Self-similarity of the negative binomial multiplicity distributions

G. Calucci and D. Treleani
Dipartimento di Fisica Teorica dell'Università and INFN, Trieste, I 34014 Italy

(Received 7 August 1997; published 2 December 1997)

The negative binomial distribution is self-similar: If the spectrum over the whole rapidity range gives rise to a negative binomial, in the absence of correlation and if the source is unique, also a partial range in rapidity gives rise to the same distribution. The property is not seen in experimental data, which are rather consistent with the presence of a number of independent sources. When multiplicities are very large, self-similarity might be used to isolate individual sources in a complex production process. [S0556-2821(98)00401-9]

PACS number(s): 12.40.Ee, 13.85.Hd

One of the first basic pieces of evidence observed in the field of many-particle production and nuclear collisions is the distribution of the multiplicity of the produced particles. Multiplicity distributions are measured both by looking at the whole spectrum of the produced particles and by looking only at a restricted segment, typically a rapidity interval. Both for theoretical and experimental reasons, one of the favorite parametrizations of the multiplicity distribution [1], also in different rapidity intervals [2], is the negative binomial (NB) distributions. A very detailed discussion of the experimental evidence, the interpretations, and also the formalism used to deal with these kinds of problems has been recently published [3]. In the case of a generic distribution the relation between the multiplicities of a restricted part of the spectrum and those arising from the whole spectrum is not trivial. In the present paper we point out that for NB, on the contrary, a peculiar self-similarity property holds between the distributions obtained from different intervals of the spectrum.

We find it convenient to make use of the generating functional formalism to deal with these kinds of problems [4–6]. Let $W_n(\xi_1,\dots,\xi_n)$ be the normalized multiparticle exclusive distributions:

$$\sum_n \int W_n(\xi_1,\dots,\xi_n)d\xi_1,\dots,d\xi_n=1. \quad (1)$$

The variables ξ can have different meanings and also represent more than one physical parameter. In high-energy collisions ξ could represent the rapidity y and the transverse momentum; if the distributions refer to incoming partons ξ could represent the fractional longitudinal momentum x and the impact parameter. The distributions may be obtained in the usual way from a generating functional $\mathcal{Z}$:

$$W_n(\xi_1,\dots,\xi_n)=\frac{1}{n!}\frac{\delta}{\delta J(\xi_1)}\cdots\frac{\delta}{\delta J(\xi_n)}\mathcal{Z}[J]\big|_{J=0} \quad (2)$$

and the normalization is expressed by $\mathcal{Z}[1]=1$. Sometimes it is useful to also use an unrenormalized generator $\mathcal{G}$ with $\mathcal{Z}[J]=\mathcal{G}[J]/\mathcal{G}[1]$.

The probability of producing n particles, in any configuration, is evidently given by

$$\begin{aligned}p_n&=\int W_n(\xi_1,\dots,\xi_n)d\xi_1,\dots,d\xi_n\\&=\frac{1}{n!}\left[\int\frac{\delta}{\delta J(\xi)}d\xi\right]^n\mathcal{Z}[J]\big|_{J=0}\\&=\frac{1}{n!}\left[\frac{\partial}{\partial\lambda}\right]^n\mathcal{Z}[J+\lambda 1]\big|_{J=0,\lambda=0}\\&=\frac{1}{n!}\left[\frac{\partial}{\partial\lambda}\right]^n\mathcal{Z}[\lambda 1]\big|_{\lambda=0}=\frac{1}{n!}\left[\frac{\partial}{\partial\lambda}\right]^n z(\lambda)\big|_{\lambda=0}.\end{aligned} \quad (3)$$

Let us now consider the situation where the interval in which the variables ξ lie is divided into two parts. Then for a particular choice of these variables $W_n(\xi_1,\dots,\xi_n)=W_r(\xi'_1,\dots,\xi'_r)W_s(\xi''_1,\dots,\xi''_s)$ with $r+s=n$. Taking into account all the possible choices of ξ' and ξ'' it results that

$$\begin{aligned}&W_r(\xi'_1,\dots,\xi'_r)W_s(\xi''_1,\dots,\xi''_s)\\&=\frac{1}{r!}\frac{\delta}{\delta J(\xi'_1)}\cdots\frac{\delta}{\delta J(\xi'_r)}\frac{1}{s!}\frac{\delta}{\delta J(\xi''_1)}\cdots\frac{\delta}{\delta J(\xi''_s)}\mathcal{Z}[J]\big|_{J=0}.\end{aligned}$$

