# What can be deduced from the multiplicity distributions

# <u>G. Wilk</u>,

NCNR, Warsaw, Poland in collaboration with **Z. Włodarczyk and M. Rybczyński,** UJK, Kielce, Poland

# Content:

- (1) Introduction
- (2) Closer look at P(N): R=fit/data ratio not flat , why?
- (3) Combinants obtained from data are oscillating, why?
- (4) Enhanced void probability: P(0) > P(1), why?
- (5) Compound distributions
- (6) Summary

<u>Based on</u>: J. Phys. G44 (2017) 015002, EPJ Web of Conf. 141 (2017) 01005, Int. J. Mod. Phys. A33 (2018)1830008, arXiv: 1811.07197, 1812.08840 (N + 1)P(N + 1) = g(N)P(N) where  $g(N) = \alpha + \beta N$ , (1)

The most popular forms of P(N) emerging from this recurrence are (p denotes probability of particl emission) :

$$P(N) = \frac{K!}{N!(K-N)!} p^{N}(1-p)^{K-N}, \quad \alpha = \frac{Kp}{1-p}, \quad \beta = -\frac{\alpha}{K}; \quad (BD)$$

$$P(N) = \frac{\lambda^{N}}{N!} \exp(-\lambda), \quad \alpha = \lambda, \quad \beta = 0; \quad (PD)$$

$$P(N) = \frac{\Gamma(N+k)}{\Gamma(N+1)\Gamma(k)} p^{N}(1-p)^{k}, \quad \alpha = kp, \quad \beta = \frac{\alpha}{k}; \quad (NBD)$$





2-NBD improves only agreement at large N, the ratio R = data/fit deviates dramatically from unity at small N for all fits.

There must be some additional information hidden in the small N region, not investigated yet [ J. Phys. G44 (2017) 015002 - Modified Negative Binomial - (MNB)].



### Some comments on NBD and MNBD:



Fig. 5 (Color online) Charged particle multiplicity dependence of the multiplicity in a clan,  $n_C(N)$  in the DCM scenario and in the standard NBD. The limiting value  $n_C = (ak)^{-1}$  for the DCM case is indicated.

In the clan model (CM) one allows for the production of very heavy clans, in the dynamical CM (DCM) there is a limit for this growth corresponding to a mass of 2.1 GeV.

The growth of the energy results in the formation of new clusters rather than in the production of heavier ones .

Newly produced particles find themselves more likely in new clans than remaining in the old ones.

This is reflected in the k-1 = P1(2)/P2(2), < 1.</li>
→ that particles more likely occur in two, not in one, clans.

This limiting mass corresponds to the very old idea of the so called H-quanta as intermediate objects being produced in high energy collisions: S. Hasegawa, PTP26(1961)150; 29(1963)128

Some comments on NBD and MNBD (cont):

- (\*) The clan model of multiparticle production employs a cascading mechanism of particle production in which one distinguishes groups of particles of common ancestry, called clans.
- (\*) Clans are supposed to be produced in an independent way, therefore their multiplicity is poissonian.
- (\*) Assuming that particles inside a clan are produced according to a logarithmic distribution, one obtains NBD. In the standard scenario the multiplicity of clans for a given energy Js and in some fixed rapidity window remains constant,  $\langle N_c \rangle = const$ , and, in a single event, the average number of particles per clan, increases with multiplicity N.
- (\*) In such approach the parameters m and k are related to the numbers of clans and to the amount of aggregation between clans and the parameter 1/k is interpreted as a measure of the aggregation of particles into clans with P1,2(2) being the probabilities to have two particles in, respectively, 1 or 2 clans.  $\frac{1}{k} = \frac{P_1(2)}{P_2(2)}$

- (\*) There is a room for changes in the P(N) resulting in agreement with data in the whole region of N → The recurrence relation (1) is too restricted to be helpful.
- (\*) Use a more general form of the recurrence relation used in counting statistics when dealing with cascade stochastic processes [B.E.A.Saleh and M.K.Teich, Proc. IEEE 70, 229 (1982)]

$$(N+1)P(N+1) = \langle N \rangle \sum_{j=0}^{N} C_j P(N-j).$$
 (2)

(\*) Contrary to (1), it connects now all multiplicities by means of some coefficients  $C_j \rightarrow$  they contain the memory of particle N + 1 about all the N - j previously produced particles.

