FROM PHOTONS TO HADRONS TO GALAXIES — HOW TO ANALYZE THE TEXTURE OF MATTER DISTRIBUTIONS*

P. CARRUTHERS

Department of Physics, University of Arizona
Tucson, AZ 85721 USA

(Received September 30, 1991)

We discuss the phenomenology of point matter distributions, using examples of photoelectron count arrival times, galaxy distributions and multihadron production at high energies. Count distributions in phase space cells, multiplicity moments and correlation functions are discussed along with their interconnections. The possible description of higher order cumulant correlations (both for hadrons and galaxies) by linked two-particle cumulants and negative binomial coefficients, is reviewed.

PACS numbers: 06.30.-k

1. Introduction

The basic problem in assessing the structure of and behavior of a system composed of a large number of "points" is how best to characterize its texture. (Note that from our vantage the "points" might represent trees, galaxies, photons, bacteria, hadrons or whatever.) Simply presenting a catalogue listing all the phase space positions is extremely useful, but does not give data in a form which the human mind can easily understand. The primary physical question is one of *correlations*, *i.e.*, given that one object is at a certain place, how is the likelihood affected that other objects will be at other places? Secondly, how do these correlations evolve dynamically as "time" progresses. It might be thought that such a traditional topic, at the heart of intense research activity in all branches of science, would be totally developed. That this is not the case is even true for "geometric" phenomenology, as is proven by the recent appreciation of the widespread

* Presented at the XXXI Cracow School of Theoretical Physics, Zakopane, Poland, June 4-14, 1991.

occurrence of fractal structures. In fact the techniques used in these lectures represent some combination of probability theory and fractal engineering. Some of the recommended data analysis procedures that emerge from this point of view are of general use whether or not scaling fractals are involved, although the techniques have been tuned to that case [1].

As everyone knows, concepts of probability play a crucial role in science, from the interpretation of quantum mechanics to the estimate of errors in data analysis. The development of statistical physics as a calculus of probabilities beginning with Maxwell, Boltzmann and Gibbs, not to mention Einstein and many others, represents one of the major achievements of science. A well formulated probabilistic law has the same predictive power as a symmetry group transformation law. Despite all this, accomplished theorists willing to learn the most erudite mathematical techniques remain ignorant not only of the techniques, but more importantly the spirit and intent of, probabilistic reasoning. As a consequence, the conceptual evolution of mainstream physics has developed a blind spot which in most cases is the ghost of classical determinism. Now that some basic issues of nonlinear science have been clarified, especially the sensitivity to initial conditions, no informed person can adhere to these old views. Still, their pernicious influence is very strong. It is commonly believed that an equation of motion for some set of fields is a finer construct than the discovery of a statistical law. Surely the Maxwell–Boltzmann distribution is as great a discovery as the equations of QCD.

Having expressed these philosophical opinions, let me say that in my view our job as physicists is to explain data. In order to avoid the vagueness that follows from excessive generality, we focus on three types of experiments to bring out some key issues.

Photoelectron Counts. Here we imagine a light source striking a detector (or detectors), producing well-resolved photoelectron counts (Fig. 1(a)). The x-marks on the time axis indicate the idealized times when the photon energy momentum excites the electron in the detector. Notice that Fig. 1(a) represents a particular example of the experiment, lasting for a time T. In order to understand the physics properly, we need to collect many such examples under the same conditions. The appearance of the spike distribution will differ from event to event.

Galaxy Distributions [2]. When looking at the night sky we see a highly irregular distribution of luminous matter. Disregarding all details, we shall speak of "the galaxy distribution." What we see can be regarded as projected onto the surface of a sphere, located by the usual angular coordinates (θ, ϕ). More detailed three-dimensional data are currently becoming available, largely due to red-shift measurements (see Fig. 1(b).)

Following fifty years of belief in a homogeneous, isotropic universe, as-

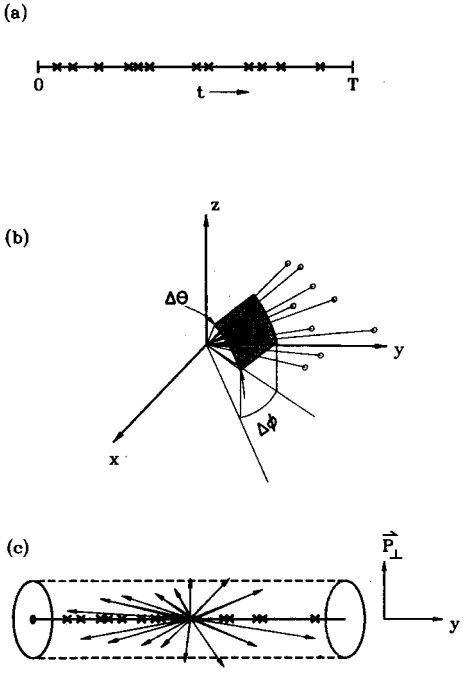

Fig. 1. In (a) the arrival times of photocounts are shown for one experiment in which the detector is on for the interval $0 \leq t \leq T$. (b) indicates the projection of galaxies in three dimensions (open circles) onto the sphere (θ, ϕ), solid circles. (c) shows a typical multiparticle production event at high energy in the c.m. frame. The vectors indicate that the transverse momentum $p_\perp$, is typically (*i.e.*, except for scarce jet events) limited to a tube of radius ~400 MeV/C, while the longitudinal momentum is parametrized by the rapidity $y = 1/2ln(E + p_z/E - p_z)$.

tronomers have suddenly begun to scrutinize every imperfection in the matter distribution. Despite the complications involved (*e.g.*, dark matter) it is believed that studies of these irregularities will illuminate the dynamics and evolution of the universe. I agree with this optimism.

