Remarks on statistical model fits

Francesco Becattini[*][†]

University of Florence and INFN Sezione di Firenze
E-mail: becattini@fi.infn.it

In this paper we point out a seemingly overlooked problem in some well known fits of hadron abundances to the statistical hadronization model published in literature. This is concerned with the replacement of measured hadron yields with ratios of yields in the fit data sample. We show that this is a statistically incorrect method implying a significant bias in the assessments of freeze-out parameters.

---

[*]Speaker.
[†]

1. Introduction

The statistical model is succesfull in describing the mean hadron multiplicities in both elementary and heavy ion collisions at high energy. In heavy ion collisions, many groups have carried out analysis of the available measurements for almost 10 years by now, reporting success in reproducing the data with few parameters. The usual employed technique is a χ^2 fit, where the data is compared to the model prediction to determine the best values of those parameters (temperature, baryon-chemical potentials, volume and, possibly, non-equilibrium parameters γ_S and γ_q). This analysis technique was first used for heavy ion collision data in ref. [1] (I beg reader's pardon for self-quotation). The calculations performed by different groups agree to a very good degree of accuracy *with the same data input*, showing that the implementations of the hadron gas model are consistent. The main difference between the calculations and conclusions of those groups is concerned with the used data input and the handling of statistical methods. The first difference is related to *a priori* physical assumptions and can be settled after a careful study of the data. One of the main controversies was related to the fitting of midrapidity or full phase space integrated yields, for which we refer to the discussion in the paper [2]. On the other hand, there should be no difference in statistical methods, which are an established and universally accepted tool. Yet, there is a misuse of fitting techniques which has led to biased results in the literature. Particularly, I will show that forming particle ratios out of *experimentally measured* yields and plugging them in statistical model fits involves a significant bias and it is an essentially incorrect procedure.

2. Why not to form particle ratios

When trying to determine the best fit parameters of the statistical model, one is given particle multiplicities or ratios of multiplicities, either at midrapity (i.e. dN/dy for $y = 0$ in the centre-of-mass frame) or in full phase space. The experiments sometimes quote ratios because a cancellation of some systematic error is implied, so that the effective error on the ratio is consistently smaller than that one would get if the yields had uncorrelated errors. In most cases, though, instead of ratios, experiments have been quoting yields, which were obtained by means of extrapolating fits to both p_T and rapidity spectra. Unless some definite information is provided by the experimenters, one is supposed to assume the errors on the yields to be fully uncorrelated, so that the no further information can be obtained just by manipulating the yields.

Nevertheless, some authors have been forming ratios out of quoted experimental multiplicities and plugged them in their statistical fit machinery to determine temperature and baryon-chemical potential, even without taking into account relevant correlations between different ratios. In some cases, this has been done for a full set of measurements, i.e. at centre-of-mass energies (typically at AGS and SPS) where experiments only quoted multiplicities and no ratio, like in ref. [3]. This procedure is meant to get rid of the volume parameter in the fit, which is an overall normalization factor cancelling out in ratios, leaving only the intensive parameters T, μ_B and γ_S. However, this method would be incorrect even if correlations were taken into account and leads to a bias in the determination of these parameters themselves. We will illustrate how this problem comes about by first giving three simple examples and presenting a more realistic one in next section.

2.1 Example 1

The simplest example is the weighted average. Consider four independent measurements of the same quantity x with different normal errors, say $x_1 = 1.2 \pm 0.2$, $x_2 = 0.8 \pm 0.2$, $x_3 = 0.8 \pm 0.2$, $x_4 = 0.8 \pm 0.2$. It is well known that the problem of determining the best estimate of x through maximum likelihood method leads to the minimization of the χ^2:

$$\chi^2 = \sum_{i=1}^{4} \frac{x_i - x_0}{\sigma_i^2} \tag{2.1}$$

and has the weighted average as solution, which is in this case 0.9 ± 0.1 with a $\chi^2/dof = 1$, meaning a very good fit.

If, on the other hand, we want to assess the consistency of the four measurements by taking ratios of pairs, we soon face an ambiguity: how many ratios should one take? The naive answer is taking as many as degrees of freedom in the χ^2 minimization, that is 3 in our example. Yet, there are 6 different triplets ($N(N-1)/2$ in general) which can be formed out of 4 objects, considering as equivalent a ratio x_i/x_j and its inverse x_j/x_i. Therefore, a choice has to be made; for instance, if we took x_1/x_2, x_1/x_3 and x_1/x_4, we would get three times 1.5, whereas if we took x_3/x_2, x_3/x_4 and x_2/x_1 we would get 1.0 twice and 0.66. The two triplets of ratios (1.5,1.5,1.5) and (1.0,1.0,0.66) submitted to a consistency test yield different answers in terms of statistical significance, even taking into account the correlations between them. The deep reason of this is an information loss in using ratios of measurements instead of measurements themselves; by retaining only three ratios out of six to avoid redundancy, one is forced to give up some information and the statistical significance happens to depend on the particular chosen subset of ratios.