- (\*) Coefficients C<sub>j</sub> replace ratio R = data/fit in quality assessment of P(N)
- (\*) They can be directly calculated from the experimentally measured P(N) by reversing (2):

$$\langle N \rangle C_j = (j+1) \left[ \frac{P(j+1)}{P(0)} \right] - \langle N \rangle \sum_{i=0}^{j-1} C_i \left[ \frac{P(j-i)}{P(0)} \right]$$
(3)

(\*) So far they were used already in: V.D.Rusov et al. PLB504(2001)213 and NPA764(2006)460 to multiparticle phenomenology.

# Result:

- (\*) C<sub>j</sub> obtained from data show distinct oscillatory behaviour gradually disappearing with N.
- (\*) It can be reproduced only by the MNB model (for which R(N) = 1 for all N).



- (\*) Oscillations of  $C_j$  are seen for different pseudorapidity windows and in data from all LHC experiments and energies.
- (\*) The only condition is that statistics of the experiment is high enough, for small statistics oscillations became too fuzzy to be recognized.

### It means that:

(\*) The single NBD is not able to reproduce data because its **C**<sub>j</sub> do not oscillate, they are equal to:

$$C_j = \frac{k}{\langle N \rangle} p^{j+1} = \frac{k}{k+m} \exp(j \ln p).$$

(\*) If we limit ourselves only to NBD, oscillations occur only for multi-NBD

### Summarizing the NBD case:

- (\*) Single NBD is not able to reproduce data, oscillations can occur only for combinations of NBD , however for the 2-NBD we were not able to find parameters of 2-NBD allowing for reasonable description of P(N) and C<sub>j</sub> at the same time.
- (\*) The best result so far is the 3-NBD fit proposed by I. J. Zborovsky, J.Phys. G40(2013) 055005, in which only P(N) and R were considered and  $C_j$  were not used. It is based on the claim that there is a place in data for a third component aiming to describe the low N events (what agrees with our observation that the cause of this effect seems to be localized at small N ).



Our fits using 3NBD with (Z) parameters [IJMPA33 (2018)1830008].

Note: (\*) improved agreement with data,

But: (\*) the low N region of P(N) still shows some deviations resulting in R departing from unity at small N and in  $C_j$  missing data for large  $j_{.14}^{.14}$ 



Our fits using 3NBD with (Z) parameters [IJMPA33 (2018)1830008].

Note: (\*) improved agreement with data, But:

(\*) the low N region of P(N) still shows some deviations resulting in R departing from unity at small N and in  $C_j$  missing data for large  $j_{.15}^{.15}$ 

(\*) It turns out that coefficients  $C_j$  oscillate in all distributions of the BD type, they are equal to:

$$C_j = (-1)^j \frac{K}{\langle N \rangle} \left( \frac{\langle N \rangle}{K - \langle N \rangle} \right)^{(j+1)} = \frac{(-1)^j}{1 - p} \left( \frac{p}{1 - p} \right)^j$$

- (\*) The amplitude of oscillations depends on the emission probability p: it increases with rank j for p > 0.5 and decreases for p < 0.5, period of oscillations is 2.
- (\*) However, they lacks the fading down feature of the C<sub>j</sub> observed experimentally





### BD used alone cannot explaindata

From recurrence relation: modified combinants C<sub>j</sub>.

<u>But from generating functions</u>:  $G(z) = \sum_{N=0}^{\infty} P(N) z^N$  : <u>combinants C<sup>\*</sup></u> (#)



(#) S.K.Kauffmann,M.Gyulassy,J. Phys. A11(1978)1715 (1978);
 R.Vasudevan et al., J.PA17(1984)989;
 S. Hegyi, PIB309(1993)443, PLB318(1993)642, PIB463(1999)126;
 R.Botet,M.Płoszajczak, Universal fluctuations, The phenomenology of hadronic matter,(WS 2002);
 W.Kittel and E.A.De Wolf, Soft Multihadron Dynamics, (WS 2005)

# What oscillations tell us?



C<sup>\*</sup>, were used in "Description of pion multiplicities using combinants" by A.B.Balantekin, J.E.Seger, PIB266(1991)231, arguing that:

"Combinants can be a useful tool to distinguish between bosons coming from the secondary decay of other particles such as deltas and bosons emitted from thermally equilibrated sources."