Note that we have only one universe available for inspection, unlike the photocount experiment. However, since the size of the universe is so large, it is useful to regard carefully chosen subspaces as members of an ensemble of statistically independent entities. But care is needed since correlated structures of enormous extent have been discovered. New ideas about the best way to deal with this situation are currently in a state of intense development. In fact this nontrivial complication shows up in many branches of science.

Multihadron Momentum Distributions. Consider the production of hadrons (usually dominated by pions) at energies so large that the fate of the projectile and target are of small consequence. In typical events a large

fraction of the initial kinetic energy is transformed — in a small space-time region — to final state hadrons, plus some secondary leptons and photons not of interest for this discussion. Except for scarce jet events the momentum of secondaries transverse to the collision axis is small — one speaks of longitudinal phase space (Fig. 1(c)) and focuses on the dependence of the final particles on their rapidity $y = 1/2 \ln(E + p_z)/(E - p_z)$. Usually one integrates over the $p_\perp$ dependence of the inclusive differential cross sections, or simply ignores $p_\perp$ altogether. The locations of the individual rapidities of a particular event are strewn along the kinematically allowed rapidity axis in the c.m. frame

$$-Y_M \leq y \leq Y_M, \tag{1.1}$$

$$Y_M = \tfrac{1}{2} \ln \frac{s}{4m_\pi^2}, \tag{1.2}$$

where the maximum allowed center of mass rapidity of the typical particle (pion) is given by Eq. (1.2). Typically data are presented in histogram form (Fig. 2) since the experimental resolution may not resolve the exact location of the particles in a bin. In that case one simply counts the particles in the bin. As we shall explain, it is also informative to study the irregularities of the histograms as a function of bin resolution. Current experiments have 10^4-10^5 events and resolution going to a rapidity bin size of around 0.1.

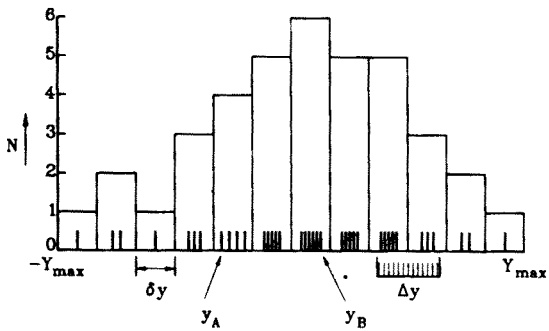

Fig. 2. A typical rapidity histogram is shown. For a given event we imagine n particles to be located at the points denoted by short vertical segments of equal length. Since the resolution is δy, the height N counts the number of particles in a given bin. Collecting many events, we can for example find the count distribution in the interval Δy, find the correlation between y_A and y_B, *etc.*

From the foregoing discussion it is clear that fluctuations occur in the number of counts in the individual events: the number of photons arriving in time T; the number of galaxies on a given photographic plate; the number of charged hadrons in a particular part of momentum phase space (here usually

a given rapidity interval ΔY). The probability P_n of the count distribution can provide crucial information about the nature of the physical process under consideration, as we shall explain.

A related measure of fluctuations is given by the number moments in bins i of the phase space $\langle n_i \rangle$, $\langle n_i(n_i-1) \rangle$, $\langle n_i(n_i-1)(n_i-2) \rangle$, ... Here we are introducing the factorial moments $\langle n^{[p]} \rangle = \langle n(n-1)\dots(n-p+1) \rangle$ as explained below. The value of this approach for multihadron production has been developed by Bialas and Peschanski [3], who stressed the importance of measuring the dependence of the factorial moments on resolution (bin size). One can also evaluate correlations between bins, *e.g.*, $\langle n_i n_j \rangle$, $i \neq j$.

In fact all the aforementioned measures of fluctuations can be computed from the density correlation functions $\rho_1(x)$, $\rho_2(x,x')$, $\rho_3(x,x',x'')$,... should they be available. Before explaining these connections we list a variety of methods currently in use to analyze the structure of matter distributions. Here we can only describe a few of these methods.

a) *Count Probability Distributions.* In this case the probability of finding n particles in a given part of space is computed.

b) *Histogram Analysis.* Particle counts and moments in given space bins are analyzed.

c) *Correlation Functions.* Correlated counts in different bins define the correlation functions.

d) *Pair (Multiplet) Counts* [4]. In this approach the emphasis is on counting pairs, triplets, of particles closer than some variable distance ϵ. There is a close connection with the sequence of correlation dimensions for scaling fractals.

e) *Information Correlations* [5]. Beyond the usual information entropy of a probability distribution, one can define a sequence of information correlations, the second order version being the well-known mutual information.

f) *Power Spectrum* [6,7]. The nomenclature refers to the connection of the autocorrelation function of a time-dependent function $x(t)$ to the squared Fourier transform $|x(\omega)|^2$ *via* the Wiener–Khinchin theorem.

g) *Multifractals.* Since the scaling properties of fractals can be mixed in character, various methods are under development to quantify them. One such is closely related to the Renyi entropy, although in applications true probabilities are replaced by event counts [8,9].

h) *Wavelet Analysis.* Few of the measures mentioned above give adequate information about *what* is happening *where* in space. (The master set of correlation functions do so in principal, but are very hard to measure precisely.) For example, given a strange attractor (fractal set) with mixed scaling properties, we might want to also know where on the attractor certain scaling properties occur. Wavelets, analogous to Wigner transforms, allow one to localize and scale the various components of a set. Again,

practical use depends on the existence of precise data.