2.2 Example 2

In this example we provide a concrete numerical example showing the awkwardness involved in using ratios instead of measurement. Consider a simple linear model $y = x + c$ to be fitted to the measurements $(x,y) = (1, 2.3 \pm 0.1), (2, 2.8 \pm 0.13), (3, 4.3 \pm 0.08)$. The correct fit procedure to determine the best value of c is the minization of:

$$\chi^2 = \sum_{i=1}^{3} \frac{(y_i - x_i - c)^2}{\sigma_i^2} \tag{2.2}$$

If we adopted the ratio method, we should choose two ratios formed out of the three measurements, e.g. y_1/y_3 and y_2/y_3 and minimize the χ^2:

$$\chi^2 = \sum_{i,j=1}^{2} (R_i - \frac{x_{k(i)} + c}{x_{h(i)} + c}) C_{ij}^{-1} (R_j - \frac{x_{k(j)} + c}{x_{h(j)} + c}) \tag{2.3}$$

where $k(i), h(i)$ are the indices of the measurements used to form the ith ratio and C is the co-variance matrix with non-vanishing off-diagonal elements estimated by means of the usual error propagation rules. For this example, the χ^2 profiles as a function of c are compared in fig. 1. One can clearly see that both the minimum and the curvature of the χ^2 around the minimum are different for the two functions (2.2) and (2.3). This is reflected in different estimates of the best fit value and its error. It is especially worth remarking that the error estimate related to (2.3) is much larger than the correct error.
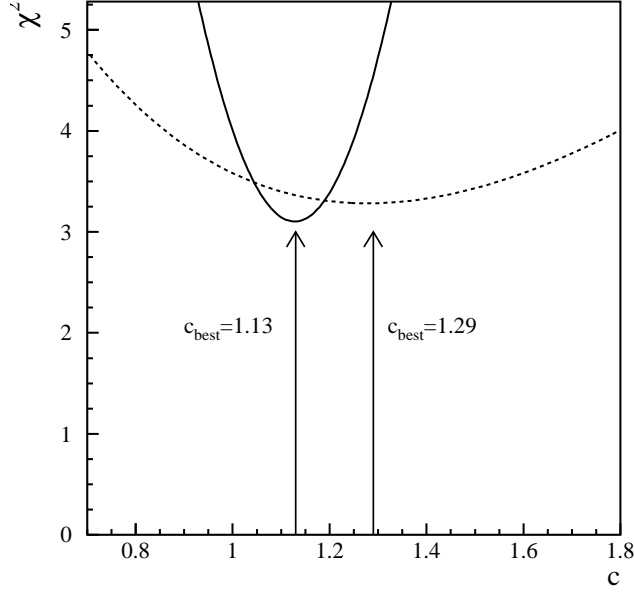
Figure 1: χ^2 profile for the example 2.

2.3 Example 3

We are now going to consider an example which is very close to the actual problem of fitting multiplicities to the statistical model, which will make it clear, beyond any doubt, that replacing N multiplicities with $N-1$ ratios is incorrect also for an exponential fit like the statistical-thermal model. Consider a model $y = a\exp(bx)$ and three measurements $(x,y) = (0, 1.8 \pm 0.1), (1, 2.71 \pm 0.13), (2, 6.5 \pm 0.08)$. Here, the parameter a corresponds to a volume and b to an inverse temperature. The idea is to get rid of parameter a and fitting just b by taking ratios of measurements. The correct *chi*² now reads:

$$\chi^2 = \sum_{i=1}^{3} \frac{(y_i - a\exp(bx_i)^2}{\sigma_i^2} \tag{2.4}$$

whereas the *chi*² for the ratios $R_1 = y_1/y_3$, $R_2 = y_2/y_3$ reads:

$$\chi^2 = \sum_{i,j=1}^{2} (R_i - \exp[bx_{k(i)} - x_{h(i)}])C_{ij}^{-1}(R_j - \exp[bx_{k(j)} - x_{h(j)}]) \tag{2.5}$$

The parameter a has disappeared from Eq. (2.5). If we have a look at χ^2 profiles of (2.4) and (2.5) as a function of b, we can see that also in this case both minima and curvature around the minima differ. The introduced bias, i.e. the difference between the correct b and that estimated with (2.5) is 2% which is little but cannot be neglected if one aimes at reaching great accuracy.

3. An example at RHIC

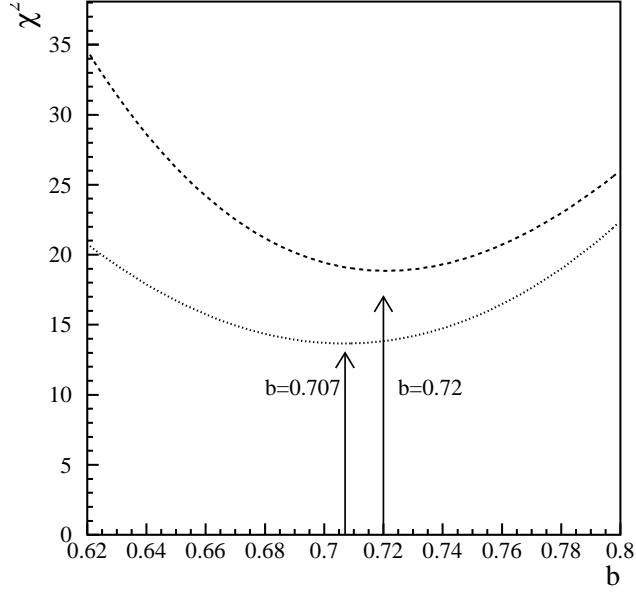We will now provide a realistic example of how the use of ratios instead of yields alters the

