NEGATIVE BINOMIAL MULTIPLICITY DISTRIBUTIONS
IN CONTINUOUS DOMAINS

L. Van Hove
CERN - Geneva

ABSTRACT

Prompted by experimental evidence on multiplicity distributions of particles produced in high energy collisions, we study the class of distributions of points in a continuous domain D_0 , characterized by the property that the multiplicity of points in D_0 and in any subdomain D of D_0 has a negative binomial distribution. We show that a distribution of this class is completely determined by the two parameters $Q_1(y) = dn/dy$ and $k(y)$ of the multiplicity distribution in infinitesimal neighbourhoods of all points y of D_0 . We prove that the distribution can be constructed by N -fold convolution of identical uncorrelated "clan" distributions, the number N of clans being itself Poisson-distributed. The clan distribution, which involves all correlations of the original distribution, is found to be very simply expressed in terms of $Q_1(y)$ and $k(y)$.

(To be published in Physica A)

1. Introduction

Recent work on multiplicity distributions of particles (mostly pions) produced in high energy collisions has revealed that in many cases they are very well described by the negative binomial (NB) distribution, a law frequently used in statistics to describe distributions broader than Poisson.¹⁻⁶ The NB distribution depends on two parameters only, usually taken to be the average multiplicity $\bar{n}$ and a positive parameter k related to the dispersion by

$$\langle n^2 \rangle - \bar{n}^2 = \bar{n}^{-1} + k^{-1}$$

The NB distribution reduces to the Poisson law for $k \rightarrow \infty$ and transforms into the ordinary binomial law when k is taken to be a negative integer.

Most studies have concentrated on total multiplicities of charged particles and on charged particle multiplicities in symmetric intervals $|y| \leq y_0$ of the centre-of-mass (c.m.) rapidity y . Although no full information is yet available, the data suggest the possibility that the NB law may hold with good precision for the multiplicity in *every* subdomain D of such an interval, the parameter k_D varying with D . The object of this paper is to study the class of probability distributions obeying this property. The main result will be that they are fully characterized by two functions of y , the one-particle distribution $Q_1(y) = \bar{n}_\Delta/\Delta$ and the value $k(y)$ of k_Δ , where Δ is an infinitesimal neighbourhood of y .

The paper is organized as follows. Section 2 gives the definitions, notations and assumptions. In section 3, the full distribution is determined in terms of $Q_1(y)$ and $k(y)$. The structure of the distribution and the positivity condition for all probabilities are studied in section 4. The concluding section discusses briefly the correlations contained in the distribution and its applicability to particle production processes. The proof of the necessary condition for positivity, which is due to Dr. T.T. Wu (Harvard), is presented in the appendix.

2. Definitions and Assumptions

Let D_0 be a domain in a space of one or more continuous variables which we denote by the single symbol y (for example, y can be the c.m. rapidity variable of particles and D_0 a symmetric interval $|y| \leq y_0$). Consider a distribution of points in D_0 (physically, they

can be the y -values of particles treated as indistinguishable), characterized by continuous probability densities $P_n(y_1 \dots y_n)$; $n = 1, 2, \dots$. The $P_n(y_1 \dots y_n)$ can be taken fully symmetric in $y_1, \dots, y_n$. They describe the distribution in D_0 when the multiplicity is exactly n . As far as D_0 is concerned, they are what high energy physicists call *exclusive distributions*. The corresponding *inclusive distributions* are given for $n = 1, 2 \dots$ by

$$Q_n(y_1 \dots y_n) = P_n(y_1 \dots y_n) + \sum_{m=1}^{\infty} (m!)^{-1} \int_{D_0} P_{n+m}(y_1 \dots y_n y'_1 \dots y'_m) dy'_1 \dots dy'_m \quad (2.1)$$

the inverse formula being

$$P_n(y_1 \dots y_n) = Q_n(y_1 \dots y_n) + \sum_{m=1}^{\infty} (-1)^m (m!)^{-1} \int_{D_0} Q_{n+m}(y_1 \dots y_n y'_1 \dots y'_m) dy'_1 \dots dy'_m \quad (2.2)$$

$Q_n(y_1 \dots y_n)$ is the probability density for n points to be at $y_1, \dots, y_n$ irrespective of the presence and location of any further points. The probability P_0 to have multiplicity zero is given by

$$P_0 = 1 - \sum_{n=1}^{\infty} (n!)^{-1} \int_{D_0} P_n(y_1 \dots y_n) dy_1 \dots dy_n \quad (2.3)$$

Extending (2.1) to $n = 0$ therefore suggests to define $Q_0 = 1$.

