\rangle$).

TABLE IV. Results of fitting multiplicity distributions measured by the PHENIX Collaboration in Au-Au collisions at $\sqrt{s_{NN}} = 62.4$ GeV, $f_{\rm geo} = 0.33 \pm 0.031$ [1]. Fitting ranges are limited to the bins with $n_i > 40$, where n_i is the number of events in the ith bin.

Centrality (%)	N	$\hat{k}$	$\hat{\bar{n}}$	$1/\hat{k}_{\rm dyn}$	$\omega_{\rm dyn}$	χ^2_λ/n_d χ^2_λ (n_d)	P value (%)	$\chi^2_{\rm PHEN}/n_d$ $\chi^2_{\rm PHEN}$	P value (%)
0–5	607 075	227.9 ±2.5	44.67 ±0.01	1.45×10^{-3} $\pm 0.14 \times 10^{-3}$	1.06 ±0.01	5.55 294.3 (53)	0	0.19 10.2	5.6×10^{-3}
5–10	752 263	143.9 ±1.1	37.96 ±0.01	2.29×10^{-3} $\pm 0.22 \times 10^{-3}$	1.09 ±0.01	7.80 382.4 (49)	0	0.12 5.95	14.4
10–15	752 739	116.2 ±0.9	31.53 ±0.01	2.84×10^{-3} $\pm 0.27 \times 10^{-3}$	1.09 ±0.01	5.67 260.8 (46)	0	0.13 6.08	7.0
15–20	752 492	88.5 ±0.6	26.07 ±0.01	3.73×10^{-3} $\pm 0.35 \times 10^{-3}$	1.10 ±0.01	5.97 250.9 (42)	0	0.11 4.60	30.9
20–25	752 182	69.2 ±0.5	21.35 ±0.01	4.77×10^{-3} $\pm 0.45 \times 10^{-3}$	1.10 ±0.01	10.2 377.2 (37)	0	0.22 8.27	2.4×10^{-3}
25–30	752 095	53.6 ±0.4	17.30 ±0.01	6.16×10^{-3} $\pm 0.58 \times 10^{-3}$	1.11 ±0.01	8.2 279.2 (34)	0	0.23 7.92	1.8×10^{-3}
30–35	751 324	40.3 ±0.3	13.84 ±0.005	8.19×10^{-3} $\pm 0.77 \times 10^{-3}$	1.11 ±0.01	7.40 229.3 (31)	0	0.26 7.92	4.3×10^{-4}
35–40	751 639	31.8 ±0.2	10.89 ±0.004	1.04×10^{-2} $\pm 0.98 \times 10^{-3}$	1.11 ±0.01	9.43 254.7 (27)	0	0.15 4.17	3.5
40–45	750 852	25.2 ±0.2	8.42 ±0.004	1.31×10^{-2} $\pm 0.12 \times 10^{-2}$	1.11 ±0.01	50.7 1166.3 (23)	0	0.22 5.13	0.062
45–50	751 348	22.0 ±0.2	6.41 ±0.003	1.50×10^{-2} $\pm 0.14 \times 10^{-2}$	1.10 ±0.01	259.8 4936.4 (19)	0	343.1 6519.1	0

TABLE V. Results of fitting multiplicity distributions measured by the PHENIX Collaboration in Cu-Cu collisions at $\sqrt{s_{NN}} = 200$ GeV, $f_{\text{geo}} = 0.40 \pm 0.047$ [1]. Fitting ranges are limited to the bins with $n_i > 5$, where n_i is the number of events in the ith bin.

Centrality (%)	N	$\hat{k}$	$\hat{\bar{n}}$	$1/\hat{k}_{\text{dyn}}$	ω_{dyn}	χ^2_λ/n_d χ^2_λ (n_d)	P value (%)	χ^2_{PHEN}/n_d χ^2_{PHEN}	P value (%)
0–5	368 510	59.6 $\pm$0.6	19.80 $\pm$0.01	6.72×10^{-3} $\pm 0.79 \times 10^{-3}$	1.13 $\pm$0.02	94.8 3887.0 (41)	0	2.1 87.1	0
5–10	369 206	49.6 $\pm$0.5	16.74 $\pm$0.01	8.06×10^{-3} $\pm 0.95 \times 10^{-3}$	1.13 $\pm$0.02	16.5 628.5 (38)	0	0.66 25.3	0
10–15	369 945	41.5 $\pm$0.4	14.05 $\pm$0.01	9.64×10^{-3} $\pm 0.11 \times 10^{-2}$	1.14 $\pm$0.02	6.8 225.5 (33)	0	0.38 12.6	0
15–20	370 066	34.5 $\pm$0.3	11.78 $\pm$0.01	1.16×10^{-2} $\pm 0.14 \times 10^{-2}$	1.14 $\pm$0.02	3.0 92.0 (31)	5.8×10^{-8}	0.24 7.53	1.5×10^{-3}
20–25	371 877	29.2 $\pm$0.3	9.81 $\pm$0.01	1.37×10^{-2} $\pm 0.16 \times 10^{-2}$	1.13 $\pm$0.02	6.6 186.0 (28)	0	3.4 93.9	0
25–30	368 876	24.9 $\pm$0.2	8.14 $\pm$0.01	1.60×10^{-2} $\pm 0.19 \times 10^{-2}$	1.13 $\pm$0.02	19.3 502.4 (26)	0	11.5 298.9	0
30–35	368 072	21.9 $\pm$0.2	6.72 $\pm$0.005	1.83×10^{-2} $\pm 0.22 \times 10^{-2}$	1.12 $\pm$0.01	65.6 1704.8 (26)	0	42.3 1098.5	0

TABLE VI. Results of fitting multiplicity distributions measured by the PHENIX Collaboration in Cu-Cu collisions at $\sqrt{s_{NN}} = 200$ GeV, $f_{\text{geo}} = 0.40 \pm 0.047$ [1]. Fitting ranges are limited to the bins with $n_i > 80$, where n_i is the number of events in the ith bin.

Centrality (%)	N	$\hat{k}$	$\hat{\bar{n}}$	$1/\hat{k}_{\text{dyn}}$	ω_{dyn}	χ^2_λ/n_d χ^2_λ (n_d)	P value (%)	χ^2_{PHEN}/n_d χ^2_{PHEN}	P value (%)
0–5	368 271	61.5 $\pm$0.6	19.79 $\pm$0.01	6.50×10^{-3} $\pm 0.77 \times 10^{-3}$	1.13 $\pm$0.02	122.2 4398.3 (36)	0	2.3 82.7	0
5–10	368 869	52.0 $\pm$0.5	16.74 $\pm$0.01	7.69×10^{-3} $\pm 0.91 \times 10^{-3}$	1.13 $\pm$0.02	20.5 613.9 (30)	0	0.39 11.7	0
10–15	369 825	42.3 $\pm$0.4	14.05 $\pm$0.01	9.46×10^{-3} $\pm 0.11 \times 10^{-2}$	1.13 $\pm$0.02	16.2 470.9 (29)	0	0.43 12.6	0
15–20	369 964	35.1 $\pm$0.3	11.77 $\pm$0.01	1.14×10^{-2} $\pm 0.13 \times 10^{-2}$	1.13 $\pm$0.02	11.4 296.8 (26)	0	0.24 6.36	5.4×10^{-3}
20–25	371 752	29.8 $\pm$0.3	9.80 $\pm$0.01	1.34×10^{-2} $\pm 0.16 \times 10^{-2}$	1.13 $\pm$0.02	16.1 370.4 (23)	0	0.20 4.51	0.38
25–30	368 708	25.6 $\pm$0.3	8.14 $\pm$0.01	1.56×10^{-2} $\pm 0.18 \times 10^{-2}$	1.13 $\pm$0.01	42.7 853.2 (20)	0	0.21 4.27	0.23
30–35	367 869	22.6 $\pm$0.2	6.72 $\pm$0.005	1.77×10^{-2} $\pm 0.21 \times 10^{-2}$	1.12 $\pm$0.01	126.4 2274.4 (18)	0	0.62 11.1	0

TABLE VII. Results of fitting multiplicity distributions measured by the PHENIX Collaboration in Cu-Cu collisions at $\sqrt{s_{NN}} = 62.4$ GeV, $f_{\text{geo}} = 0.32 \pm 0.063$ [1]. Fitting ranges are limited to the bins with $n_i > 5$, where n_i is the number of events in the ith bin.