2. Interlude: some useful definitions from probability theory

The material in this Section is well known to those who know it well. In any case it will establish our notations for the subsequent Sections. For a more extensive account we refer to the treatise [11] of Kendall and co-workers, as well as a physics-oriented review in Ref. [12].

We refer to the basic relations connecting count probabilities, moments, and the generating functions which collect the moments into a single structure. Given a count probability distribution P_n, $n = 0, 1, 2, \ldots$ we can define $(p = 0, 1, 2, \ldots)$

ordinary moments μ_p

$$\mu_p \equiv \langle n^p \rangle \equiv \sum_{n=0}^{\infty} n^p P_n, \tag{2.1}$$

moment generating function $M(\lambda)$

$$M(\lambda) \equiv \sum e^{-\lambda n} p_n \equiv \sum_{p=0}^{\infty} \frac{(-\lambda)^p}{p!} \mu_p, \tag{2.2}$$

cumulant moment $\mathcal{K}_p$ *and generating function*

$$\ln M(\lambda) \equiv \sum_{p=1}^{\infty} \frac{(-\lambda)^p}{p!} \mathcal{K}_p, \tag{2.3}$$

$$\begin{aligned} &\mathcal{K}_1 = \langle n \rangle, \quad \mathcal{K}_2 = \langle n^2 \rangle - \langle n \rangle^2 \\ &\mathcal{K}_3 = \langle n^3 \rangle - 3\langle n^2 \rangle \langle n \rangle + 2\langle n \rangle^3, \ldots . \end{aligned} \tag{2.4}$$

For our discussion of the structure of point distributions of one species of particle it is more convenient to use a variant set of moments and cumulant moments in which ordinary powers are replaced by "factorial" powers, as indicated in the introduction. We define the factorial moment ξ_p, its generating function $Q(\lambda)$ and cumulant factorial moment f_p by

factorial moment

$$\begin{aligned} \xi_p &\equiv \langle n(n-1)\ldots(n-p+1) \rangle \equiv \langle n^{[p]} \rangle \\ &\equiv \sum_{n=0}^{\infty} n^{[p]} P_n \end{aligned} \tag{2.5}$$

factorial moment generating function

$$Q(\lambda) \equiv \sum_{n=0}^{\infty}(1-\lambda)^n P_n \equiv \sum_{p=0}^{\infty}\frac{(-\lambda)^p}{p!}\xi_p \tag{2.6}$$

factorial cumulant moment generating function

$$\ln Q(\lambda) \equiv \sum_{p=1}^{\infty}\frac{(-\lambda)^p}{p!}f_p \tag{2.7}$$

$$\begin{aligned} &f_1 = \langle n\rangle\,, \quad f_2 = \langle n(n-1)\rangle - \langle n\rangle^2 \\ &f_3 = \langle n(n-1)(n-2)\rangle - 3\langle n(n-1)\rangle\langle n\rangle + 2\langle n\rangle^3\,, \ldots\,. \end{aligned} \tag{2.8}$$

Each set of moments, generating functions and cumulant moments has its own merits and most naturally describes certain distributions. Consider, for example the Poisson distribution and its predictions for Eqs (2.5)–(2.8):

$$\begin{aligned} &P_n = \bar{n}^n\frac{\exp(-\bar{n})}{n!}, \\ &Q(\lambda) = \exp(-\lambda\bar{n}), \\ &\ln Q(\lambda) = -\lambda\bar{n}, \\ &\xi_p \equiv \langle n^{[p]}\rangle = \bar{n}^p, \\ &f_p = \bar{n}\delta_{p1}\,. \end{aligned} \tag{2.9}$$

Note that $\xi_p/\bar{n}_p = 1$ and that all cumulant factorial moments vanish beyond the first. The Poisson distribution describes a minimal fluctuation structure.

For purposes of data analysis it is useful to employ both the normalized factorial moments F_p, which are easily measured:

$$\begin{aligned} &F_1 = 1\,, \quad F_2 = \frac{\langle n(n-1)\rangle}{\langle n\rangle^2}\,, \\ &F_3 = \frac{\langle n(n-1)(n-2)\rangle}{\langle n\rangle^3}\,, \ \ldots \end{aligned} \tag{2.10}$$

as well as the normalized cumulant factorial moments $K_p = f_p/\langle n\rangle^p$:

$$\begin{aligned} &K_1 = 1\,, \quad K_2 = \frac{\langle n(n-1)\rangle - \langle n\rangle^2}{\langle n\rangle^2}, \\ &K_3 = \frac{\langle n(n-1)(n-2)\rangle - 3\langle n(n-1)\rangle\langle n\rangle + 2\langle n\rangle^3}{\langle n\rangle^3}\,. \end{aligned} \tag{2.11}$$