Figure 2: χ^2 profile for the example 3.

actual estimation of thermodynamical parameters. We will use data collected by STAR experiment at RHIC, at $\sqrt{s}_{NN} = 130$ GeV, shown in table 1. We stress that experimental numbers here serve as an instrument to compare different fit procedures; for a more exhaustive discussion on how they have been collected and other comments, see Jakko Manninen's talk in this conference [4]. The aim of this exercise is to show how dangerous the choice of a particular set of ratios can be in terms of the fit outcome.

In this particular case, we have to pick 11 ratios out of 66 and we then have much freedom. Let us first perform a correct fit to the midrapidity yields, as quoted in table 1, and construct an array of residuals, i.e. a set of differences between fitted values and actual measurements, scaled by the experimental error. Indeed, it is fairly easy to realize that if we systematically choose ratios of light particles with positive residual to heavy particles with negative residual, the ensuing value of T from a new fit to the ratios will tend to be larger and, as a consequence, the best fit value will be biased.

To show this, we have first performed a fit to midrapidity densities in table 1 by fixing the strangeness suppression factor γ_S to 1. We then had a look at residuals of all particles and took the heaviest particle with negative residual, i.e. $\bar{\Xi}^-$. We then chose a set of 11 ratios $\langle X \rangle / \langle \bar{\Xi}^+ \rangle$, X being any particle lighter than Ξ^- in table 1, and added as last ratio in the data sample $\langle p \rangle / \langle \Omega + \bar{\Omega}$, also showing a fluctuation in the same direction. As has been mentioned, the expectation for this kind of analysis is to artificially *enhance* the temperature in this fit, because all lighter particles had a residual larger than $\bar{\Xi}$'s. This is indeed what is found, as shown in table 2. If we do not include correlations, as it is done e.g. in ref. [3], the difference between fit to ratios and correct fit to the

Particle	$dN/dy|_{y=0}$
π^+	239±10.6
π^-	239±10.6
K^+	45.8±6.7
K^-	43.2±6.0
p	26.2±6.0
$\bar{p}$	18.9±4.3
Λ	17.2±1.8
$\bar{\Lambda}$	12.3±1.3
ϕ	6.09±0.85
Ξ^-	2.13±0.27
$\bar{\Xi}^+$	1.78±0.24
$\Omega + \bar{\Omega}$	0.586±0.128

Table 1: List of midrapidity yields of different hadrons measured by STAR in Au-Au collisions at $\sqrt{s}_{NN} = 130$ GeV (see ref. [4]).

Parameters	Fit to yields	Fit to ratios w/o correlations	Fit to ratios with correlations
T [MeV]	168.4±3.2	175.6±3.0	170.3±3.0
μ_B [MeV]	36.2±0.6	23.4±0.4	33.1±0.6
χ^2/dof	6.3/10	6.6/10	7.6/10

Table 2: Results of statistical model fits to hadron midrapidity densities in table 1. The first column shows the best fit parameter in a direct fit to the yields. The second column shows the results of a fit to the set of ratios described in the text. The thirs column shows the results of a fit to the same set of ratios taking into account correlations in the χ^2.

yields is 8.2 MeV, i.e. about 2.6 σ; the situation is dramatically worse for the baryon-chemical potential, with a difference of 21 σ! On the other hand, by including correlations, the situation gets improved and one is left with a slight discrepancy from the main fit. Yet, the fit method is still conceptually wrong and should be avoided.

4. Conclusions

We have shown that the use of hand-made particle ratios instead of measured yields in fits is an incorrect method to determine the parameters of a model like, e.g. the statistical hadronization model in heavy ion collisions. This has been done in several cases and it is especially relevant for self-claimed "2nd generation analyses" [3] aiming at reaching the best accuracy but using this incorrect method.

We stress that this does not mean that experiments should not quote ratios, if a smaller systematic error is involved: this is a perfectly correct and welcome procedure. What is wrong is taking experimental yields and forming ratios out of them *a posteriori*. We have shown in a realistic ex-

ample that the introduced bias may be as large as $\mathcal{O}(10)$ MeV for temperature and baryon-chemical potential if correlations are not taken into account.

References

[1] F. Becattini, M. Gazdzicki and J. Sollfrank, Eur. Phys. J. C **5** (1998) 143.

[2] F. Becattini, J. Manninen and M. Gazdzicki, Phys. Rev. C **73** (2006) 044905.

[3] A. Andronic, P. Braun-Munzinger and J. Stachel, Nucl. Phys. A **772** (2006) 167.

[4] J. Manninen, this conference.

PoS(CPOD2006)033