Centrality (%)	N	$\hat{k}$	$\hat{n}$	$1/\hat{k}_{\text{dyn}}$	ω_{dyn}	χ^2_λ/n_d χ^2_λ (n_d)	P value (%)	χ^2_{PHEN}/n_d χ^2_{PHEN}	P value (%)
0–5	298 182	41.6 ± 0.4	13.35 ± 0.01	7.69×10^{-3} $\pm 0.15 \times 10^{-2}$	1.10 ± 0.02	9.3 279.9 (30)	0	0.65 19.4	0
5–10	307 150	26.5 ± 0.2	11.67 ± 0.01	1.21×10^{-2} $\pm 0.24 \times 10^{-2}$	1.14 ± 0.03	9.7 290.7 (30)	0	0.78 23.3	0
10–15	309 874	20.5 ± 0.2	9.90 ± 0.01	1.56×10^{-2} $\pm 0.31 \times 10^{-2}$	1.15 ± 0.03	9.3 261.1 (28)	0	4.4 122.5	0
15–20	312 530	17.8 ± 0.1	8.27 ± 0.01	1.80×10^{-2} $\pm 0.36 \times 10^{-2}$	1.15 ± 0.03	26.0 677.1 (26)	0	31.6 821.7	0
20–25	312 884	16.0 ± 0.1	6.89 ± 0.01	1.99×10^{-2} $\pm 0.39 \times 10^{-2}$	1.14 ± 0.03	75.8 1744.0 (23)	0	80.9 1861.4	0

III. RESULTS AND DISCUSSION

The method described in Sec. II requires that all bins in a given data set have widths equal to 1, so that the experimental probability P_i^{ex} to measure a signal in the ith bin was equivalent to the probability of the measurement of $(i-1)$ charged particles if the first bin is the bin of zero charged particles detected. This is fulfilled for all bins of the considered data sets.

Since the test statistic $\chi^2_{\lambda,\min}$ has a χ^2 distribution approximately in the large sample limit, it can be used as a test of the goodness of fit. The result of the test is given by the so-called P value, which is the probability of obtaining the value of the statistic, Eq. (14), equal to or greater than the value just obtained for the present data set, when repeating the whole experiment many times (see Appendix B):

$$P = P\left(\chi^2 \geqslant \chi^2_{\lambda,\min}; n_d\right) = \int_{\chi^2_{\lambda,\min}}^{\infty} f(z; n_d) dz, \qquad (21)$$

where $f(z; n_d)$ is the χ^2 PDF and n_d is the number of degrees of freedom, $n_d = m - 2$ here.

TABLE VIII. Results of fitting multiplicity distributions measured by the PHENIX Collaboration in Cu-Cu collisions at $\sqrt{s_{NN}} = 62.4$ GeV, $f_{\text{geo}} = 0.32 \pm 0.063$ [1]. Fitting ranges are limited to the bins with $n_i > 60$, where n_i is the number of events in the ith bin.

Centrality (%)	N	$\hat{k}$	$\hat{n}$	$1/\hat{k}_{\text{dyn}}$	ω_{dyn}	χ^2_λ/n_d χ^2_λ (n_d)	P value (%)	χ^2_{PHEN}/n_d χ^2_{PHEN}	P value (%)
0–5	298 131	42.0 ± 0.5	13.35 ± 0.01	7.62×10^{-3} $\pm 0.15 \times 10^{-2}$	1.10 ± 0.02	14.7 411.9 (28)	0	0.67 18.9	0
5–10	307 061	26.8 ± 0.2	11.66 ± 0.01	1.19×10^{-2} $\pm 0.24 \times 10^{-2}$	1.14 ± 0.03	19.7 512.5 (26)	0	0.86 22.5	0
10–15	309 798	20.7 ± 0.2	9.90 ± 0.01	1.54×10^{-2} $\pm 0.30 \times 10^{-2}$	1.15 ± 0.03	19.4 465.5 (24)	0	0.38 9.08	1.1×10^{-7}
15–20	312 434	18.0 ± 0.1	8.27 ± 0.01	1.78×10^{-2} $\pm 0.35 \times 10^{-2}$	1.15 ± 0.03	46.5 976.4 (21)	0	0.40 8.37	1.9×10^{-7}
20–25	312 758	16.3 ± 0.1	6.89 ± 0.01	1.96×10^{-2} $\pm 0.39 \times 10^{-2}$	1.14 ± 0.03	118.1 2243.4 (19)	0	0.63 12.05	0

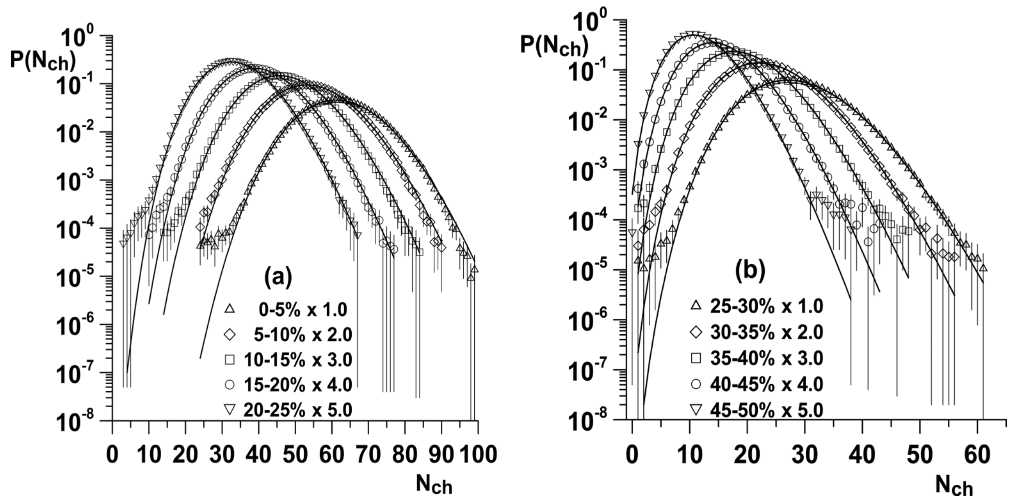

FIG. 1. Uncorrected multiplicity distributions of charged hadrons for 200-GeV Au-Au collisions [1] within ranges limited to the bins with $n_i > 5$. The lines are fits to the NBD. The data are scaled by the amounts in the legend. Errors represent the statistical and systematic errors added in quadrature.

The results of the analysis are presented in Tables I–VIII and illustrated with Figs. 1–6. In fact, the whole analysis was done for the two kinds of histograms: (i) bins with the number of entries $n_i \leqslant 5$ excluded (Tables I, III, V, and VII) and (ii) bins with the number of entries $n_i \leqslant 40$ (Table IV), $n_i \leqslant 60$ (Tables II and VIII), or $n_i \leqslant 80$ (Table VI) excluded. In practice this corresponds to cutting off (i) less or (ii) more the tails of the full measured histograms. The tails break the visual agreement between the data and the NBD (cf. Figs. 1 and 2). The condition that only bins with $n_i > 5$ are taken into account is the minimal condition imposed on a histogram to do any statistical inference without Monte Carlo simulations [4]. Condition (ii) corresponds roughly to the choice made originally by the PHENIX Collaboration in their analysis [1]. It has turned out that the results of this analysis are qualitatively the same for both choices.