Both the F_p and K_p moments will be shown to be averages over related correlation functions, with the K_p allowing the expression of the easily measured F_p moments in terms of mostly lower order K_p moments:

$$\begin{aligned} F_1 &= K_1\,,\ F_2 = 1 + K_2\,,\\ F_3 &= 1 + 3K_2 + K_3\,,\\ F_4 &= 1 + 6K_2 + 3K_2^2 + 4K_3 + K_4\,,\\ F_5 &= 1 + 10K_2 + 15K_2^2 + 10K_2K_3 + 10K_3 + 5K_4 + K_5\,. \end{aligned} \tag{2.12}$$

The identities (2.12) are in a sense akin to a phase shift decomposition of a scattering amplitude. These equations apply to the particular phase space domain for which the probabilities P_n are defined. We shall explain how these relations and their generalizations, averaged over larger parts of phase space, are useful in studying fluctuations.

Besides the Poisson distribution, the Negative Binomial Distribution (NBD) is of major significance in describing count probabilities in arbitrary partitions of phase space for photon, galaxy and hadron count distributions. Known in mathematical statistics for some time, this distribution appeared in physics in 1924 when Planck [13] considered the distribution of n Bose–-Einstein (B-E) particles in k cells of average equal occupancy $\bar{n}/k$. The geometric B-E distribution $P_n = \bar{n}^n/(1+\bar{n})^{n+1}$ has generating function $Q(\lambda) = (1 + \bar{n}\lambda)^{-1}$ so that for the k cells ($n = \sum_{i=1}^{k} n_i$)

$$Q_k(\lambda) = \left(1 + \frac{\lambda\bar{n}}{k}\right)^{-k}\,. \tag{2.13}$$

The corresponding NBD count distribution is

$$P_n^k = \frac{(n+k-1)!}{n!(k-1)!}\,\frac{(\bar{n}/k)^n}{(1+\bar{n}/k)^{n+k}}\,. \tag{2.14}$$

In modern times (2.14) was derived by Mandel [14] for semiclassical photocount distributions for k sources of gaussian random fields. In the early 70's Giovannini [15] showed its success in describing hadronic multiplicity data. In 1983 we [16] rediscovered this distribution after the publication of the UA5 data. This group subsequently presented an impressive fit of hadronic data as a function of energy, in which the k parameter decreases with energy [17]. The meaning of, and the dependence of this parameter on energy and rapidity interval remains somewhat obscure.

It is easy to derive from (2.14) the *scaling form*

$$\bar{n}P_n^k \cong \psi_k(n/\bar{n}) \quad n, \bar{n} \to \infty\,, \tag{2.15}$$

$$\psi_k(x) = \frac{k^k}{(k-1)!} x^{k-1} e^{-kx}, \quad x = \frac{n}{\bar{n}}. \tag{2.16}$$

In fact ψ_k, a special case of the gamma-distribution, is a good approximation for moderately large $n, \bar{n}$.

Eq. (2.15) is a particular expression of Koba–Nielsen–Olesen (KNO) scaling [18]. According to UA5, this scaling is broken (but in a simple manner) *via* the energy dependence of the k parameter.

For subsequent use, we derive the factorial cumulant moments for the NBD. Comparing Eqs (2.7) and (2.13) we find $f_p = \bar{n}^p (p-1)! \,/\, k^{p-1}$, *i.e.* $K_p = (p-1)! \,/\, k^{p-1}$. Hence the ordinary factorial moments are given by [19]

$$\begin{aligned} F_2 &= 1 + 1/k, \\ F_3 &= 1 + 3/k + 2/k^2, \\ F_4 &= 1 + 11/k^2 + 6/k^3, \ldots \end{aligned} \tag{2.17}$$

i.e. as a power series in $1/k$. In application, we will consider these formulas in terms of their dependence on the considered volume, in which $1/k$ is the average of the two- particle cumulant correlation function defined below.

3. Counts and their fluctuations

In the simplest type of measurement we simply count the number of objects appearing in a given fixed volume (*e.g.*, galaxies), time interval (*e.g.*, photons) or rapidity interval (hadrons). By accumulating many such examples we can find the frequency of occurrence of a particular number of objects. The form of the probability distribution P_n summarizes the fluctuation in the count (n) observable. (Of course there can be subtleties in the preparation of the ensemble; in the case of galaxies we have to sample different volumes of the same universe, a more subtle process than allowing the same accelerator to produce statistically independent collisions under basically identical conditions.) Note that the count probability does not determine *where* the individual contributor is located in the count *cell* (cell=volume, rapidity bin or time interval).

First consider how the nature of the fluctuating electromagnetic field determines the photocount distribution. Consider highly coherent laser light scattered by a fluid at its critical point, at which the density undergoes strong fluctuations. The scattering process changes the form of the photocount distribution P_n from Poisson to a much broader distribution, [20] closely approximated by the negative binomial. Similar results were found in the early days of laser physics. Below the threshold for coherent light emission, the single mode noisy photocount distribution is well described

by a Bose–Einstein (NBD with $k = 1$) distribution, while the Poissonian appears in the coherent regime.