If we sum over all configurations in ξ'' the distributions in ξ' are $W_r(\xi'_1,\dots,\xi'_r)\cdot\Sigma\int W_s(\xi''_1,\dots,\xi''_s)d\xi''_1,\dots,d\xi''_s$. A set of semi-inclusive distributions are obtained in this way[1] since everything referring to the variables ξ'' are not observed. The generator of these new distributions is $\mathcal{Z}'=\mathcal{Z}[J'+\Theta'']$, where J' has as an argument only ξ', i.e., $J'(\xi'')=0$, Θ'' is 1 for $\xi=\xi''$ and 0 for $\xi=\xi'$, Θ' is 1 for $\xi=\xi'$ and 0 for $\xi=\xi''$. The probability of finding n particles in the observed part of the spectrum is then

$$p'_n=\frac{1}{n!}\left[\frac{\partial}{\partial\lambda}\right]^n\mathcal{Z}[\lambda\Theta'+\Theta'']\big|_{\lambda=0}=\frac{1}{n!}\left[\frac{\partial}{\partial\lambda}\right]^n z'(\lambda)\big|_{\lambda=0}. \quad (4)$$

Two particular cases of interest are as follows. The Poissonian distribution, which is obtained by defining

$$\mathcal{U}=\int J(\xi)\cdot D(\xi)d\xi,\quad \tilde{u}=\int D(\xi)d\xi,$$
$$\mathcal{G}=e^{\mathcal{U}[J]},\quad \mathcal{G}_1=e^{\tilde{u}}$$

and finally $\mathcal{Z}=e^{\mathcal{U}[J]-\tilde{u}}$.

[1] A similar treatment has been proposed by van Hove [7], in dealing with continuous NB distributions.

If one looks only at the spectrum in ξ' by integrating over ξ'', the new generator is $\mathcal{G}'=e^{\mathcal{U}[J(\xi')]+\mathcal{U}[\Theta'']}$; since $\widetilde{u}=\mathcal{U}[\Theta']+\mathcal{U}[\Theta'']$, $\mathcal{Z}'=\mathcal{Z}$.

The NB distribution, whose generating functional is

$$f(\mathcal{U})=[1-\mathcal{U}]^{-k}/[1-\widetilde{u}]^{-k}, \tag{5}$$

while the generator of the semi-inclusive spectra in ξ' is

$$\{1-\mathcal{U}[J(\xi')]-\mathcal{U}[\Theta'']\}^{-k}/[1-\widetilde{u}]^{-k}.$$

This corresponds to a pure redefinition of $\mathcal{U}$ since one gets the new generator by going from $\mathcal{Z}=f\{\mathcal{U}\}$ to $\mathcal{Z}'=f\{\mathcal{U}/(1-\mathcal{U}[\Theta''])\}$. This means that the NB is transformed into a NB, with the same exponent as the original one. Clearly, in both cases, the mean multiplicity is changed.

The generating function of the multiplicity distribution in this case is explicitly given as

$$z(\lambda)=[1-\lambda u'-u'']^{-k}/[1-\widetilde{u}]^{-k} \tag{6a}$$

or, after defining $r=u'/(1-\widetilde{u})$, in a different and sometimes more convenient form

$$z(\lambda)=[1+(1-\lambda)r]^{-k}. \tag{6b}$$

In terms of these parameters one gets for the mean multiplicity $\overline{n}=kr$ and for the dispersion $D^2=kr(r+1)$.

A survey of other kinds of one-body distributions shows that this property of self-similarity if only a part of the spectrum is observed is quite unlikely.[2] One may therefore wonder whether this property is peculiar to the NB distribution, with the Poissonian distribution as a limiting case, or if it is also found in other cases.