We are interested in $P_n(D, y_1 \dots y_n)$, the probability to have exactly $n \geq 1$ points in a subdomain D of D_0 , at positions $y_1, \dots, y_n$ (all in D), irrespective of the presence and location of points in the complement, $D_0 - D$. It is given by

$$P_n(D, y_1, \dots, y_n) = P_n(y_1 \dots y_n) + \sum_{m=1}^{\infty} (m!)^{-1} \int_{D_0-D} P_{n+m}(y_1 \dots y_n y'_1 \dots y'_m) dy'_1 \dots dy'_m \quad (2.4)$$

The probability $P_n(D)$ to have exactly n points in D is therefore

$$P_n(D) = (n!)^{-1} \int_D P_n(D, y_1, \dots, y_n) dy_1 \dots dy_n \quad (2.5)$$

For $n = 0$ it is

$$P_0(D) = P_0 + \sum_{m=1}^{\infty} (m!)^{-1} \int_{D_0-D} P_m(y'_1 \dots y'_m) dy'_1 \dots dy'_m \quad (2.6)$$

Using (2.3) to (2.5) one easily verifies

$$\sum_{n=0}^{\infty} P_n(D) = 1$$

As far as the subdomain D is concerned, the $P_n(D, y_1 \dots y_n)$ are the exclusive distributions, whereas the inclusive distributions are the same functions (2.1) as for D_0 . This implies that if all integrations in (2.2) are restricted to D , the lefthand side becomes $P_n(D, y_1 \dots y_n)$. Similarly, (2.1) also holds when all $P_i(y_1 \dots y_n)$ are replaced by $P_i(D, y_1, \dots, y_i)$ and the integrations are restricted to D . This gives

$$Q_n(D) \equiv \int_D Q_n(y_1 \dots y_n) dy_1 \dots dy_n = \sum_{m=0}^{\infty} (n+m)! (m!)^{-1} P_{n+m}(D) \quad (2.7)$$

The special case $n = 0$ reduces to $Q_0 = 1$. The integral (2.7) is also equal to the following average

$$Q_n(D) = \langle n_D (n_D - 1) \dots (n_D - n + 1) \rangle \quad (2.8)$$

where n_D is the multiplicity in subdomain D . One sees this first for given n_D , where it immediately follows from the definition of an inclusive distribution, and one then averages over n_D .

We now formulate our assumptions:

Assumption I - In D_0 and in every subdomain D of D_0 , the multiplicity distribution $P_n(D)$ is of negative binomial type, i.e.

$$P_0(D) = (1 - b_D)^{k_D}, P_n(D) = k_D (k_D + 1) \dots (k_D + n - 1) b_D^n (1 - b_D)^{k_D} / n! \quad (2.9)$$

$$b_D = \bar{n}_D / (\bar{n}_D + k_D), \quad \bar{n}_D = \int_D Q_1(y) dy, \quad k_D > 0 \quad (2.10)$$

$\bar{n}_D$ is the average multiplicity in D . It is related to $Q_1(y)$ by (2.7), (2.8).

Assumption II - For each point y of D_0 we have $Q_1(y) > 0$ and for each sequence of neighbourhoods Δ of y tending to y , k_Δ tends to a limit $k(y) > 0$ which depends only on y and not on the sequence of Δ 's.

Note that for such a sequence of neighbourhoods

$$\bar{n}_\Delta / \Delta \rightarrow Q_1(y) \quad (2.11)$$

as follows from the second equation (2.10). We now proceed to show that the two functions $Q_1(y)$, $k(y)$ completely determine the distribution $P_n(y_1 \dots y_n)$ for all n .