A simple model — the coherent states [21] of a simple harmonic oscillator — provides a clear illustration of the dependence of the count distribution on the density matrix of the field ensemble. The usual coherent states $|\alpha\rangle$ (α is complex)

$$|\alpha\rangle = \exp\left(-\tfrac{1}{2}|\alpha|^2\right) \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle \tag{3.1}$$

are produced by a classical external force [22] coupled linearly to the oscillator and in any case are prototype quantum states giving classical coherent motion $\langle\alpha|x(t)\alpha\rangle \propto \cos(\omega t - \phi)$. The pure density matrix ρ_α and the count probability corresponding to the coherent state are

$$\begin{aligned} \rho_\alpha &= |\alpha\rangle\langle\alpha| \\ P_n &= |\langle n|\alpha\rangle|^2 = |\alpha|^{2n} \frac{\exp\left(-|\alpha|^2\right)}{n!}, \end{aligned} \tag{3.2}$$

confirming what is clear from (3.1): $\bar{n} = |\alpha|^2$ for a Poisson distribution. Next consider a Gaussian probability in the complex α plane, with density matrix

$$\rho = \int d^2\alpha \frac{\exp(-|\alpha|^2/N)}{\pi N} \rho^\alpha \tag{3.3}$$

and count distribution

$$P_n = \langle n|\rho|n\rangle = \frac{N^n}{(1+N)^{n+1}}. \tag{3.4}$$

The Gaussian random oscillator variable α is the prototype of noisy and thermal field ensembles.

Eq. (3.3) is a special case of the "P-representation" of the (single-mode) density matrix [21]

$$\rho = \int d^2\alpha P(\alpha) |\alpha\rangle\langle\alpha| \,. \tag{3.5}$$

This representation is of rather general validity; in quantum domains one has to realize that P is not necessarily positive. Note that (integrating out the phase angle), Eq. (3.5) leads to the useful Poisson transform

$$P_n = \int_0^\infty dx f(x) \frac{(x\bar{n})^n \exp(-x\bar{n})}{n!}. \tag{3.6}$$

This expression, and its generalization to higher dimensions in $\vec{n}$ (and $\vec{x}$) is extremely useful. We note that for large n and $\bar{n}$, $\bar{n}p_n \sim f(n/\bar{n})$ is a useful scaling form. The reader can confirm that for the choice $f(x) = \psi_k(x)$ of Eq. (2.16), the NBD (2.14) is recovered. One can also confirm this by compounding k independent oscillators with gaussian weights corresponding to equal average occupancy N/k (*i.e.*,

$$\rho(\alpha_1 \cdots \alpha_2) = \frac{\exp\left(-\sum_{i=1}^{k} |\alpha_i|^2/(N/k)\right)}{(\pi N/k)^k}$$

subject to $n = \sum_{i=1}^{k} n_i$ one recovers (2.13) and hence (2.14).) (However in physical problems the effective oscillators are correlated and k can be nonintegral.) Note that the photocount distribution for a superposition of k independent Gaussian sources is negative binomial.

Naturally the connections of photocount statistics and their relation to the fluctuating light sources have been thoroughly worked out since the advent of the laser [20]. We must also mention the Hánbury–Brown, Twiss effect on intensity correlations. In high energy physics the analogous effect for like-sign particles goes under the name of Bose–Einstein correlations.

Next consider counts of charged hadrons for a fixed rapidity interval. Originally (*e.g.*, in a bubble chamber) one derived n-prong cross sections and hence the probabilities in full phase space. In other detectors (say UA1, UA5)) the forward and backward cones — where diffractive events live — are excluded. Most published data do not distinguish plus from minus, Ks from πs, *etc.* When angles rather than true momenta are measured the variable of choice is the pseudorapidity η rather than the rapidity y.

By now there exist data for many energies and for many choices of target and projectile. Here we only discuss multihadron production in hadron-hadron and $e^+ - e^-$ collisions at high enough energies that statistical methods make sense (say c.m. energy $W = \sqrt{s} \gtrsim 20$ GeV, with charged multiplicity $\gtrsim 10$.

For hadron-hadron collisions π^+-p, K^+-p, p-p, p-$\bar{p}$ for c.m. energies ranging up to 600 GeV, the NBD provides a remarkably good fit to the count data. The k-parameter depends on the energy and the size ΔY and location of the rapidity bin. For old timers this is known a "violation of KNO (Koba–Nielsen–Olesen) scaling" [18] which would be expressed by Eq. (2.16) for constant k. At 900 GeV the UA5 group has difficulty accommodating their data with NBD formula. Perhaps this is real, perhaps not.

The value of k decreases from order 20 to 3 as energies increase to collider values. Unfortunately there is no truly persuasive theory of the NBD, or its parameter k. Many cluster, cascade and stochastic evolutions lead to

such a result, but we are a great distance from the analogue of, say, deriving the Maxwell distribution as the stationary solution of the Boltzmann equation. Later we shall see the connection with an equally mysterious structure of the correlation functions. For the time being we accept the result as a phenomenological law.