It will be shown that in the simplest conditions the property of self-similarity is unique of the NB distribution. In this case one can give for the non-normalized generating functional the representation $\mathcal{G}=g(\mathcal{U})$; the probabilities p', Eq. (4), can be obtained from a generating function $g(\lambda u'+u'')$, where

$$u'=\int D(\xi')d\xi', \quad u''=\int D(\xi'')d\xi'', \quad u'+u''=\widetilde{u}. \tag{7}$$

The invariance of the functional form of the distribution, when considering only limited parts of the spectrum is expressed as $g(x+y)=N(y)g(x\cdot f(y))$ because in this way the relation $p'_n=c^n p_n/C$ is produced, and this property can be expressed by saying that the distribution remains the same. The arbitrary normalization $g(0)=1$, which is always possible, gives $N(y)=g(y)$. So finally,

$$g(x+y)=g(y)g(x\cdot f(y)). \tag{8a}$$

By taking the first and the second derivative with respect to x and setting then $x=0$, two differential equations for $g(y)$ are obtained:

$$\dot{g}(y)=\dot{g}(0)g(y)f(y), \quad \ddot{g}(y)=\ddot{g}(0)g(y)f(y)^2. \tag{8b}$$

[2]E.g., the NB is a particular case of a hypergeometric distribution, but a generic hypergeometric distribution does not have this kind of self-similarity.

It then follows that $g(y)\ddot{g}(y)=R\dot{g}(y)^2$ with $R=\ddot{g}(0)/\dot{g}(0)^2$. With the usual position

$$g(y)=\exp\left[\int_0^y q(w)dw\right],$$

which ensures the correct normalization $g(0)=1$, the equation becomes

$$\dot{q}(y)=(R-1)q(y)^2. \tag{8c}$$

The solution of Eq. (8c) is $q(y)=[(1-R)y+S]^{-1}$. Redefining the constants as $k=1/(R-1)$ and $u=(R-1)/S$ one obtains

$$g_u(y)=[1-uy]^{-k}. \tag{9}$$

This expression is the generating function of a binomial distribution whose exponent is, in general, not integer. The meaning of the function $g(\xi)$ requires that it be positive together with all its derivatives in the origin; this certainly happens if the exponent is negative, i.e., $R>1$ and the parameter u is positive. A different possibility is given by positive integer exponent and negative u. This corresponds, however, to a distribution with only a finite number of terms.

The two differential equations (8b) are not completely equivalent to the functional relation Eq. (8a), but they follow from it. The conclusion is that the self-similarity implies the NB (which could be not sufficient) but it has already shown that the NB implies the self-similarity, so the two properties are equivalent. The generating functional of the NB is more conveniently expressed by writing $g_u(\lambda)$ as $g_1(\lambda u)$ and suppressing from now on the index 1; the normalized distribution is given by $z(\lambda)=g(\lambda u)/g(u)$.

The limit $R\to 1$ gives rise to the solution $g(y)=\exp[y/S]$, i.e., it yields the generating function for a Poissonian distribution.

The experimental evidence and their elaboration [8,9] show that the NB distribution holds well for different intervals of observed rapidity but that the parameters present strong variations. The real world does not show the sharp self-similarity property discussed above. The actual analysis was done in a frame where $\mathcal{Z}=f\{\mathcal{U}\}$ so that case genuine two-body correlation was absent.

When correlations are present the relation between exclusive and semi-inclusive distribution is more complicated and there is no obvious reason for the self-similarity to hold. However, this does not seem too promising: either the effect of the correlations is so strong that the NB distribution is destroyed or the overall effect is not very important; but then the parameters of the NB distribution are changed too little to agree with the experimental evidence. An example will be shown in the Appendix.