Two scenarios were considered:

- (i) N sources emitting bosons without any restrictions on their number **NBD and smooth and diminishing combinants**.
- (ii) M sources emitting only a limited number of bosons each
   BD and oscillating combinants.

Modified combinants  $C_{i}$  should follow the same behavior

### Comments:

(\*) Multiplicity distributions P(N) are usually studied by analyzing factorial moments:

$$F_q = \sum_{N=a}^{\infty} N(N-1)(N-2)\dots(N-q+1)P(N)$$

cumulant factorial moments:

$$K_q = F_q - \sum_{i=1}^{q-1} {\binom{q-1}{i-1}} K_{q-i}F_i$$

or their ratios:

$$H_q = \frac{K_q}{F_q}$$

which are very sensitive to the details of the multiplicity distribution.

- (\*) They seem to be well described by perturbative QCD considerations, especially their oscillations in sign as a function of the rank q.
- (\*) Note that  $K_q$  can be expressed as an infinite series of  $C_j$  and  $C_j$  can be expressed as an infinite series of  $K_q$ :

$$K_q = \sum_{j=q}^{\infty} \frac{(j-1)!}{(j-q)!} \langle N \rangle C_{j-1} \quad \text{and} \quad C_j = \frac{1}{\langle N \rangle} \frac{1}{(j-1)!} \sum_{p=0}^{\infty} \frac{(-1)^p}{p!} K_{p+j}$$

- (\*) By analogy to factorial cumulants, the combinants can be understood as *exclusive correlation integrals* .
- (\*) However, cumulants and combinants differ in the region of phase space they are most suitable to study:
  - cumulants are particularly well suited for the study of densely populated phase-space bins,
  - combinants are better suited for the study of sparsely populated regions, and their calculation requires only a finite number of P(N), with N < j; which compensates the drawback caused by the requirement that one must have P(O) > O.
- (\*) Combinants are finite combinations of the probability ratios P(N)/P(O), therfore they do not suffer from a bias (empty-bin effect) present at high resolution in factorial moments and cumulants.
- (\*) Combinants share with cumulants property of additivity.

### Enhanced void probability: P(0) > P(1)



Experimental smooth multiplicity distributions P (N) displayed for low multiplicities and for energies ranging from 0.2 TeV up to 8 TeV. Note the peculiar enhancement of the void probability P (0) (rather small at 0.2 TeV but quite substantial at 8 TeV).

It can be reproduced only by using a 2-component compound binomial distribution (BD and NBD).

**Enhanced void probability:** Void probability is strongly correlated with the modified combinants:

$$P(\mathbf{0}) = \exp\left(-\sum_{j=0}^{\infty} \frac{\langle N \rangle}{j+1} C_j\right)$$

(\*) The P(0) > P(1) property is possible only when  $\langle N \rangle C_0 < 1$ . For most multiplicity distributions we also have that P(2) > P(1), which results in additional condition:  $C_1 > C_0 (2 - \langle N \rangle C_0)$ , together they result  $C_1 > C_0$ .

(\*) This initial increase of Cj cannot continue for all ranks j; because of the normalization condition  $\sum_{j=0}^{\infty} C_j = 1$  we should observe some kind of nonmonotonic behaviour of Cj with rank j.

(\*) Therefore, all multiplicity distributions for which the modified combinants Cj decrease monotonically with rank j (like, for example, the NBD) do not exhibit the enhanced void probability.

Neither NBD or BD alone can describe data

**CD**: production process consists of M objects produced according to distribution f(M) (defined by generating function F(z)), which subsequently decay independently into a number of secondaries,  $n_{i=1,...,M}$ , following some other (the same for all M) distribution, g(n) (defined by a generating function G(z)) (#).