High energy multihadron production in $e^+ - e^-$ collisions is known to be *via* jet production and decay. The multiplicity distribution is much narrower than for h-h collisions at the same energy. In fact the shapes are quite similar to almost coherent laser photocount distributions (which are very sensitive to small amounts of noise.) The deviation from Poisson is, however, worth more discussion because of the apparent validity of KNO scaling in $e^+ - e^-$ production. The pure Poisson distribution does not scale, and the agreement with scaling requires special tuning of the signal/noise ratio. For this reason, the alternative log-normal distribution $\propto \exp(-A\ln^2(n\bar{n}))$ is appealing. Actually the log-normal is close to negative binomial except for large n. The NBD is turn goes over to the Poisson for large k (this can be seen by comparing Eqs (2.9) and (2.13)).

Phenomenologically it is interesting that photo count distributions of Gaussian light sources resemble those of hadronic multiparticle production, and that photocount distributions of nearly coherent laser sources resemble hadronic counts from jet sources. Although we have argued [24] that a quark jet resembles a decelerated classical current source, which produces coherent (multiple-Poisson) counts in QED, the absence of a persuasive hadronization theory prevents taking the analogy too seriously. Even so, the mystery deepens when we consider galaxy counts.

As we all know, the luminous matter distribution in the universe, as measured by galaxy distributions is highly irregular, unlike the homogeneous models written up in textbooks. In fact it is not really clear that there is a scale beyond which uniformity exists. Primitive measures of irregularity are given by count distributions and by low order correlation functions.

In 1934 Hubble published [25] count statistics for galaxies as seen in photographic plates (*i.e.*, two dimensional projections). Instead of multinomial $\rightarrow$ Poisson (for dilute samples, he found a log-normal count P_n)

$$\psi(n/\bar{n}) \equiv \bar{n}P_n \equiv \frac{\exp\left(-(\ln(n/\bar{n}))^2/2\sigma^2\right)}{\sqrt{2\pi}\sigma}, \tag{3.7}$$

where $\sigma \simeq 0.45$ and the normalization is $\int \psi(x)dx/x = 1$.

This simple result gives the simplest measure of non-randomness and corresponds to Fig. 1(b).

Later, stimulated by our work on hadronic counts, we constructed [26]. the (conditional) count distribution based on the Zwicky catalogue [27]

of clusters. After some hand- waving in data interpretation to remove anisotropy in the count distributions we were surprised to find (Fig. 3) a good fit with the gamma distribution with $k \approx 6$. (Note that since the average cluster size is fixed, we cannot vary the cell size, as in the Hubble case.) Yet the general shape of the curves are quite similar. Another distribution has been derived from thermodynamic considerations. [28] Although it is not very different from the NBD/Γ distribution in appearance, it seems less good when confronted by the void probability data.

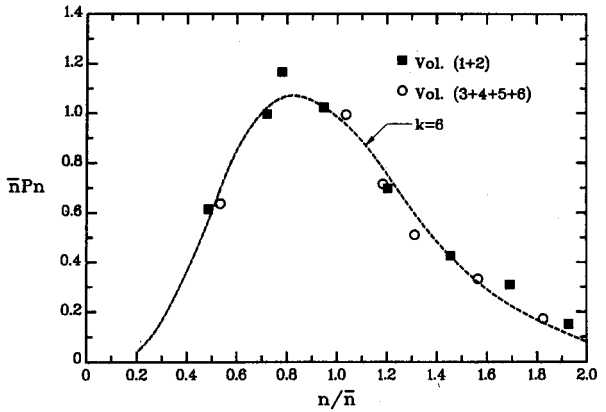

Fig. 3. The probability of finding n galaxies in a Zwicky cluster is given in scaled form. The average number of galaxies is $\bar{n} \simeq 97$. The solid squares denote angular regions of the heaven found in volumes 1 and 2 of the catalogue; the open circles volumes 3 to 6. The ambiguities for small $N \lesssim 20$ account for the absence of points. The dashed curve is the gamma distribution for $k = 6$: $\psi_6 \propto x^5 \exp(-6x)$.

From Eq. (2.6) we note that $Q(1) = P_0$. From (2.7) and (2.11) we find

$$P_0(v) = \exp\left(\sum_{p=1}^{\infty} \frac{(-\bar{n})^p}{p!} K_p\right), \tag{3.8}$$

where $\bar{n}$ and K_p also refer to the void volume v. Clearly P_0 depends on moments (fluctuations) of all orders. In the case of the negative binomial $K_p = (p-1)!/k^{p-1}$ giving agreement with (2.14) as it must. This can be rewritten as a scaling law

$$-\frac{\ln P_0}{\bar{n}} = \frac{\ln(1+\bar{n}/k)}{\bar{n}k}. \tag{3.9}$$

This NBD prediction has been nicely confirmed by Fry *et al.* [29] It also has been confirmed for multihadron rapidity-gap distributions [30].

Having begun to look around, we have found negative binomial count distributions everywhere [12]. Even the Poissonian and near-Poissonians, as well as log-normals are limits of or closely related to the NBD. Since the physical details and circumstances of these systems are so distinct, any search for a universal explanation would seem to depend on a statistical assumption.

Without looking too closely into details, we can notice similarities. In the cases we have discussed, the basic force laws are scale invariant (QED, QCD and gravity) to first approximation. There are many active degrees of freedom; the systems are driven (by lasers, collisions or big bangs) and dissipative. Cascading or branching dynamics seems appropriate. (Note that fluid turbulence shares these features, although we shall not discuss that problem here.)