A more interesting possibility is given by the often considered possibility [1,2,9] of considering multiple sources in the rapidity range. Let us consider a simple case where a source extends in rapidity from y_0 to y_1 and another source is present from y_1 to y_2: when we observe the produced particles in a rapidity range that ends at $y_f<y_1$ then the second source in inactive, the parameter r grows with y_f and does the multiplicity, the parameter k stays evidently constant. When y_f goes beyond y_1 the first source is frozen (r

has attained its final value) and the second gives a contribution still growing with y_f. The generating function is now

$$z(\lambda)=[1+(1-\lambda)r]^{-k}\cdot[1+(1-\lambda)r_f]^{-k} \tag{10}$$

and does not yield a NB distribution. One could force the function $z(\lambda)$ to become a NB-generating function:

$$z_e(\lambda)=[1+(1-\lambda)r_e]^{-k_e} \tag{11}$$

by defining the equivalent parameters in such a way that multiplicity and dispersion acquire the correct values. The prescription is expressed through the auxiliary parameter $\rho = r_f/r$: $r_e=r(1+\rho^2)/(1+\rho)$, $k_e=k(1+\rho)^2/(1+\rho^2)$. In order to explore how good this representation is it is useful to calculate the higher central momenta $\mu_s=\langle(n-\langle n\rangle)^s\rangle$. The third central momentum indicates that the worst situation is produced for $r_f\approx\frac{1}{3}r$ and a similar indication is obtained by examining the fourth cumulant [10] $\kappa_4=\mu_4-3D^2$; in this situation the error cannot exceed 12%. One can also examine in detail the individual distribution of the multiplicity produced, respectively, by the generating functions Eq. (10) and Eq. (11); it results that the approximation is better than it could seem at first sight because large deviations between the two series of numbers is found for multiplicities very large, typically a discrepancy of the order 12% arises for multiplicities of the order of 25 which gives sizable contributions to the higher momenta but are not very relevant in the analysis of the data; for values from 6 to 9, where the maximum of the production rate lies in the difference, it is less than 1%. These values are obtained for $r_f\approx\frac{1}{3}r$, in other cases the discrepancy is definitely smaller. Anyhow, without dwelling furthermore on a particular form of approximation the conclusion that we are trying to draw is that a number of sources, each of them giving rise to a strict NB distribution within a definite range of rapidity, yields a distribution not very different when taken over the whole rapidity range.

If one would try to construct a model for high-energy particle production which implies sources extended in rapidity, one would like to determine the extension in rapidity of the individual sources. A qualitative examination of the distributions associated to events with 2, 3, 4 jets suggests that the extension of the individual source cannot be the same in the different families of events but, better, that it is larger in the 2-jet events and becomes narrower and narrower in passing to the configurations with 3 and 4 jets. The extension in y of the sources cannot become too narrow, if this should happen and the number of sources grows too high, the generating function would approach the corresponding expression for the Poissonian distribution.

When many sources are active the present description of the multiple production acquires many similarities with the ''clan'' description [9]. On the other hand, a feature of the two-source model discussed previously is that it is possible that only a part of the source is active. The description we start with is in fact differential in y. The model lacks information on the transverse dynamics which certainly enters also in the multiplicity distributions. In fact the total multiplicity is larger when the jets number is larger [8,9]. In the description presented here this would require that more than one source is active in the same rapidity interval, which looks very artificial if we neglect the transverse degrees of freedom but becomes quite natural when transverse degrees of freedom are taken into account. The model of multiple sources just described is still rather rough; in particular, one would not expect a sharp beginning and a sharp end for the rapidity range where the source is active. The present accuracy of the experimental data, however, does not allow us to discriminate the actual model from different possibilities. A further point is that the sources have been taken as equivalent: the presence of internal quantum numbers, which may affect the production mechanism [11], have not been taken into account.

A rather general feature, associated with the presence of different sources ordered in rapidity is a weak, long-range correlation in rapidity among the particles. This may be seen in the following way. The generating functional Eq. (5) is substituted by a product

$$f(\mathcal{U})=\prod_n [1-\mathcal{U}_n]^{-k}/[1-\tilde{u}_n]^{-k}; \tag{12}$$

every factor n acts in a different range of rapidity. If the two particles lie in the same rapidity interval, the two-body distribution is $D(\xi_1,\xi_2)=Ak(k+1)D(\xi_1)D(\xi_2)[1-\tilde{u}_{n'}]^2$; whereas if the two particles lie in different rapidity intervals, $D(\xi_1,\xi_2) = Ak^2D(\xi_1)D(\xi_2)[1-\tilde{u}_{n'}]\,[1-\tilde{u}_{n''}]$. In both cases $A=\Pi_n[1-\tilde{u}_n]^k$.