3. Determination of k_D and $Q_n(y_1 \dots y_n)$

By substituting (2.9) in (2.7) and summing the series

$$\sum_{m=0}^{\infty} (k_D + n) \dots (k_D + n + m - 1) b_D^m / m! = (1 - b_D)^{-k_D - n} \quad (3.1)$$

one obtains for any subdomain D ,

$$\begin{aligned} Q_n(D) &= k_D (k_D + 1) \dots (k_D + n - 1) b_D^n (1 - b_D)^{-n} \\ &= \bar{n}_D^n [1 + k_D^{-1}] \dots [1 + (n - 1) k_D^{-1}] \end{aligned} \quad (3.2)$$

We apply this first to $D = \Delta$ with Δ a small neighbourhood of a point y . Dividing by Δ^n and going to the limit $\Delta \rightarrow 0$ by using (2.11) and Assumption II, we obtain

$$Q_n(y \dots y) = [Q_1(y)]^n [1 + k(y)^{-1}] \dots [1 + (n - 1) k(y)^{-1}] \quad (3.3)$$

Take now m distinct points $y_1, \dots, y_m$ with non-overlapping neighbourhoods $\Delta_1, \dots, \Delta_m$ around them. Apply (3.2) to $D = \Delta$ with Δ the union $\Delta = \Delta_1 + \dots + \Delta_m$ of these neighbourhoods. For small Δ_i we obtain

$$\begin{aligned} \sum_{i_1=1}^m \dots \sum_{i_n=1}^m Q_n(y_{i_1} \dots y_{i_n}) \Delta_{i_1} \dots \Delta_{i_n} &= \bar{n}_\Delta^n [1 + k_\Delta^{-1}] \dots [1 + (n - 1) k_\Delta^{-1}] \\ \bar{n}_\Delta &= \sum_{i=1}^m \bar{n}_i, \quad \bar{n}_i = Q_1(y_i) \Delta_i \end{aligned}$$

This can be rewritten

$$\sum_{i_1=1}^m \dots \sum_{i_n=1}^m Q'_n(y_{i_1} \dots y_{i_n}) \alpha_{i_1} \dots \alpha_{i_n} = [1 + k_\Delta^{-1}] \dots [1 + (n - 1) k_\Delta^{-1}] \quad (3.4)$$

with the notations

$$Q'_n(y_1 \dots y_n) = Q_n(y_1 \dots y_n) / \prod_{j=1}^n Q_1(y_j) \quad (3.5)$$

$$\alpha_i = \bar{n}_i / \bar{n}_\Delta, \quad \sum_1^m \alpha_i = 1 \quad (3.6)$$

Eliminating one of the α_i , for example $\alpha_1 = 1 - \alpha_2 - \dots - \alpha_m$, we find that $[1 + k_\Delta^{-1}] \dots [1 + (n - 1) k_\Delta^{-1}]$ is a polynomial of order n or less in $\alpha_2, \dots, \alpha_m$. The case $n = 2$ implies that k_Δ^{-1} is at most of order 2. But order 2 is excluded, because it would make

$[1 + k_{\Delta}^{-1}] [1 + 2k_{\Delta}^{-1}]$ of order 4 instead of 3. Hence for $n = 2$ the lefthand side of (3.4) has no quadratic terms after the substitution $\alpha_1 = 1 - \alpha_2 \dots - \alpha_m$, which is equivalent to the property that it vanishes after the substitution $\alpha_1 = -\alpha_2 \dots - \alpha_m$. Consequently we have

$$\sum_{i_1=1}^m \sum_{i_2=1}^m Q'_2(y_{i_1} y_{i_2}) \alpha_{i_1} \alpha_{i_2} = (\alpha_1 + \dots \alpha_m) (\alpha_1 V_1 + \dots \alpha_m V_m) \quad (3.7)$$

for all values of $\alpha_1, \dots \alpha_m$. Identifying the terms in α_i^2 we obtain

$$V_i = Q'_2(y_i y_i) = 1 + k(y_i)^{-1} \quad (3.8)$$

where we have used (3.3) and (3.5). Re-introducing again the values (3.6) in (3.7), we find with (3.8) and (3.4) for $n = 2$

$$k_{\Delta}^{-1} = \sum_1^m \bar{n}_i k(y_i)^{-1} / \sum_1^m \bar{n}_i$$

This can be introduced in (3.4) for general n and gives

$$\begin{aligned} \sum_{i_1=1}^m \dots \sum_{i_n=1}^m Q'_n(y_{i_1} \dots y_{i_n}) \bar{n}_{i_1} \dots \bar{n}_{i_n} = \\ \left(\sum_1^m \bar{n}_i \right) \left\{ \sum_1^m \bar{n}_i [1 + k(y_i)^{-1}] \right\} \dots \left\{ \sum_1^m \bar{n}_i [1 + (n-1) k(y_i)^{-1}] \right\} \end{aligned}$$