As one can see, the hypothesis in question should be rejected in all considered cases, but it was claimed that charged particle multiplicities measured in Au-Au and Cu-Cu collisions at $\sqrt{s_{NN}} = 62.4$ and 200 GeV are distributed according to the NBD [1]. However, that conclusion was the result of the application of the LS method. Therefore, it seems to be reasonable to check what are the values of the LS test statistic at the ML estimators listed in the third and fourth columns of Tables I–VIII. For the sample described in Sec. II one can define the LS test statistic (commonly called the χ^2 function) as

$$\chi^2_{\mathrm{LS}}(\mathbf{n};\bar{n},k) = \sum_{i=1}^{m} \frac{(n_i - \nu_i(\bar{n},k))^2}{\mathrm{err}^2_{n_i}} = \sum_{i=1}^{m} \frac{\left(P_i^{\mathrm{ex}} - P(Y_i;\bar{n},k)\right)^2}{\mathrm{err}^2_i}, \tag{22}$$

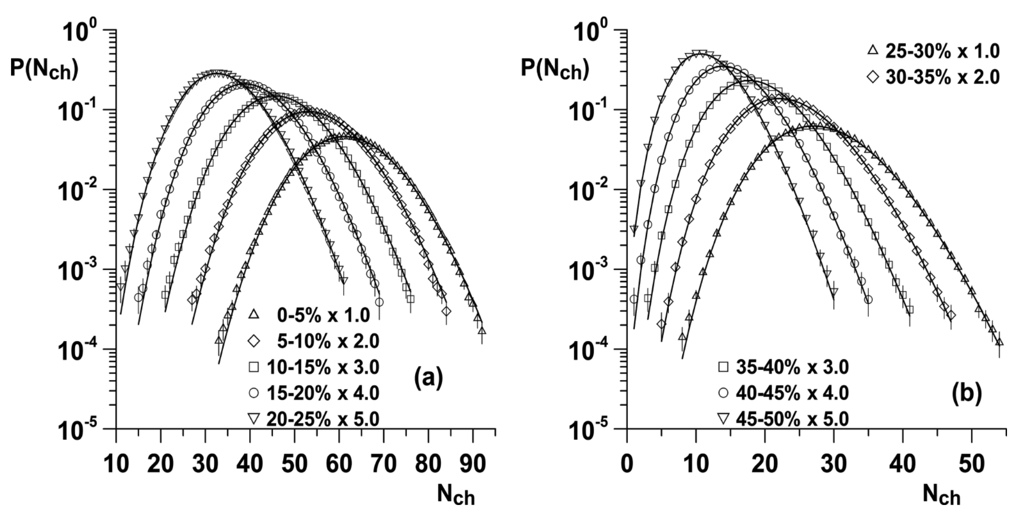

FIG. 2. Uncorrected multiplicity distributions of charged hadrons for 200-GeV Au-Au collisions [1] within ranges limited to the bins with $n_i > 60$. The lines are fits to the NBD. The data are scaled by the amounts in the legend. Errors represent the statistical and systematic errors added in quadrature.

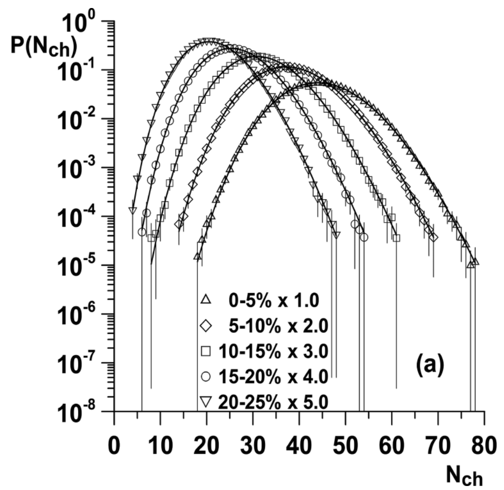

FIG. 3. Uncorrected multiplicity distributions of charged hadrons for 62.4-GeV Au-Au collisions [1] within ranges limited to the bins with $n_i > 5$. The lines are fits to the NBD. The data are scaled by the amounts in the legend. Errors represent the statistical and systematic errors added in quadrature.

where $\nu_i(\bar{n}, k)$ is given by Eq. (7) with $\nu_{\rm tot} = N$ and ${\rm err}_{n_i}$ (${\rm err}_i = {\rm err}_{n_i}/N$) is the uncertainty on n_i ($P_i^{\rm ex}$, respectively). Note that for ${\rm err}_{n_i}^2 = \nu_i$ the above equation is Pearson's χ^2 test statistic, whereas for ${\rm err}_{n_i}^2 = n_i$ this is Neyman's χ^2 test statistic (also called the modified chi-square or modified least-squares method), both well known in mathematical statistics [4–6,10]. The advantage of the use of these statistics is that both follow a χ^2 distribution asymptotically. The errors given by $\sqrt{\nu_i}$ or $\sqrt{n_i}$ are interpreted as theoretical or experimental statistical errors, respectively (for the discussion of the pros and cons of both see [4,9]). It should be stressed that when ${\rm err}_{n_i}$ includes also a systematic error (e.g., by adding in quadrature to a statistical one) then the statement about the asymptotic form of the distribution of the test statistic is no longer valid.

In the present analysis $\chi^2_{\rm LS}$ function, Eq. (22), *is not minimized* with respect to $\bar{n}$ and k (or p and k) as in the LS method but is calculated at ML estimates of $\bar{n}$ and k. Generally, this is allowed in statistics and is equivalent to test a single point in the parameter space. Then the tested point might not be the best estimate of the true value but the hypothesis in question becomes the hypothesis only about a particular distribution (a *simple* hypothesis). At first sight, $\chi^2_{\rm LS}/n_d$ values of the ninth column of Tables I–VIII seem to be significant for almost all centrality classes, which agrees with the results of Ref. [1], but this contradicts the results of the likelihood ratio test, which are expressed by χ^2_λ/n_d and P values listed in the seventh and eight columns of Tables I–VIII. The crucial question is now why the conclusions from χ^2_λ and $\chi^2_{\rm LS}$ test statistics are entirely opposite for PHENIX measurements. The main difference between both statistics is that χ^2_λ does not depend on the actual errors but $\chi^2_{\rm LS}$ does. Additionally, χ^2_λ depends explicitly on the number of events whereas $\chi^2_{\rm LS}$ does not [cf. Eqs. (19) and (22)]. In principle, one can conclude that χ^2_λ statistic implicitly assumes errors of the type $\sqrt{n_i}$ because the statistic originated from the likelihood function, Eqs. (8) and (9), which is the product of Poisson distributions, but there is no

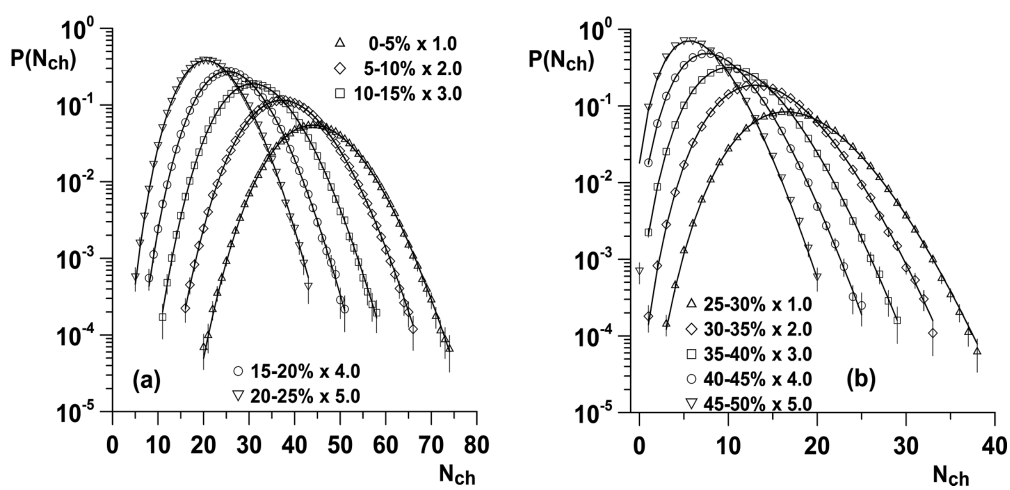

FIG. 4. Uncorrected multiplicity distributions of charged hadrons for 62.4-GeV Au-Au Collisions [1] within ranges limited to the bins with $n_i > 40$. The lines are fits to the NBD. The data are scaled by the amounts in the legend. Errors represent the statistical and systematic errors added in quadrature.