The resultant multiplicity distribution,  $h(N) = f \otimes g$ ,  $N = \sum_{i=0}^{n} n_i$  is compound distribution of f and g

with generating function:

H(z) = F[G(z)]

<u>Compound distributions (CD)</u>

for which :  $\langle N \rangle = \langle M \rangle \langle n \rangle$ ,  $Var(N) = \langle M \rangle Var(n) + Var(M) \langle n \rangle^2$ 

(#) Note that NBD is compound Poisson distribution with the number of clusters given by a Poissonian distribution and the particle inside the clusters distributed according to a logarithmic distribution.

### <u>CD in action</u>:

The immediate consequence of relation H(z)=F[G(z)] is that, in the case where f(M) is a Poisson distribution with generating function  $F(z) = \exp[\lambda(z-1)]$  then, for any other distribution g(n) with generating function G(z), the combinants obtained from the compound distribution  $h(N) = f \otimes g$ , (and calculated using generating function approach) are not oscillating and are

equal to 
$$C_j = \frac{\lambda(j+1)}{\langle N \rangle} \mathbf{g}(j+1).$$

This explains why  $C_j$  from the NBD (which is compound distribution of Poisson with logarithmic) are not oscillating and why  $C_j$  from any compound distribution based on NBD will not oscilate as well.

The choice of a BD as the basis of the used CD is therefore crucial to obtain the oscillatory behavior of  $C_{j}$ .

Note: This result explains also the apparent success of the multi-NBD type of P (N) in fitting data on the C<sub>i</sub>.

(\*) This happens because the sum of the NBD, with weights given by the BD and with the respectively chosen values of k and  $\langle N \rangle$  for each component, gives the same P(N) as the compound distribution of the NB-NBD type (i.e., based on the NB).

(\*) This is because the sum of M variables, each from the NBD characterized by parameters (p, k), is described by a NBD characterized by (p, Mk). In the case where  $M = 1, \ldots, K$  is distributed according to a BD, we have a K-component NBD  $P(N) = \sum_{M=0}^{K} P_{BD}(M)P_{NBD}(N;p,Mk)$ , (where consecutive NBD have precisely defined parameters k), which leads in a natural way to the appearance of oscillations.

(\*) However, folding several NBD is not exactly the same as compound distribution of of the type (BD and NBD) in which one can have situation with naught NBD present. It means than **OLNY COMPOUND BD** can reproduce the enhanced void probability.

### <u>CD in action</u>: example of Compound Binomial Distributions (CBD = BD and Poisson)

Poisson: 
$$G(z) = \exp[\lambda(z-1)]$$
 ( $C_0 = 2, C_{j>0} = 0$ )

BD: 
$$F(z) = (pz + 1 - p)^{K}$$

CBD: 
$$H(z) = \{p \exp[\lambda(z-1)] + 1 - p\}^{K}$$



<u>Note</u>: the choice of a BD as the basis of the used CD is crucial to obtain the oscillatory behavior of  $C_j$ .

<u>For example</u>: a compound distribution formed from a NBD and some other NBD provides smooth  $C_j$ .



<u>CD in action</u>: example of **3-component CBD (BD&Poisson)** 

$$P(N) = \sum_{i=1,2,3} w_i h(N; p_i, K_i, \lambda_i);$$

$$\sum_{i=1,2,3} w_i = 1$$



Parameters:  $w_1 = 0.34$ ,  $w_2 = 0.4$ ,  $w_3 = 0.26$ ;  $p_1 = 0.22$ ,  $p_2 = 0.22$ ,  $p_3 = 0.12$ ;  $K_1 = 10$ ,  $K_2 = 12$ ,  $K_3 = 30$ ;  $\Lambda_1 = 4$ ,  $\Lambda_2 = 9$ ,  $\Lambda_3 = 14$  <u>CD in action</u>: example of **3-component CBD (BD&Poisson)** 

$$P(N) = \sum_{i=1,2,3} w_i h(N; p_i, K_i, \lambda_i);$$

$$\sum_{i=1,2,3} w_i = 1$$



This time the fit to P(N) is quite good and the modified combinants  $C_j$  follow an oscillatory pattern as far as the period of the oscillations is concerned, albeit their amplitudes still decay too slowly. <u>CD in action</u>: example of 2-component CBD (BD&NBD)

$$H(z) = \left[ p \left( \frac{1 - p'}{1 - p'z} \right)^k + 1 - p \right]^k, \quad p' = \frac{m}{m + k}$$



<u>CD in action</u>: example of 2-component CBD (BD&NBD)



This improves substantially behaviour of  $C_j$ .  $\implies$  One has to use multicomponent CD based on BD (responsible for oscillations) and some other distribution providing damping of oscillations for large N.