Remembering the symmetry of Q'_n , we obtain by identification

$$\begin{aligned} Q'_n(y_1 \dots y_n) &= \left[\prod_1^n Q_1(y_i) \right]^{-1} Q_n(y_1 \dots y_n) \\ &= \left\{ [1 + k(y_2)^{-1}] \dots [1 + (n-1) k(y_n)^{-1}] \right\}_s \end{aligned} \quad (3.9)$$

where the bracket $\{\dots\}_s$ indicates that the expression must be fully symmetrized in $y_1 \dots y_n$ (average over all relevant permutations). Eq. (3.9) reduces to (3.3) when $y_1 = \dots = y_n$.

This completes the determination of all $Q_n(y_1 \dots y_n)$ and hence of all $P_n(y_1 \dots y_n)$ in terms of $Q_1(y)$ and $k(y)$. We can now return to (3.2) for $n = 2$ and general subdomain D . With (3.9) for $n = 2$ it leads to

$$\bar{n}_D / k_D = \int_D [Q_1(y) / k(y)] dy \quad (3.10)$$

Since $\bar{n}_D = \int_D Q_1(y) dy$, (3.10) means that k_D^{-1} for any subdomain is the average of $k(y)^{-1}$ over that subdomain with the inclusive one-point distribution $Q_1(y)$ as weight.

4. Generating Functional and Clan Distribution

To analyze the distribution determined by (3.9) and (2.2), it is convenient to consider the generating functional F defined by

$$F \equiv P_0 + \sum_1^{\infty} (n!)^{-1} \int_{D_0} P_n(y_1 \dots y_n) z(y_1) \dots z(y_n) dy_1 \dots dy_n \quad (4.1)$$

where $z(y)$ is an arbitrary function of y in D_0 . The substitution

$$z(y) = 1 + u(y) \quad (4.2)$$

gives through (2.1) the alternative expansion

$$F = 1 + \sum_1^{\infty} (n!)^{-1} \int_{D_0} Q_n(y_1 \dots y_n) u(y_1) \dots u(y_n) dy_1 \dots dy_n \quad (4.3)$$

Inserting (3.9) into this equation, one obtains a series involving the two linear functionals

$$R \equiv \int_{D_0} Q_1(y) u(y) dy = \int_{D_0} Q_1(y) z(y) dy - \bar{n}_0 \quad (4.4)$$

$$S \equiv \int_{D_0} Q_1(y) k(y)^{-1} u(y) dy = \int_{D_0} Q_1(y) k(y)^{-1} z(y) dy - \bar{n}_0 k_0^{-1} \quad (4.5)$$

By $\bar{n}_0$, k_0 we denote the values of $\bar{n}_D$ and k_D for the complete domain $D = D_0$. The series is

$$F = 1 + R + R \sum_2^{\infty} (n!)^{-1} [R + S] \dots [R + (n-1)S] \quad (4.6)$$

and it can be summed explicitly: it is the Taylor expansion of

$$F = [1 - S]^{-R/S} \quad (4.7)$$

in powers of the S appearing in the bracket.

It is now apparent that the structure of our distribution is a complex one due to the exponent R/S in the functional (4.7), and in general the $P_n(y_1 \dots y_n)$ are quite complicated. Only the case of constant $k(y) = k_0$ is simple, because R/S then reduces to k_0 and is no longer a functional. In this special case the generating functional F reduces to

$$F = \left[1 + \bar{n}_0 k_0^{-1} - k_0^{-1} \int_{D_0} Q_1(y) z(y) dy \right]^{-k_0} \quad (4.8)$$

and $k_D = k_0$ for all subdomains D . Such distributions have been considered by many authors, under the names of multivariate negative binomial or negative multinomial distributions.* In high energy physics the multiplicity data clearly show that k_D varies with D , and our main concern is to elucidate this general case. We shall find in particular the amount of variation of $k(y)$ which is compatible with the requirement that all $P_n(y_1 \dots y_n)$ are non-negative.