Unless these similarities are purely accidental, there ought to be a way to analyze them all in a unified way. In order to make this point of view more persuasive we go on to consider the behavior of moments and correlation functions.

4. Histograms, multiplicity moments and correlations

In order to better analyze the texture of count distributions it is useful to subdivide the phase space into pieces of variable size. In addition to assessing the size dependence we can learn about correlations through the mutual dependence of counts in various phase space cells.

In Fig. 2 we exhibit the main features of this analysis, for one dimensional bins in time or rapidity y. The number of individual occurrences in each bin are counted and give rise to the bar graph (histogram). By assembling many equivalent examples for differing δy we can learn about many interesting features of the data. These include:

1. Correlations: If a particle is known to be at y_A, how likely is another at y_B (and $y_C, y_D, \ldots$)?
2. Counts: For an arbitrarily centered interval of arbitrary size Δy, what is the count probability?
3. Fluctuations: These are measured by the variation of number of produced particles as well as their distribution y_i.

When the bin size is large enough so that several or many particles are counted, the irregularity of the point distribution is smoothed out. At first, the appearance of the histogram becomes more irregular as the resolution is increased, as shown in Fig. 4(a)-(b). However if we keep decreasing δy, eventually every bin is empty or has one particle (Fig. 4(c)). This elementary fact presents a number of subtle data analysis problems [31] for the modest number of particles characterizing many experiments of current interest.

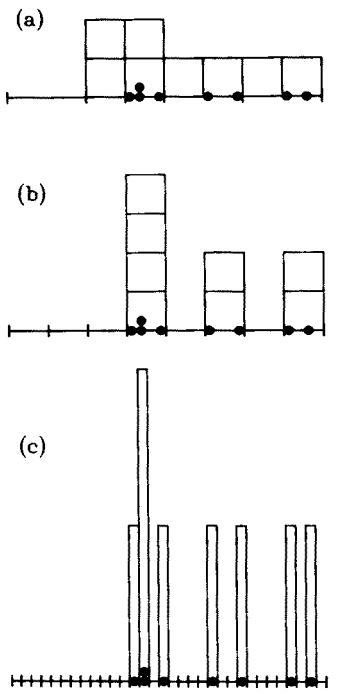

Fig. 4. Histograms are shown for bin widths $8\delta y$, $4\delta y$, δy ((a),(b),(c) respectively). For sufficiently small bin size the finite multiplicity event (here eight particles) will be resolved into vertical bars of equal height. At first the number fluctuations increase as resolution improves, but eventually the irregularities are measured by the point distribution, which is more fundamental.

Returning to Eq. (2.10), let us consider the bin-averaged factorial moments of Bialas and Peschanski [3]. Here a large interval ΔY is broken into M equal segments δy with

$$F_2 = \frac{1}{M}\sum_{i=1}^{M} F_2^i(\delta y)\,, \tag{4.1}$$

$$F_3 = \frac{1}{M}\sum_{i=1}^{M} F_3^i(\delta y)\,, \tag{4.2}$$

etc. (The individual bin moments can be normalized locally or to the overall average.) The averaging process improves statistics but can be misleading if the bin-dependence is strong.

As mentioned in Section 2, the deviation of the F_p^i from unity signifies non-Poisson Statistics in bin i. Fig. 4(a), (b) show intuitively how the factorial moments should increase as δy decreases if fluctuations are nontrivial,

until the mean bin population drops to order unity (Fig. 4). This example has eight particles distributed among eight bins:

$$\text{a)} \quad \sum_i n_i(n_i - 1) = 2 \times 1 + 2 \times 1 + 1 \times 0 + 1 \times 0 = 4\,,$$

$$\text{b)} \quad = 4 \times 3 + 2 \times 1 + 2 \times 1 = 16\,,$$

$$\text{c)} \quad = 2 \times 1 = 2\,.$$

Higher factorial moments are dependent on large bin population for their survival. As $\delta y \to 0$, $n_i = 0, 1$ and no factorial moment survives. Methods to correct for this effect have been developed by Lipa *et al.* [31].

The quantitative dependence of the $F_p(\delta y)$ moments on bin size δy is shown in Fig. 5 for the UA1 experiment (640 GeV c.m. energy $p\bar{p}$ collisions). We note that the F_p increase with decreasing δy and increase for fixed δy with moment order. (Not discussed here is the increase of the F_p with c.m. collision energy for fixed δy, or the decrease with increasing complexity of the collision, due to the introduction of nuclei as projectiles or targets.)

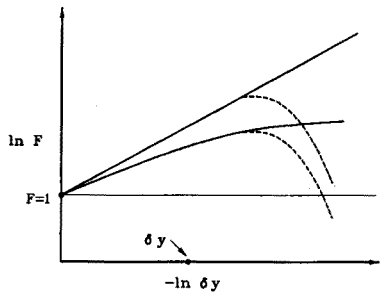

Fig. 5. Schematic representations of possible factorial moment dependence on bin resolution δy is shown. The horizontal line $F = 1$ corresponds to the absence of correlations (Poisson count statistics). The sloped straight line corresponds to scaling $F(\delta y) \propto (\delta y)^{-\nu}$ and the curved line to "saturation" expected for a finite correlation length. The dashed segments indicate the falling off of the F-moments due to the population per bin dropping to one or zero for small δy (see Fig. 4).