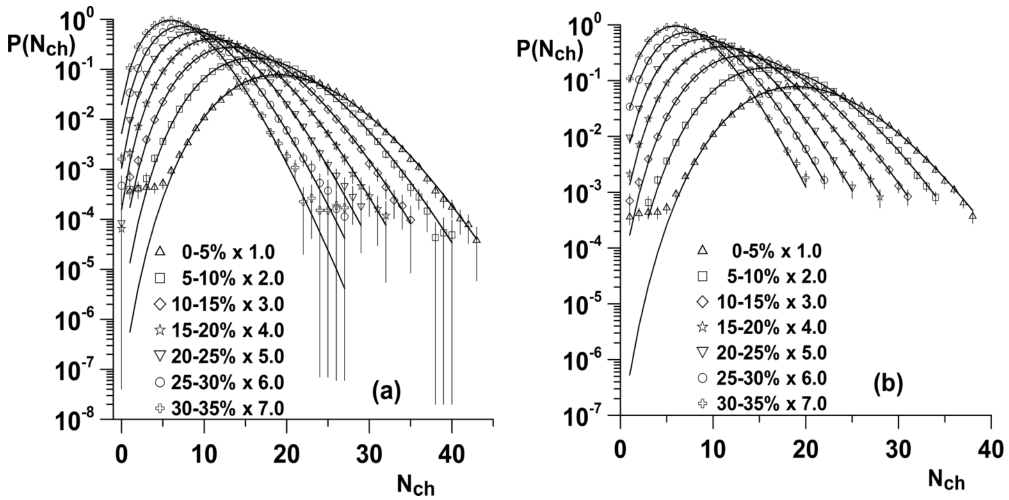

FIG. 5. Uncorrected multiplicity distributions of charged hadrons for 200-GeV Cu-Cu collisions [1] within ranges limited to the bins with $n_i > 5$ (left) and $n_i > 80$ (right). The lines are fits to the NBD. The data are scaled by the amounts in the legend. Errors represent the statistical and systematic errors added in quadrature.

place to insert actual experimental errors into the χ^2_λ statistic, Eqs. (13) and (14); this test statistic by definition does not take the experimental errors into account. Finally, the distribution of $\chi^2_{\lambda,\mathrm{min}}$ is known asymptotically, whereas the distribution of χ^2_{LS} at the minimum, when systematic errors are included, is not known, even asymptotically.

In the PHENIX analysis [1], errors err_{n_i} (err_i) in Eq. (22) are represented by the quadrature sum of the statistical and systematic components, and the statistical error on the number of entries n_i is equal to $\sqrt{n_i}$ exactly [11] (the statistical error on P_i^{ex} is $\sqrt{n_i}/N$ then). The systematic errors were mostly caused by time-dependent variation of results. Data sets were taken over spans of several days to several weeks, during which the total acceptance and efficiency were changing, mainly because of degradation of the tracking detectors [1,12]. To estimate these systematic errors, the entire data set was divided into ten subsets of approximately equal sizes. Then plots from these subsets were overlaid with each other, from which bin-by-bin systematic errors were estimated as 3.0 times the statistical errors, the same for all data sets and centralities [11,12].[4] This leads to the relation $\mathrm{err}^2_{n_i} = \sigma^2_{\mathrm{stat},n_i} + 9\sigma^2_{\mathrm{stat},n_i} = 10\sigma^2_{\mathrm{stat},n_i} = 10n_i$ ($\mathrm{err}_{n_i} = \sqrt{10} \times \sigma_{\mathrm{stat},n_i} \approx 3.0\sigma_{\mathrm{stat},n_i}$), where $\sigma_{\mathrm{stat},n_i} = \sqrt{n_i}$ is the statistical error of the ith measurement. Hence, if statistical errors only were taken into account the values of χ^2_{LS}/n_d would be ten times greater than those listed in Tables I–VIII, so it seems that the acceptance of the NBD hypothesis by χ^2_{LS} test is entirely due to the magnitude of systematic errors, but in fact this is the result of confused inference, as will be shown further.

If one inserts explicit values of PHENIX errors, $\mathrm{err}^2_{n_i} = 10n_i$, into Eq. (22), then the χ^2_{LS} test statistic takes the form called χ^2_{PHEN} from now on (the author strongly advises the reader to read Appendix B first, before going further):

$$\chi^2_{\mathrm{PHEN}}(\mathbf{n};\bar{n},k) = \sum_{i=1}^{m} \frac{(n_i - \nu_i(\bar{n},k))^2}{10n_i} = \frac{1}{10}\sum_{i=1}^{m} \frac{(n_i - \nu_i(\bar{n},k))^2}{n_i} = \frac{1}{10}\chi^2_N(\mathbf{n};\bar{n},k). \tag{23}$$

However, this is exactly Neyman's χ^2 test statistic, χ^2_N, multiplied by 0.1. Therefore, the PHENIX test statistic estimators of parameters $\bar{n}$ and k are Neyman's χ^2 estimators, $\hat{\bar{n}}_N$ and $\hat{k}_N$, respectively. Further, the distribution of Neyman's χ^2 test statistic $t_N(\mathbf{n}) \equiv \chi^2_N(\mathbf{n};\hat{\bar{n}}_N,\hat{k}_N)$ asymptotically approaches a χ^2 distribution with $n_d = m - 2$ [6,13,15]. Now, the more rigorous justification for inserting ML estimates into χ^2_{LS}, Eq. (22), can be given. The likelihood χ^2, Pearson's χ^2, and Neyman's χ^2 test statistics are asymptotically equivalent; i.e., their estimators are consistent, asymptotically normal, with the same minimum variance ([5], p. 192; [10], Sec. 18.58; [13], pp. 457 and 458). Moreover, "So far as the χ^2's considered for tests of significance are concerned, any can be used with any of the estimates" ([14], p. 464; also see p. 444). This means that, e.g., ML estimates could be put into Neyman's χ^2 test statistic and still the distribution of such a test statistic would approach a χ^2 distribution asymptotically. Since PHENIX samples are very large (see the second column in Tables I–VIII) one can reasonably approximate the distribution of $t_N(\mathbf{n})$ by the corresponding χ^2 distribution, but what is the distribution of the PHENIX test statistic $t_{\mathrm{PHEN}}(\mathbf{n}) = \chi^2_{\mathrm{PHEN,min}}(\mathbf{n}) \equiv \chi^2_{\mathrm{PHEN}}(\mathbf{n};\hat{\bar{n}}_N,\hat{k}_N) = 0.1t_N(\mathbf{n})$ then? The solution to this question can be easily found with the use of the general rule of finding the distribution $g(t)$ of a function $t(z)$ of a random variable z with the known PDF $f(z)$ ([4], p. 14):

$$g(t) = f(z(t))\left|\frac{dz}{dt}\right|, \tag{24}$$

[4]This detailed information is from Ref. [12], but there is a short note in Ref. [11] below the data for Fig. 2(c): "On average, the systematic + statistical errors are a factor of 3 larger than the statistical errors."