While the generating functional (4.7) is quite complicated when $k(y)$ varies with y , it turns out that its logarithm is much simpler. It has also a probabilistic interpretation which gives considerable insight in the various correlations contained in the $P_n(y_1 \dots y_n)$. We write

$$F = \exp [\bar{N} (f - 1)] = \exp (-\bar{N}) \sum_0^{\infty} f^N \bar{N}^N / N! \quad (4.9)$$

where $\bar{N}$ is a positive number and f a functional

$$\begin{aligned} f &= p_0 + \sum_1^{\infty} (n!)^{-1} \int_{D_0} p_n(y_1 \dots y_n) z(y_1) \dots z(y_n) dy_1 \dots dy_n \\ &= 1 + \sum_1^{\infty} (n!)^{-1} \int_{D_0} q_n(y_1 \dots y_n) u(y_1) \dots u(y_n) dy_1 \dots dy_n \end{aligned} \quad (4.10)$$

We interpret the $p_n(y_1 \dots y_n)$ as describing a new distribution which we call the *clan distribution* (following the terminology introduced in Ref. 5). The $p_n(y_1 \dots y_n)$ are the exclusive distributions of a clan, the $q_n(y_1 \dots y_n)$ the corresponding inclusive distributions, and one can repeat for these functions all statements made in Section 2 for the $P_n(y_1 \dots y_n)$ and $Q_n(y_1 \dots y_n)$.

The probabilistic interpretation of (4.9) is that the original distribution $P_n(y_1 \dots y_n)$ is generated by combining a variable number N of clan distributions, with for N a Poisson distribution of mean $\bar{N}$. There are no correlations between different clans, and all the correlations of the $P_n(y_1 \dots y_n)$ are therefore induced by those in a single clan. This interpretation is valid if and only if p_0 and all $p_n(y_1 \dots y_n)$ are non-negative. The probability p_0 of having zero particle in a clan is obtained by putting $z(y) = 0$ in (4.7), (4.9) and (4.10). The result is

$$(1 + \bar{n}_0 k_0^{-1})^{-k_0} = \exp [\bar{N} (p_0 - 1)]$$

* The author is indebted to Prof. J.P. Imhof (Geneva) for bringing this work to his attention. For reviews, see Refs. 7,8.

or equivalently

$$1 - p_0 = k_0 \ell_0 / \bar{N}, \quad \ell_0 = \ln(1 + \bar{n}_0 / k_0) \quad (4.11)$$

The condition $p_0 \geq 0$ is equivalent to

$$\bar{N} \geq k_0 \ell_0 \quad (4.12)$$

For $n \geq 1$ all $p_n(y_1 \dots y_n)$, $q_n(y_1 \dots y_n)$ will be proportional to $\bar{N}^{-1}$, i.e., to $1 - p_0$, and one can choose $p_0 = 0$, $\bar{N} = k_0 \ell_0$ without loss of generality.

From (4.7) and (4.9) one obtains

$$f = 1 - \bar{N}^{-1} R \ln[1 - S] / S = 1 + \bar{N}^{-1} R \sum_0^{\infty} S^n / (n + 1)$$

The identification with (4.10) using (4.4), (4.5) gives immediately the $q_n(y_1 \dots y_n)$:

$$q_1(y) = Q_1(y) / \bar{N} \quad (4.13)$$

$$q_n(y_1 \dots y_n) = (n - 1)! \{Q_1(y_1) v(y_2) \dots v(y_n)\}_s / \bar{N} \quad (4.14)$$

where we defined

$$v(y) = Q_1(y) / k(y) \quad (4.15)$$

and $\{\dots\}_s$ denotes again symmetrization in $y_1 \dots y_n$. When one inserts these $q_n(y_1 \dots y_n)$ in the analog of (2.2) for $p_n(y_1 \dots y_n)$, the resulting series can be simplified by means of the identity

$$\begin{aligned} s^n / n + \sum_{m=1}^{\infty} (-1)^m (m!)^{-1} n(n+1) \dots (n+m-1) s^{m+n} / (m+n) = \\ \int_0^s (1+x)^{-n} x^{n-1} dx = \int_0^{s/(1+s)} (1-b)^{-1} b^{n-1} db \end{aligned} \quad (4.16)$$