In conclusion the main points of the present analysis are summarized. The success of the NB in describing the multiparticle distributions supports the possibility that the NB is the actual distribution arising from a single source. The characterizing property of the NB is the self-similarity: if the source is unique, when considering a part of the spectrum one obtains the same NB distribution which describes the total spectrum. The large variation of the NB parameters as a function of the rapidity interval in multiparticle production is therefore a strong indication for the presence of many sources. The alternative possibility is the presence of a correlation within a single source. If the distribution in the whole spectrum is a NB, correlations most probably produce different distributions when looking at different parts of the spectrum. On the contrary, as in the model discussed above, the superposition of different sources, each giving rise to a NB distribution, can easily produce distributions which are close to a NB with altered parameters.

Hence one could consider, in high-energy processes with very high multiplicity, to use the self-similarity property in order to isolate different sources which are active in a complex production process. Events could be organized by considering different topologies, e.g., number of jets, impact parameter (in heavy-ion collisions), etc., and one could look at multiplicity distributions in different regions of phase space. The individual sources are isolated when, subdividing further the phase space regions, the corresponding multiplicity distributions are self-similar.

This work was partially supported by the Italian Ministry of University and of Scientific and Technological Research by means of the Fondi per la Ricerca scientifica–Università di Trieste.

APPENDIX

In this appendix only the two-body correlations are studied; so beyond the linear term $\mathcal{U}[J]=\int J(\xi)D(\xi)d\xi$, a term $\mathcal{V}[J,J]=1/2\int C(\xi_1,\xi_2)J(\xi_1)J(\xi_2)d\xi_1 d\xi_2$ is also used with the condition $\mathcal{V}[1,1]=0$. Then a generating functional, with these restrictions, can be expressed as $\mathcal{Z}=g(\mathcal{U}[J], \mathcal{V}[J,J])/g(\mathcal{U}[1])$ so that the corresponding generating function for the multiplicities is $z(\lambda)=g(\lambda\tilde{u})/g(\tilde{u})$. If one looks only at one part of the spectrum, then one can define the corresponding multiplicities according to Eq. (4), and the result is

$$z(\lambda)=g(\lambda u'+u'',\lambda^2 v'+2\lambda\overline{v+}v'')/g(\tilde{u}). \qquad (A1)$$

The terms $\tilde{u},u',u''$ have been already defined in Eq. (5). The definition of the v terms, where the symmetry of C has been used, is

$$v'=\frac{1}{2}\int C(\xi_1',\xi_2')d\xi_1' d\xi_2', \quad \overline{v}=\frac{1}{2}\int C(\xi_1',\xi_2'')d\xi_1' d\xi_2'',$$

$$v''=\frac{1}{2}\int C(\xi_1'',\xi_2'')d\xi_1'' d\xi_2''; \qquad (A2)$$

the initial condition $\mathcal{V}[1,1]=0$ is translated into $v'+2\overline{v}+v''=0$ which will be used in order to eliminate the term $\overline{v}$.

Now one can look to particular cases, the most interesting of which seems to be precisely a distribution which produces a NB multiplicity when integrated over the whole spectrum but contains two-body correlations. The simplest form in which the generating functional may be written is

$$\mathcal{Z}=f(\mathcal{U})=[1-\mathcal{U}-\mathcal{V}]^{-k}/[1-\tilde{u}]^{-k}, \qquad (A3)$$

and when only a part of the spectrum is observed and the rest is integrated over the generating function of the multiplicity is

$$z(\lambda)=[1-\tilde{u}]^k/[1-(u''+\lambda u')-(\lambda^2 v'+2\lambda\overline{v+}v'')]^k. \qquad (A4)$$

It is useful to write the same expression in a more compact form: i.e.,

$$z(\lambda)=N\cdot[1-\lambda a-\lambda^2 b]^{-k}, \qquad (A5)$$

having definitions

$$a=u'-v'-v''/(1-u''-v''), \quad b=v'/(1-u''-v''),$$
$$N=[1-a-b]^k=((1-\tilde{u})/(1-u''-v''))^k. \qquad (A6)$$