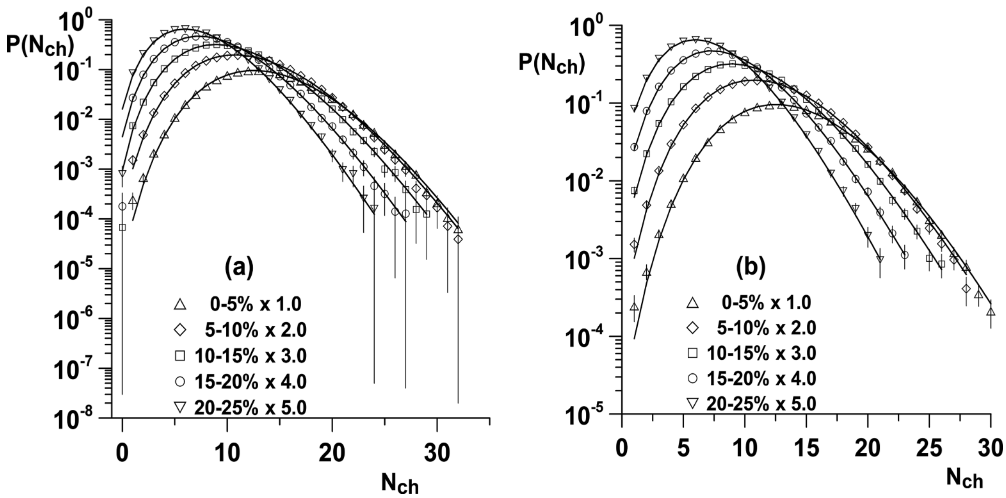

FIG. 6. Uncorrected multiplicity distributions of charged hadrons for 62.4-GeV Cu-Cu collisions [1] within ranges limited to the bins with $n_i > 5$ (left) and $n_i > 60$ (right). The lines are fits to the NBD. The data are scaled by the amounts in the legend. Errors represent the statistical and systematic errors added in quadrature.

if $t(z)$ has a unique inverse. In the present case $t(z) = 0.1z$ and $f(z) = f(z;n_d)$, so $z(t) = 10t$ and $g(t;n_d) = 10f(10t;n_d)$ is the distribution in question. The expectation value of the PHENIX test statistic is $E[t_{\rm PHEN}] = E[0.1t_N] = 0.1\cdot E[t_N] = 0.1n_d$. Thus, $E[t_{\rm PHEN}/n_d] = 0.1$ or, rewriting it in a more familiar way, $E[\chi^2_{\rm PHEN}/n_d] = 0.1$, *not* 1. Therefore, if the (PHENIX) experiment is "reasonable" and the hypothesis is true, one should expect to obtain $\chi^2_{\rm PHEN}/n_d \approx 0.1$—values of $\chi^2_{\rm PHEN}/n_d$ much greater than 0.1 suggest that the hypothesis (of the NBD) should be rejected. In the language of Appendix B, the decision boundary for the PHENIX test statistic $\chi^2_{\rm PHEN}$ should be placed at $0.1n_d$, *not* $\mathbf{n_d}$. In the case of the $\chi^2_{\rm PHEN}$ statistic the P value for the hypothesis is given by

$$P = \int_{t_{\rm PHEN}}^{\infty} g(t;n_d)dt = \int_{10\chi^2_{\rm PHEN,min}}^{\infty} f(t;n_d)dt, \qquad (25)$$

where $f(z;n_d)$ is the χ^2 PDF with n_d degrees of freedom. The corresponding values are given in the tenth column of Tables I–VIII. Altogether there are 33 classes of collision systems and centralities of the PHENIX measurements [1] considered here. They are doubled because of two possibilities of cutting tails in full histograms. The assessment of the quality of fits presented in Tables I–VIII depends on the assumed significance level. Following the choice made by the UA5 Collaboration [3], the 0.1% level is fixed here. There are eight cases in which the PHENIX test is significant at the 0.1% level at least for one of the two histograms corresponding to the same class. It is interesting that half of them belong to the case of Au-Au collisions at $\sqrt{s_{NN}} = 62.4$ GeV and are significant for both kinds of histograms with P values greater than 1% (see Tables III and IV). The next two happen for Au-Au collisions at $\sqrt{s_{NN}} = 200$ GeV (Table II), and the last two happen for Cu-Cu collisions at $\sqrt{s_{NN}} = 200$ GeV (Table VI), but only in the case of narrower histograms and with P values smaller than 1%. In contrast, the case of Cu-Cu collisions at $\sqrt{s_{NN}} = 62.4$ GeV has no any significant fit at all (see Tables VII and VIII). Thus, one

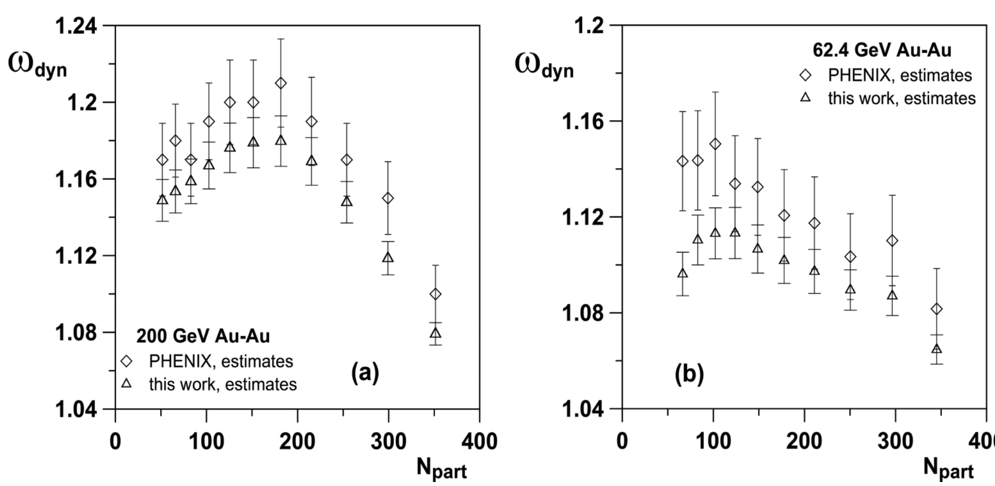

FIG. 7. Scaled variance for 200-GeV (a) and 62.4-GeV (b) Au-Au collisions. PHENIX estimates are from [1]. Estimates from this work are for the cases listed in Tables II and IV.