One then obtains for $n \geq 1$,

$$p_n(y_1 \dots y_n) = n! \left\{ [\ell_{n-1} Q_1(y_1) - \ell_n k_0 v(y_1)] v(y_2) \dots v(y_n) \right\}_s k_0^n / (\bar{N} \bar{n}_0^n) \quad (4.17)$$

with

$$\ell_n = \sum_{m=n+1}^{\infty} b_0^m / m = \ell_0 - \sum_1^n b_0^m / m, \quad b_0 = \bar{n}_0 / (\bar{n}_0 + k_0) \quad (4.18)$$

and $\ell_0 = \ln(1 + \bar{n}_0/k_0) = -\ln(1 - b_0)$ as in (4.11). From (4.17) one immediately sees that $p_n(y_1 \dots y_n) \geq 0$ for all y_i if and only if for all y , $\ell_{n-1}Q_1(y) \geq \ell_n k_0 v(y)$ or equivalently

$$k(y)/k_0 \geq \ell_n/\ell_{n-1} \quad (4.19)$$

It is easy to show that ℓ_n/ℓ_{n-1} increases with n ; show instead that $\ell_{n-1}(\ell_{n-1} - \ell_n)^{-1} = 1 + nb_0(n+1)^{-1} + nb_0^2(n+2)^{-1} + \dots$ increases. For $n \rightarrow \infty$, the latter series tends to $(1 - b_0)^{-1}$ and ℓ_n/ℓ_{n-1} therefore to b_0 (remember that one has always $b_0 < 1$). The necessary and sufficient condition for all $p_n(y_1 \dots y_n)$ to be non-negative is therefore

$$k(y) \geq k_0 \bar{n}_0 / (\bar{n}_0 + k_0) \quad (4.20)$$

Remarkably this is also the necessary and sufficient condition for all $P_n(y_1 \dots y_n)$ to be non-negative. It is sufficient, because $p_n(y_1 \dots y_n) \geq 0$ for all $n, y_1, \dots, y_n$ implies $P_n(y_1 \dots y_n) \geq 0$ by (4.9) and the first equation (4.10). It is also necessary, because $P_n(y \dots y) \geq 0$ for all n implies (4.20). The latter result is due to Dr. T.T. Wu and is proven in the appendix. This completes the characterization of the class of distributions defined by Assumptions I and II.

5. Concluding Remarks

The distributions defined and characterized in this paper are of a novel type and further work will be needed to study their properties and their applicability. All we do here is to present a few remarks.

The positivity condition (4.20) must be considered together with the relations (2.10) and (3.10) taken for $D = D_0$:

$$\bar{n}_0 = \int_{D_0} Q_1(y) dy, \quad \bar{n}_0/k_0 = \int_{D_0} Q_1(y) k(y)^{-1} dy \quad (5.1)$$

The latter mean that

$$\langle k_0/k(y) \rangle_{D_0} = 1 \quad (5.2)$$

where the average over D_0 is taken with the weight $Q_1(y)$. Condition (4.20) says that $k_0/k(y)$ cannot be larger than $b_0^{-1} = 1 + k_0 n_0^{-1}$, and (5.2) that $k_0/k(y)$ must have average value 1. This implies that for $k_0 \ll n_0$ the function $k(y)$ must be close to k_0 almost everywhere, whereas it can vary over a wide range for $k_0 \gg n_0$.

As mentioned already in connection with (4.9), the distribution $P_n(y_1 \dots y_n)$ is generated by a Poisson distribution of uncorrelated clans. All correlations are therefore contained within the clan distribution, eqs. (4.13) to (4.18). Correlations can best be examined on the inclusive distributions $q_n(y_1 \dots y_n)$, especially on the relation between any q_n and the q_m for $m < n$. From (4.13) and (4.14) one derives the simple recurrence relation ($n \geq 1$),

$$q_{n+1}(y_1 \dots y_{n+1}) = n \{q_n(y_1 \dots y_n) v(y_{n+1})\}_s \quad (5.3)$$

It involves two types of correlations. One is expressed by the factor n and reflects the fact that the multiplicity distribution of a clan

$$p_n(D_0) = (n!)^{-1} \int_{D_0} p_n(y_1 \dots y_n) dy_1 \dots dy_n$$

obeys the recurrence relation ($n \geq 1$),

$$(n+1) p_{n+1}(D_0) / p_n(D_0) = b_0 n \quad (5.4)$$

A constant value of the ratio $(n+1) p_{n+1}(D_0) / p_n(D_0)$ would correspond to a Poisson distribution. This first correlation already occurs for $k(y)$ independent of y . Its intuitive meaning is that the presence of n points enhances by a factor n the probability of having an additional point. In particle physics it can be interpreted in terms of a cascade-type mechanism for particle production.^{2,5}