### <u>CD in action</u>: example of 2-component CBD (BD&NBD)



Note that the enhancement P(0) > P(1) is also reproduced in this approach.

Recent ALICE data J.Adam et al. EPJC 77(2016)852 (NSD events, 7 TeV, 3 rapidity windows);



- (\*) Increase of the period of oscillations and their amplitude with the width of rapidity windows .
- (\*) The previous fading down of amplitude is replaced by the (almost) constant behavior or dramatic increase.
- (\*)  $\langle N \rangle \sim \Delta \eta$ , one  $\Rightarrow$  at least part of this increase comes from the increase of  $\langle N \rangle$  with  $\Delta \eta$ .
- (\*) This can be only partially true.

# To summarize:

- (\*) We argue that only compound distributions based on the BD (like) and the NBD (like) components can fit adequately observed oscillations of modified combinants C<sub>i</sub> and enhaced void probability.
- (\*) The question of which particular theoretical mechanism is in work remains, however, (at the moment...) still open.
- (\*) Hint (?):
  - NBD belongs to the class of the so-called infinite divisible distributions, BD does not.
  - In literature one finds that " combinants of all ranks are all non-negative if and only if the probability distribution is infinitely divisible"

→ modified combinants should share this property.

[However, for a uniform distribution in the interval (0, K), which is not infinitely divisible, one has  $C_i = 2/(K + 1)....(?)$ ]

<u>Comparison of  $p\overline{p}$  results and the *e+e-* annihilations case<sup>(#)</sup>:</u>



- (\*) A large number of papers suggest some kind of universality in the mechanisms of hadro production in e+e- annihilations and in pp and  $\underline{pp}$  collisions.
- (\*) However, the modified combinant analysis reveals differences between these processes. Namely: in e+e- annihilations we observe oscillations of C<sub>j</sub> with period 2, in pp and pp<sup>-</sup> collisions the period of oscillation is ~ 10 times longer and the amplitude of oscillations in both types of processes differs drastically.

# Thank you for your attention

# Дякуємо Вам за увагу

# Supplementary material

### **Estimations of errors -1:**

- (\*) A detailed discussion of the sensitivity of the modified combinants C<sub>j</sub> to the measurement uncertainties is given in Zborovsky, EPJC78(2018)816.
- (\*) In our case one observes that statistical errors cause only some chaotic spread of the measured C<sub>j</sub> but do not result in periodic oscillations. In the case of monotonic behavior of C<sub>j</sub> as function of the rank j (for example, when one uses the NBD) one gets no oscillations from errors.
- (\*) However, in the case when one observes oscillations, systematic errors can blur the whole picture of oscillation (making them invisible). The most important point is that the oscillations of C<sub>j</sub> are highly correlated (they are not chaotically scattered). Statistical errors do not give such oscillations.

### **Estimations of errors -2a:**



(a) Monte Carlo evaluated coefficients  $C_j$  emerging from NBD with parameters:  $\langle N \rangle = 25.5$  and k = 1.45. With increasing statistics points are merging to a continuous line.

### **Estimations of errors -2b:**



(b) Errors of <N>C<sub>j</sub> evaluated using the systematic and statistical uncertainties of P(N) given by ALICE.

**Estimations of errors -2c:** 



(c) Monte Carlo evaluated coefficients  $<N>C_j$  emerging from the systematic and statistical errors of P(N). The curve presented here denotes the fit to the original coefficients  $C_j$  obtained from the measured P(N), it is not the fit to the points shown.

### **Estimations of errors -2d:**



(d) For the same data as before the errors were evaluated assuming only statistical uncertainties of the measured P(N) with a poissonian distribution of events in each bin, i.e.,  $Var[P(N)] = P(N)/N_{stat}$ . Note that in this case statistical errors do not give any noticeable errors of C<sub>i</sub>.