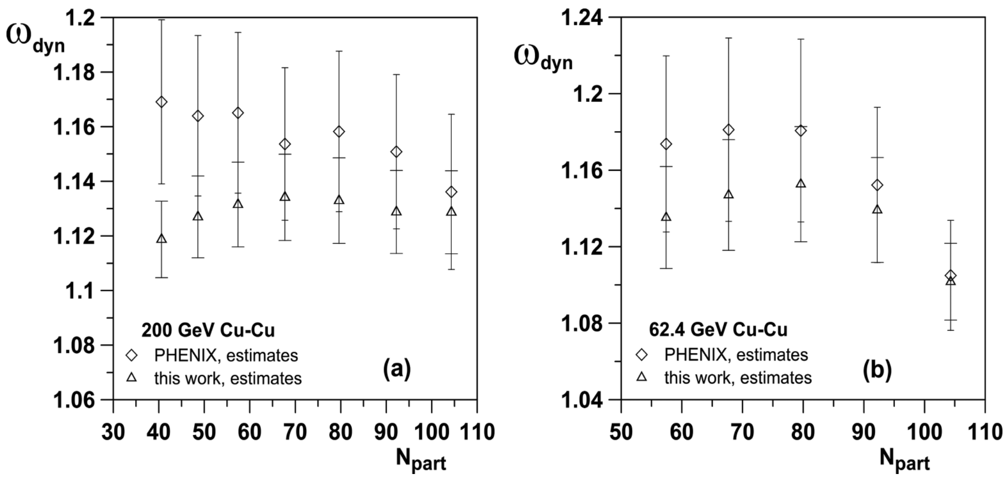

FIG. 8. Scaled variance for 200-GeV (a) and 62.4-GeV (b) Cu-Cu collisions. PHENIX estimates are from [1]. Estimates from this work are for the cases listed in Tables VI and VIII.

can conclude that only for the PHENIX collision system Au-Au at $\sqrt{s_{NN}} = 62.4$ GeV the hypothesis of the NBD could not be rejected. For other systems the hypothesis of the NBD seems to be very unlikely. What distinguishes the case of Au-Au collisions at $\sqrt{s_{NN}} = 62.4$ GeV from others? The only thing which can be noticed from Tables I–VIII is that the number of events is substantially greater (about 14%) in this case.

In principle, the accuracy with which experimental distributions approximate the NBD should increase with the sample size because if the hypothesis is true the postulated form of distribution is exact for the whole population. So with the growing number of events the experimental distribution should be closer to the postulated one. This is also seen in the form of $\chi^2_{\lambda,\min}$, Eq. (19), where the linear dependence on N is explicit. To keep $\chi^2_{\lambda,\min}$ at least constant when N (the sample size) is growing the relative differences between $P(Y_i)$ and P_i^{ex} have to decrease. The PHENIX test statistic χ^2_{PHEN}, Eq. (23), reveals the same feature because relative errors behave like $\sqrt{n_i}/N$. So the results of fits for the collision system Au-Au at $\sqrt{s_{NN}} = 62.4$ GeV are even more valuable.

Another surprising point is the comparison of the values of the PHENIX test statistic χ^2_{PHEN} divided by n_d, the ninth column of Tables I–VIII, with the corresponding values of Ref. [1]. For choice ii (Tables II, IV, VI, and VIII), the $\chi^2_{\mathrm{PHEN}}/n_d$ values obtained here are lower than the corresponding ones in Ref. [1]. Values of the parameters $\hat{k}, \hat{n}$ are also different from those in Ref. [1], which has resulted in slightly different (1–3% lower) values of the scaled variance ω_{dyn} (see Figs. 7 and 8). To make the comparison easier, also values of $\hat{k}^{-1}_{\mathrm{dyn}}$ are presented in the fifth column of Tables I–VIII. Generally, $\hat{n}$ is greater but the difference does not exceed 10% and decreases with the centrality. $\hat{k}^{-1}_{\mathrm{dyn}}$ is smaller, especially for case ii, and the difference also decreases with the centrality—from about 20–30% for the least central classes to about 5–10% for the most central ones.

IV. CONCLUSIONS

Results of the likelihood ratio test (likelihood χ^2) suggest that the hypothesis of the NBD of charged particle multiplicities measured by the PHENIX Collaboration in Au-Au and Cu-Cu collisions at $\sqrt{s_{NN}} = 62.4$ and 200 GeV should be rejected for all centrality classes. However, it must be stressed that the maximum likelihood method and the likelihood ratio test do not take actual experimental errors into account. This could be seen as a drawback but, in fact, only the LS test statistic takes actual experimental errors into account. Thus, the problem with the size of errors might occur when the LS method is used not only to fit parameters of a theoretical model but also to assess how confident the rejection or acceptance of a hypothesis is. This is because too big or too small errors cause a false inference in this case. But the judgment whether errors are too big already or still adequate is subjective. When errors are large enough it is likely that a false hypothesis would be accepted (this situation is called "error of the second kind" in statistics [4,5,7]). Also, one can encounter serious difficulties when one tries to express somehow the goodness of fit when the LS method is applied, as explained in Appendix B.

The goodness of fit expressed by the P value is necessary to assess the quality of fit. Here is an example: let $\chi^2/n_d = 1.5$ for a test which is χ^2 distributed. Is this fit good or bad? It depends on n_d. But how does one find any quantitative measure to decide? This measure is the P value. For $n_d = 10$, $P = 0.13$, so the fit should be accepted at the significance level 0.1%, but, for $n_d = 100$, $P = 0.0009$, so the fit should be rejected at the same significance level ([4], p. 62). However, to calculate the P value one has to know the distribution of the test statistic at the parameter estimates. In the general case of the LS test statistic this distribution is unknown, unless very specific assumptions are fulfilled, as shown in Appendix B. Certainly, assumptions 1 and 3 are not fulfilled when the NBD hypothesis is tested and systematic errors are added in quadrature to statistical ones. Thus, at the beginning of the investigations the situation is the following: the likelihood χ^2 does not take the errors into account, but its distribution is known asymptotically; the LS test statistic takes errors (including systematic ones) into account but its distribution is not known, even asymptotically. In the PHENIX case and with their estimations of systematic errors, these problems have been resolved naturally; i.e., both goals have been achieved—statistical and systematic errors are taken into account and the test statistic distribution is known.

The LS method, as the PHENIX Collaboration applied it, i.e., with their systematic errors included, has revealed a few very interesting things. First, it has turned out that the corresponding LS test statistic (the PHENIX test statistic χ^2_{PHEN}) equals Neyman's χ^2 test statistic multiplied by 0.1. This enables to use the well-known asymptotic properties of Neyman's χ^2 to find the asymptotic distribution of the PHENIX test statistic, so the goodness of fit can be now calculated because sample sizes are very large here. Additionally, PHENIX test statistic estimators of NBD parameters are Neyman's χ^2 estimators, but likelihood χ^2 and Neyman's χ^2 test statistics are asymptotically equivalent, so for a very large sample their estimators (and estimates) should coincide. Therefore, determination of NBD parameters with the use of the ML method and then insertion of them into the PHENIX test statistic is reasonable. Note that this method of the determination of NBD parameters has turned out to be much simpler than with the use of the LS method, e.g., the optimal $\bar{n}$ equals $\langle N_{\text{ch}}\rangle$ (see Appendix A). Finally, because the likelihood χ^2 converges faster to efficiency than Neyman's χ^2, this method should be preferable when estimation of parameters and errors on estimates are considered ([5], p. 193; [10], Sec. 18.59).

The correct inference from the results of the PHENIX test statistic χ^2_{PHEN}, i.e., the test statistic which in contrast to the likelihood χ^2 takes the systematic errors into account, shows that the hypothesis of the NBD of charged particle multiplicities measured in Au-Au and Cu-Cu collisions at $\sqrt{s_{NN}} = 62.4$ and 200 GeV should be accepted roughly in one fourth of PHENIX classes of the collision system and centrality. In particular, for the PHENIX collision system Au-Au at $\sqrt{s_{NN}} = 62.4$ GeV as a whole the hypothesis of the NBD could not be rejected, whereas for the Cu-Cu system at the same energy it should be rejected. For two other systems (both at $\sqrt{s_{NN}} = 200$ GeV) the hypothesis of the NBD seems to be very unlikely.

ACKNOWLEDGMENTS

The author thanks Jeffery Mitchell for helpful explanations of the PHENIX data. This work was supported in part by the Polish Ministry of Science and Higher Education under Contract No. N N202 0523 40.