The second correlation effect in (5.3) results from the variation of $k(y)$ with y , i.e., the non-proportionality of $q_1(y)$ and $v(y)$. In the n -point distributions of a clan (inclusive and exclusive), there are $n-1$ points distributed as $v(y)$ and one point with a different distribution (always $q_1(y)$ for the inclusive n -point distribution, an n -dependent function of y for the exclusive one). The interpretation of this effect requires further study. It should be noted that both correlations contained in (5.3) are of long-range type, extending over the whole domain D_0 .

In conclusion, we comment on the implications of our results for particle production. As mentioned in the introduction, many multiplicity distributions observed in high energy particle production processes are very well represented by the negative binomial (NB) law (2.9). This holds best for multiplicity distributions in intervals of c.m. rapidity y which are not significantly constrained by conservation laws. Energy-momentum conservation

always cuts off large multiplicities. This effect is only significant in rapidity intervals near the rapidities $\pm y_{inc}$ of the incident particles, and indeed multiplicity distributions in such intervals are found to be narrower than Poisson and therefore not of NB type. Charge conservation has an effect which is most significant for total multiplicities (complete suppression of charged particle numbers of one parity) and is therefore also significant for large phase space domains, although good NB fits are usually found when considering separately even and odd multiplicities of charged particles.

It is in symmetric rapidity intervals $|y| \leq y_0$ that the best NB fits have been found, especially at the lower energies. Although complete experimental information is not yet available, we can already test our assumptions I and II (Section 2) by trying them out on such intervals. The first test is provided by eq. (3.10) which implies that $\bar{n}_D/k_D$ increases when the subdomain D increases to D_0 . If we take D to be $|y| \leq y_1$ with increasing y_1 , the pp and π^+p data at $\sqrt{s} = 22\text{GeV}^4$ and the e^+e^- data at $\sqrt{s} = 29\text{GeV}^3$ show that $\bar{n}_D/k_D$ increases up to $y_1 \sim 1.5$ and decreases for larger y_1 , whereas for $\bar{p}p$ at $\sqrt{s} = 546\text{GeV}$, $\bar{n}_D/k_D$ continues to grow up to $y_1 \sim 2.5 - 3.0$ and then becomes essentially constant (in the latter experiment y is the pseudorapidity). Hence, our first test does not limit D_0 for the $\bar{p}p$ experiment but limits it to $|y| \lesssim 1.5$ for the lower energy ones.

The second test exploits (4.20). Where available, the data indicate that k_D is smallest for small symmetric intervals around $y = 0$. If our assumptions hold for the domain D_0 defined by $|y| \leq y_0$, eq. (4.20) implies that $k_0\bar{n}_0/(k_0 + \bar{n}_0)$ cannot be larger than the minimum k_D measured. In all experiments mentioned above one finds that this limits y_0 to values $\lesssim 1.5$. For the $\bar{p}p$ experiment at $\sqrt{s} = 546\text{GeV}$, this can be read off directly from Fig. 1 of Ref. 2 by noting that $a = k_0\bar{n}_0/(\bar{n}_0 + k_0)$ and that the lowest measured k value is 1.6 for $|y| \leq 0.2$ (Table 1 of Ref. 1).

These results show that our assumptions I and II hold at most in a rapidity interval $|y| \lesssim 1.5$. Whether they actually do hold in that interval requires further experimental information which is not yet available. But we know that NB multiplicity distributions occur in much wider symmetric rapidity intervals, which therefore contain subdomains where the multiplicity distribution is presumably not of NB type. It should be possible to study this experimentally in order to reach a more complete characterization of the complete distributions.

6. Acknowledgments

The author is greatly indebted to T.T. Wu for his proof of the necessary character of condition (4.20), given in the Appendix. The author also acknowledges many useful discussions, especially with A. Białas, A. Giovannini, W. Kittel, F. Meijers, and with the mathematical statisticians J.P. Imhof, F. Streit and G.S. Watson. He is indebted to Brookhaven National Laboratory and Mrs. I. Harrity for the preparation of the typescript.