### **Estimations of errors -2e:**



(e) Monte Carlo evaluated coefficients <N>C<sub>j</sub> with only statistical errors of P(N) accounted for. The continuous curve represents the fit to the original coefficients C<sub>j</sub> obtained from the measured P(N). **Estimations of errors -2f:** 



(f) The modified combinants C<sub>j</sub> emerging from the ALICE data on P(N) (continuous curve) in envelope corresponding to the systematic uncertainties of data, P(N) +/- Δ[P(N)].

### **Estimations of errors -3:**



#### **Estimations of errors -4:**

Estimation of the statistical significance of the oscillating behavior of the  $C_j$  using the periodogram-based Fisher g-statistic test (#) which determines whether a peak in the periodogram is significant or not. It proceeds as follows: given a series y(j) = <N>C of length L, the periodogram I( $\omega$ ) is first computed as

 $I(\omega) = \frac{1}{L} \left| \sum_{j=1}^{L} y(j) \exp(-i\omega j) \right|^2, \quad \omega \in [0, \pi].$  levaluated at the discrete normalized

frequencies  $\omega_l = 2\pi l/L$ , l=0,1,...,a = [(L - 1)/2], [x] denotes the integer part of x. If a series has a significant sinusoidal component with frequency  $\omega_k$ , then the periodogram will exhibit a peak at that frequency.

Fisher derived an exact test of the significance of the spectral peak by introducing the

Fisher g-statistic  $g = \frac{\max I(\omega_l)}{\sum_{l=1}^{a} I((\omega_l))}$ . One is testing the null hypothesis, H<sub>0</sub>, that the spectral peak is statistically insignificant against the alternative hypothesis, H<sub>1</sub>, that there is a periodic component in the signal y(j). Under the Gaussian noise assumption, the exact distribution of the g-statistic under the null hypothesis H<sub>0</sub> is given by

$$P(g^{\star} > g) = \sum_{k=1}^{[1/g]} (-1)^{k-1} \frac{a!}{k!(a-k)!} (1-kg)^{a-1}.$$

<sup>(#)</sup> R.A. Fisher, Proc. R. Soc. A 125, 54 (1929); P.J. Brockwell and R.A. Davis, Time Series: Theory and Methods (2nd Ed.), (Springer Verlag, 1991), Chapter 10.

### **Estimations of errors -5:**



Note the large observed value of g, a peak in the periodogram, which indicates the existence of a strong periodic component and leads us to reject the null hypothesis. The probability that the spectral peak is statistically insignificant is 10<sup>-16</sup>.

Normalized periodogram I( $\omega$ ) for <N>C<sub>j</sub> calculated from ALICE data for pp at 7 TeV and  $|\eta|$  < 3.



It means that the probability to observe oscillations as a result of errors is extremaly small. Therefore, we conclude, that they show enough power to disclose the fine details of experimentally measured multiplicity distributions, and can shed new light on the dynamics of multiparticle production processes.

# Example of the e+e- annihilations case(\*):



Data on P(N) measured in e+e– collisions by the ALEPH experiment at 91 GeV; the modified combinants C<sub>j</sub> deduced from these data on P(N) and the mean value  $\langle C_j/\xi_j \rangle$  (averaging is performed over available ranks j) for even and odd ranks j evaluated from P(N) in different rapidity windows for  $\xi = 6.85$ , 2.35, 1.2 and 0.91 for |y| < 2.0, 1.5, 1.0 and 0.5, respectively.

#### Note that:

- (\*) A large number of papers suggest some kind of universality in the mechanisms of hadro production in e+e- annihilations and in pp and pp<sup>-</sup> collisions.
- (\*) However, the modified combinant analysis reveals differences between these processes. Namely: in e+e- annihilations we observe oscillations of C<sub>j</sub> with period 2, in pp and pp<sup>-</sup> collisions the period of oscillation is ~ 10 times longer and the amplitude of oscillations in both types of processes differs drastically.

H.W.Ang, A.H.Chan, M. Ghaffar, Q.Leong, M.Rybczyński, G.Wilk, Z.Włodarczyk, arXiv.1812.08840.