APPENDIX A: ML ESTIMATES OF NBD PARAMETERS

Dropping terms not depending on the parameters in Eq. (18), one obtains the following form for the log-likelihood function under consideration:

$$\ln L(\mathbf{Y}|p,k) = N \sum_{i=1}^{m} P_i^{\text{ex}} \ln P(Y_i;p,k). \tag{A1}$$

Since the logarithm of the NBD is given by

$$\ln P(n;p,k) = \sum_{j=1}^{n} \ln(k+j-1) + n\ln(1-p) + k\ln p - \ln(n!), \tag{A2}$$

the necessary conditions for the existence of the maximum have the following form:

$$\begin{aligned}\frac{\partial}{\partial p}\ln L(\mathbf{Y}|p,k) &= N\sum_{i=1}^{m} P_i^{\text{ex}}\left[-Y_i\frac{1}{1-p}+\frac{k}{p}\right]\\ &= N\left[-\frac{1}{1-p}\sum_{i=1}^{m} P_i^{\text{ex}}Y_i + \frac{k}{p}\sum_{i=1}^{m} P_i^{\text{ex}}\right]\\ &= N\left[-\frac{1}{1-p}\langle N_{\text{ch}}\rangle + \frac{k}{p}\right] = 0,\end{aligned} \tag{A3}$$

$$\begin{aligned}\frac{\partial}{\partial k}\ln L(\mathbf{Y}|p,k) &= N\sum_{i=1}^{m} P_i^{\text{ex}}\left[\sum_{j=1}^{Y_i}\frac{1}{k+j-1}+\ln p\right]\\ &= N\left[\sum_{i=1}^{m} P_i^{\text{ex}}\sum_{j=1}^{Y_i}\frac{1}{k+j-1}+\ln p\right] = 0,\end{aligned} \tag{A4}$$

where the sum over j is 0 if $Y_i = 0$.

From Eqs. (A3) and (2) one can obtain

$$\langle N_{\text{ch}}\rangle = \frac{k(1-p)}{p} = \bar{n}. \tag{A5}$$

Expressing p as a function of k and $\langle N_{\text{ch}}\rangle$,

$$\frac{1}{p} = \frac{\langle N_{\text{ch}}\rangle}{k} + 1, \tag{A6}$$

and substituting it into Eq. (A4) the equation which determines $\hat{k}$ is obtained:

$$\frac{\partial}{\partial k}\ln L(\mathbf{Y}|p,k) = N\left[\sum_{i=1}^{m} P_i^{\text{ex}}\sum_{j=1}^{Y_i}\frac{1}{k+j-1} - \ln\left(1+\frac{\langle N_{\text{ch}}\rangle}{k}\right)\right] = 0. \tag{A7}$$

The above equation can be solved numerically. Having obtained $\hat{k}$ and substituting it into Eq. (A6), $\hat{p}$ is derived.

APPENDIX B: STATISTICAL INFERENCE IN A CAPSULE

Let $\{Y_1, Y_2, \dots, Y_N\}$ be a set of repeated observations of a random variable Y or a single observation of N-dimensional random variable $\vec{Y} = (Y_1, Y_2, \dots, Y_N)$ (this Appendix is a brief summary based on Refs. [4,5]). The null hypothesis, H_0, specifies a PDF of Y or a joint PDF of $\vec{Y}$. The test statistic t is a function of the observations (a function of N random variables equivalently): $t = t(Y_1, Y_2, \dots, Y_N)$. For simplicity let us assume that t is a scalar function. Let $g(t|H_0)$ be a given PDF for the statistic t if H_0 is true. The qualitative assessment about the compatibility of H_0 with the data is expressed as a decision to accept or reject the null hypothesis. This is done by choosing a value t_{cut}, called the cut or decision boundary. Then, for given observations $\{Y_1, Y_2, \dots, Y_N\}$ $t_O = t(Y_1, Y_2, \dots, Y_N)$ and if $t_O > t_{\text{cut}}$ the hypothesis is rejected; if

$t_O \leqslant t_{\rm cut}$, H_0 is accepted. Usually $t_{\rm cut}$ is chosen in such a way that one obtains the assumed probability α to reject H_0 if H_0 is true—this is called the significance level:

$$\alpha = \int_{t_{\rm cut}}^{\infty} g(t|H_0)dt. \tag{B1}$$

Now, let $\vec{Y}$ be an N-dimensional Gaussian random variable with known covariance matrix V but not known expectation values. $\vec{Y}$ is related to another variable $\vec{X}$ in such a way that there is a true value function ($\equiv$ a hypothesis) $\Lambda = \Lambda(X;\vec{\theta})$, which depends on unknown parameters $\vec{\theta} = (\theta_1,\ldots,\theta_m)$ and an expectation value of Y_i, $E[Y_i] = \Lambda(X_i;\vec{\theta})$. Then one defines the least-squares (LS) statistic as

$$\chi^2_{\rm LS}(\vec{Y};\vec{\theta}) = \sum_{i,j=1}^{N}(Y_i - \Lambda(X_i;\vec{\theta}))[V^{-1}]_{ij}(Y_j - \Lambda(X_j;\vec{\theta})). \tag{B2}$$

Instead, if one has N independent Gaussian random variables with different unknown means but known variances σ_i^2 and the true value function $\Lambda = \Lambda(X;\vec{\theta})$, then the LS statistic, Eq. (B2), becomes

$$\chi^2_{\rm LS}(\vec{Y};\vec{\theta}) = \sum_{i=1}^{N}\frac{(Y_i - \Lambda(X_i;\vec{\theta}))^2}{\sigma_i^2}. \tag{B3}$$

Let $\vec{Y}$ be a single measurement of the N-dimensional random variable (or a set of independent measurements of N random variables). Having replaced the variables by their measured values in Eq. (B2) [or Eq. (B3)] one converts the LS statistic $\chi^2_{\rm LS}(\vec{Y};\vec{\theta})$ into the function of $\vec{\theta}$ only. The next step is to minimize this function with respect to $\vec{\theta}$. Values of parameters at the minimum are called the LS estimators, $(\hat{\theta}_1,\ldots,\hat{\theta}_m)$. When one has replaced parameters $\vec{\theta}$ (treated as free until now) by their estimators in Eq. (B2) [or Eq. (B3)], then a test statistic $t_{\chi^2} = t_{\chi^2}(Y_1,Y_2,\ldots,Y_N) \equiv \chi^2_{\rm LS,min}(\vec{Y}) = \chi^2_{\rm LS}(\vec{Y};\hat{\theta}_1,\ldots,\hat{\theta}_m)$ is obtained. What is the decision boundary $t_{\chi^2,\rm cut}$ for this test statistic? The choice of the proper $t_{\chi^2,\rm cut}$ is the consequence of the following theorem ([4], pp. 95–96, 104; [16], Sec. 10.4.3). If (1) $(Y_1,Y_2,\ldots,Y_N)$ is an N-dimensional Gaussian random variable with known covariance matrix V or $(Y_1,Y_2,\ldots,Y_N)$ are independent Gaussian random variables with known variances σ_i^2; (2) variables $(X_1,X_2,\ldots,X_N)$ are measured with infinite precision, i.e., without any errors; (3) the hypothesis $\Lambda(X;\theta_1,\ldots,\theta_m)$ is linear in the parameters θ_i; and (4) the hypothesis is correct, then the test statistic $\chi^2_{\rm LS,min}$ is distributed according to a χ^2 distribution with $n_d = N - m$ degrees of freedom. If the hypothesis $\Lambda(X;\theta_1,\ldots,\theta_m)$ is nonlinear in the parameters, the exact distribution of $\chi^2_{\rm LS,min}$ is not known. However, asymptotically (when $N \longrightarrow \infty$) the distribution of $\chi^2_{\rm LS,min}$ approaches a χ^2 distribution as well ([16], p. 287; [17], p. 147). Thus, when assumptions 1, 2, and 4 at least are fulfilled and the sample size is large one can consider the $\chi^2_{\rm LS,min}$ test statistic as χ^2 distributed. The expectation value of a random variable Z distributed according to the χ^2 distribution with n_d degrees of freedom is $E[Z] = n_d$ and the variance $V[Z] = 2n_d$. As a result "one expects in a 'reasonable' experiment to obtain $\chi^2_{\rm LS,min} \approx n_d$" ([18], p. 15). Therefore, for the test statistic $t_{\chi^2} = \chi^2_{\rm LS,min}$ the decision boundary $t_{\chi^2,\rm cut} = E[\chi^2_{\rm LS,min}] = n_d$ is chosen. Usually the so-called "reduced χ^2" is reported, which equals $\chi^2_{\rm LS,min}/n_d$. Thus, for $\chi^2_{\rm LS,min}/n_d$ the decision boundary is just one. It must be stressed here that this choice is the consequence of the fact that the $\chi^2_{\rm LS,min}$ test statistic is χ^2 distributed. If the distribution of $\chi^2_{\rm LS,min}$ is not known at all (e.g., one of the assumptions 1, 2, or 4 is not fulfilled or the sample size is small), this choice is arbitrary—based on common belief rather than on any justification.