Appendix - For a subdomain D of D_0 , consider the probability $\tilde{P}_n(D)$ to have exactly n points in D and no points in the complement $D_0 - D$. It is given by

$$\tilde{P}_0(D) = P_0, \quad \tilde{P}_n(D) = (n!)^{-1} \int_D P_n(y_1 \dots y_n) dy_1 \dots dy_n \quad (n \geq 1) \quad (A.1)$$

The corresponding generating function

$$\tilde{F}_D = \sum_0^{\infty} \tilde{P}_n(D) z^n \quad (A.2)$$

is obtained from the functional (4.1) by the substitution $z(y) = z$ in D , $z(y) = 0$ in $D_0 - D$. With this substitution (4.4) and (4.5) become

$$R = \bar{n}_D z - \bar{n}_0, \quad S = \bar{n}_D k_D^{-1} z - \bar{n}_0 k_0^{-1}$$

Using (4.7) one therefore obtains for the function (A.2)

$$\tilde{F}_D = [1 + \bar{n}_0 k_0^{-1} - \bar{n}_D k_D^{-1} z]^{(\bar{n}_0 - \bar{n}_D z) / (\bar{n}_D k_D^{-1} z - \bar{n}_0 k_0^{-1})} \quad (A.3)$$

As will be shown presently, $\tilde{P}_n(D) \geq 0$ for all n implies

$$k_D \geq k_0 \bar{n}_0 / (\bar{n}_0 + k_0) \quad (A.4)$$

If D is a neighbourhood of y tending to y [in which case $\tilde{P}_n(D) D^{-n} \rightarrow (n!)^{-1} P_n(y \dots y)$], we have $k_D \rightarrow k(y)$ and (A.4) reduces to (4.20), which proves the necessity of the latter condition.

A very simple method to derive (A.4) was found by Dr. T.T. Wu. From (A.3) one gets

$$\ln \tilde{F}_D = (\bar{n}_0 - \bar{n}_D z) \ln (1 + s_0 - s_D z) / (s_D z - s_0) \quad (A.5)$$

with the notations

$$s_0 = \bar{n}_0 k_0^{-1}, \quad s_D = \bar{n}_D k_D^{-1} \quad (A.6)$$

As seen in (A.5), $\ln \tilde{F}_D$ has no singularity at $z = s_0/s_D$. It is holomorphic for

$$|z| < (1 + s_0) / s_D \equiv \rho_D \quad (A.7)$$

and the same holds for $\tilde{F}_D$. Take z real and let it tend to ρ_D from below. Then $\ln (1 + s_0 - s_D z)$ tends to $-\infty$, whereas

$$(\bar{n}_0 - \bar{n}_D z) / (s_D z - s_0) \rightarrow \bar{n}_0 - k_D (1 + s_0) \quad (A.8)$$

If the latter limit were positive, we would conclude $\ln \tilde{F}_D \rightarrow -\infty$ and $\tilde{F}_D \rightarrow 0$, which is impossible when all $\tilde{P}_n(D) \geq 0$. This proves the inequality $\bar{n}_0 \leq k_D (1 + s_0)$ which is identical to (A.4).

REFERENCES

1. UA5 Collaboration, G.J. Alner et al., Phys. Lett. 160B (1985)193.
2. A. Giovannini and L. Van Hove, Z. Phys. C 30 (1986)391.
3. HRS Collaboration, M. Derrick et al., Phys. Lett. B168 (1986)299.
4. NA22 Collaboration, M. Adamus et al., Phys. Lett. B177 (1986)239.
5. L. Van Hove and A. Giovannini, *Negative Binomial Multiplicity Distributions, a New Empirical Law for High Energy Collisions*, XVII International Symposium on Multi-particle Dynamics, Seewinkel, Austria, 16-20 June 1986, Torino University Preprint DFTT 11/86, to appear in the Proceedings (World Scientific Publishing Company, Singapore).
6. P. Carlson, *Multiplicity Distributions*, Review presented at the XXIII International Conference on High Energy Physics, Berkeley, California, 16-23 July 1986.
7. M. Sibuya, I. Yoshimura and R. Shimizu, Annals Institute of Statistical Mathematics, Tokyo, 16 (1964)409.
8. N.L. Johnson and S. Kotz, *Discrete Distributions*, Wiley, New York (1969)chapt. 11, sect. 3.