The new expression for the multiplicity distribution is now obtained by expanding $z(\lambda)$, as given in Eq. (A5), in powers of λ; the result is

$$z(\lambda)=N\sum_n (i\lambda\sqrt{b})^n C_n^{(k)}(ia/2\sqrt{b}), \qquad (A7)$$

where $C_n^{(k)}$ represents the Gegenbauer polynomial [10] of index k and order n. This kind of expansion does not look very transparent, but from the explicit form of the Gegenbauer polynomials it is easily seen that every term of the sum is real, as obviously it must be. It is also straightforward to verify that when the effect of the correlations vanishes, v', v'', b go to zero and the usual binomial distribution is recovered.

If the correlations are present but not very strong, the terms v will be small and one can perform an expansion in b. To the first order in b the expression of $z(\lambda)$ is

$$z(\lambda)=N\{[1-\lambda a]^{-k}(1-2kb/a^2)$$
$$+([1-\lambda a]^{-k+1}+[1-\lambda a]^{-k-1})kb/a^2\}. \qquad (A8)$$

With this expansion the original binomial distribution is reproduced, with some small correction for the parameter, but other satellite binomial distributions arise, whose exponent is shifted by ± 1, so that the distance from the original distribution increases with the power of the small parameter representing the effect of the correlations.

Also in the presence of a correlation there is the limiting relation between the NB and the Poissonian distribution. In a formal way this may be obtained through the definitions $\mathcal{U}=\mathcal{P}/k$, $\mathcal{V}=\mathcal{Q}/k$, $u=p/k$, $v=q/k$; then in the limit $k\to\infty$ out of Eqs. (A3), (A4) it results that $\mathcal{Z}=\exp[\mathcal{P}+\mathcal{Q}-p_1]$, $z(\lambda)=\exp[-(p'+q'')+\lambda(p'-q'-q'')+\lambda^2 q']$, and as far as Eq. (A7) is concerned, one can use the limiting expressions of the Gegenbauer polynomials yielding the Hermite polynomials [10].

What appears, beyond the details of the calculations that necessarily refer to simplified examples, is that in presence of two-body correlations the partial spectra are necessarily different from the complete ones: if the correlations play a minor role, then the NB distribution is approximately preserved, but with the too strong result of having a constant k parameter, strong correlations may simulate a variable k parameter but strongly modify the distribution, which is no longer a NB distribution (not even approximately).

[1] A. Giovannini and L. van Hove, Z. Phys. C **30**, 391 (1986); Acta Phys. Pol. B **19**, 495 (1988); **19**, 931 (1988); *Proceedings of the XVII International Symposium on Multiparticle Dynamics* (World Scientific, Singapore, 1989), p. 561.
[2] A. Giovannini, *Proceedings of the International Conference: Physics in Collisions VI* (World Scientific, Singapore, 1987), p. 39.
[3] E. A. DeWolf, I. M. Dremin, and W. Kittel, Phys. Rep. **270**, 1 (1996).
[4] L. S. Brown, Phys. Rev. D **5**, 748 (1972); Y. Akiyama and S. Hori, Prog. Theor. Phys. **48**, 276 (1973).
[5] R. Blankenbecler, Phys. Rev. D **8**, 1611 (1973).
[6] G. Calucci and D. Treleani, Int. J. Mod. Phys. A **6**, 4375 (1991).
[7] L. van Hove, Physica A **147**, 19 (1987).
[8] Delphi Collaboration, P. Abreu *et al.*, Z. Phys. C **56**, 63 (1992).
[9] F. Bianchi, A. Giovannini, S. Lupia, and R. Ugoccioni, Z. Phys. C **58**, 71 (1993); S. Lupia, A. Giovannini, and R. Ugoccioni, *ibid.* **66**, 195 (1994).
[10] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965), Chap. 22, p. 1968.
[11] C. K. Chew, D. Kiang, and H. Zhou, Phys. Lett. B **186**, 411 (1986); D. Ghosh *et al.*, Fizika (Zagreb) **B6**, 37 (1997).