The comparison of the actually obtained value of the test statistic $t_O = t(Y_1,Y_2,\ldots,Y_N)$ with the decision boundary $t_{\rm cut}$ gives only qualitative information about the validity of the hypothesis H_0. If one wants to express quantitatively how the null hypothesis agrees with the data a test of goodness of fit is necessary [4,5]. The value of this test shows the level of the compatibility of the observed data with the predictions of H_0. This value is given by the probability P, under the assumption that H_0 is true and the experiment would be repeated many times under the same circumstances, of obtaining results as compatible or less with H_0 than the result just observed. This probability is called the P value of the test and can be expressed as ([5], p. 300)

$$P = \int_{\vec{Y}:t\geqslant t_O} f(\vec{Y}|H_0), \tag{B4}$$

where $f(\vec{Y}|H_0)$ is the PDF of the N-dimensional random variable $\vec{Y}$ under the null hypothesis H_0. In general the above integral could be very difficult to calculate unless the PDF $g(t|H_0)$ of the test statistic t is known somehow; then one obtains ([18], p. 13)

$$P = \int_{t_O}^{\infty} g(t|H_0)dt. \tag{B5}$$

Note that this is not the same as Eq. (B1) because that expression is the equation for $t_{\rm cut}$ given the significance level α and should be solved before the measurement, whereas Eq. (B5) is calculated after the measurement and reflects the obtained (dis)agreement of the observation with the hypothesis H_0. The criterion for the rejection or acceptance of H_0 can be now formulated with the use of P and α instead of t_O and $t_{\rm cut}$: if $P \leqslant \alpha$ then the hypothesis should be rejected; otherwise it should be accepted.

However, the most interesting class of test statistics is that in which their distributions are known independently of H_0. The most important class consists of so-called "χ^2 statistics," i.e., test statistics which are distributed (at least asymptotically) in the χ^2 distribution [5,6]. Note that $\chi^2_{\rm LS}$, defined earlier, when the assumptions of the theorem are fulfilled, belongs to this class. The likelihood χ^2, Eq. (14), Pearson's χ^2, and Neyman's χ^2 mentioned in Sec. III do as well. Then the P value is given by

$$P = \int_{t_O}^{\infty} f(z;n_d)dz, \tag{B6}$$

where $f(z; n_d)$ is the χ^2 PDF and n_d is the number of degrees of freedom.

APPENDIX C: WILKS'S THEOREM

Let X be a random variable with PDF $f(X, \theta)$, which depends on parameters $\theta = \{\theta_1, \theta_2, \ldots, \theta_d\} \in \Theta$, where a parameter space Θ is an open set in R^d. For the set of N independent observations of X, $\mathbf{X} = \{X_1, X_2, \ldots, X_N\}$, one can define the likelihood function:

$$L(\mathbf{X}|\theta) = \prod_{j=1}^{N} f(X_j; \theta). \tag{C1}$$

Now consider H_0, a k-dimensional subset of Θ, $k < d$. Then the maximum likelihood ratio can be defined as

$$\lambda = \frac{\max_{\theta \in H_0} L(\mathbf{X}|\theta)}{\max_{\theta \in \Theta} L(\mathbf{X}|\theta)}. \tag{C2}$$

This is a statistic because it does not depend on parameters θ anymore; in the numerator and the denominator there are likelihood function values at the ML estimators of parameters θ with respect to sets H_0 and Θ, respectively.

Wilks's theorem says that under certain regularity conditions if the hypothesis H_0 is true (i.e., it is true that $\theta \in H_0$) then the distribution of the statistic $-2 \ln \lambda$ converges to a χ^2 distribution with $d - k$ degrees of freedom as $N \longrightarrow \infty$ [5,7]. The proof can be found in Ref. [19]. Note that $k = 0$ is possible, so one point in the parameter space (one value of the parameter) can be tested as well.

[1] A. Adare *et al.* (PHENIX Collaboration), Phys. Rev. C **78**, 044902 (2008).
[2] G. J. Alner *et al.* (UA5 Collaboration), Phys. Lett. B **160**, 193 (1985).
[3] R. E. Ansorge *et al.* (UA5 Collaboration), Z. Phys. C **43**, 357 (1989).
[4] G. Cowan, *Statistical Data Analysis* (Oxford University Press, Oxford, 1998).
[5] F. James, *Statistical Methods in Experimental Physics* (World Scientific, Singapore, 2006).
[6] S. Baker and R. D. Cousins, Nucl. Instrum. Meth. **221**, 437 (1984).
[7] P. G. Hoel, *Introduction to Mathematical Statistics*, 4th ed. (Wiley, New York, 1971).
[8] T. Abbott *et al.* (E-802 Collaboration), Phys. Rev. C **52**, 2663 (1995).
[9] L. Lyons, *Statistics for Nuclear and Particle Physicists* (Cambridge University Press, Cambridge, 1986).
[10] A. Stuart, J. K. Ord, and S. Arnold, *Kendall's Advanced Theory of Statistics, Vol. 2A: Classical Inference and the Linear Model*, 6th ed. (Wiley, Chichester, 2008).
[11] http://www.phenix.bnl.gov/phenix/WWW/info/data/ppg070_data.html; also, see note for Figs. 1 and 2, below the data for Fig. 2(c).
[12] J. T. Mitchell (private communication).
[13] J. Berkson, Ann. Stat. **8**, 457 (1980).
[14] J. Berkson, Biometrics **28**, 443 (1972).
[15] F. Beaujean, A. Caldwell, D. Kollar, and K. Kroninger, Report No. CERN-2011-006, in *Proceedings of the PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding*, edited by H. B. Prosper and L. Lyons (CERN, Geneva, 2011), pp. 177–182.
[16] A. G. Frodesen, O. Skjeggestad, and H. Tofte, *Probability and Statistics in Particle Physics* (Universitetsforlaget, Bergen, 1979).
[17] B. P. Roe, *Probability and Statistics in Experimental Physics*, 2nd ed. (Springer-Verlag, New York, 1992).
[18] J. Beringer *et al.* (Particle Data Group Collaboration), Phys. Rev. D **86**, 010001 (2012), http://pdg.lbl.gov/2012/reviews/rpp2012-rev-statistics.pdf.
[19] R. M. Dudley, *18.466 Mathematical Statistics, Spring 2003*, Massachusetts Institute of Technology OpenCourseWare, http://ocw.mit.edu/courses/mathematics/18-466-mathematical-statistics-spring-2003/lecture-notes/.