As we have stressed elsewhere, [32,33] the behavior of the moments is rigorously determined by the correlation functions. The inverse is not so strict: good data [34,35] on moments exists to $5th$ order even though correlation functions are not known beyond third order in hadronic data. Nevertheless such information gives for the first time important clues about higher correlations. The question of the behavior of the factorial moments is the question of explaining the sequence of rapidity correlation functions.

We illustrate this connection in the case of the two-particle correlation-density $\hat{\rho}_2$ and its associated moment F_2. Taking one step backwards we

introduce the single particle density $\hat{\rho}_1$ and two particle density $\hat{\rho}_2$ for a given event with n particles located at positions s_i.

$$\hat{\rho}(y;\vec{s}) = \sum_{i=1}^{n} \delta(y - s_i)\,, \tag{4.3}$$

$$\rho_2(y_1, y_2; \vec{s}) = \sum_{i \neq j=1}^{n} \delta(y_1 - s_i)\delta(y_2 - s_j)\,. \tag{4.4}$$

Averaging these densities over events [36] we arrive at the one and two-particle inclusive differential cross sections:

$$\frac{1}{\sigma}\frac{d\sigma}{dy}, \quad \frac{1}{\sigma}\frac{d^2\sigma}{dy_1 dy_2}, \quad \cdots$$

In order to see the connection between Eq. (4.4) and Eqs (2.10) and (4.1) we integrate y_1 and y_2 over identical ranges in bin i, obtaining

$$\langle n_i(n_i - 1)\rangle = \int_{\Omega_i} dy_1 dy_2 \langle \hat{\rho}_2(y_1, y_2)\rangle\,. \tag{4.5}$$

If we normalize locally we find that the second of Eqs (2.12) also holds for bin averaged moments (note $\langle n_i\rangle \equiv \rho_i \delta y$)

$$F_2(\delta y) = 1 + K_2(\delta y)\,, \tag{4.6}$$

$$K_2^i(\delta y) = \int_{\Omega_i} \frac{\rho_2(y_1, y_2) - \rho_1(y_1)\rho_1(y_2)}{(\rho_i)^2} \frac{dy_1 dy_2}{(\delta y)^2}\,. \tag{4.7}$$

The structure $\rho_2 - \rho_1\rho_1$ is the second order cumulant correlation function; its vanishing signals the absence of correlation, in this order, and to Poissonian counting statistics. Experimentalists frequently present data for the ratio

$$k_2(y_1, y_2) \equiv \frac{\rho_2(y_1, y_2)}{\rho_1(y_1)\rho_1(y_2)} - 1 \tag{4.8}$$

since many detection errors cancel in the ratio. For small bin size (4.8) equals the integrand of Eq. (4.7), so that $F_2(\delta y)$ is predicted once k_2 (often called R in the experimental literature) is known.

Although knowledge of k_2 determines F_2, the converse is not true. It is possible to get good agreement [37,38] with F_2 using translation — invariant

correlation functions that are only qualitatively correct at relatively low energy.

In order to illustrate several points with an analytic model consider the popular exponential fit to Eq. (4.8)

$$k_2(y_1, y_2) = \gamma \exp\left(\frac{-|y_1 - y_2|}{\xi}\right). \tag{4.9}$$

The correlation strength γ and correlation length ξ are of order unity and increase with c.m. energy.) Due to the translation invariance all bins give equal contributions. The integration domain for the bins is shown in Fig. 6 to be the set of squares along the diagonal. Integration of (4.9) gives

$$F_2(\delta y) = 1 + 2\gamma\xi^2\left(\delta y/\xi - 1 + e^{-\delta y/\xi}\right)(\delta y)^2. \tag{4.10}$$

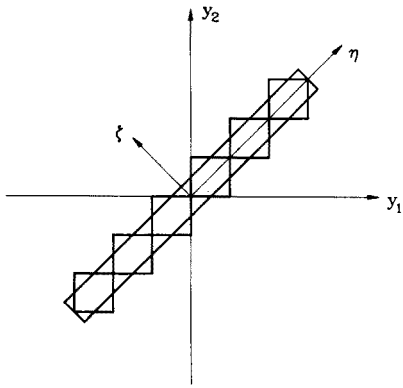

Fig. 6. This figure exhibits the geometry and integration domains that allow the expression of bin-averaged factorial moments and strip integrals in terms of the two-particle correlation function. The squares give bins (here 6 of width δy each), counting $\langle n_i(n_i - 1)\rangle$ for each bin. The strip domain has equal area to the sum of squares (but note that the variables η and ζ are stretched and contracted by $\sqrt{2}$ on this plot); it leads naturally to the pair-counting formula of Eq. 16. The rotated coordinate axes of the strip approach, expressed in the variables η and ζ, is shown.

Expressions such as (4.9), or more realistic fits such as

$$\exp\left(-\frac{(y_1 + y_2)^2}{A}\right)\exp\left(-\frac{(y_1 - y_2)^2}{B}\right)$$

suggest that the bin (box) geometry is not the most natural one to use. Instead the strip integral of equal area suggests itself. Redefining K_2 in

(4.6) for the strip geometry, we find from Eq. (4.9)

$$F_2(\delta y) = 1 + \gamma \int_{-\delta y/2}^{\delta y/2} d\zeta \exp(-\frac{|\zeta|}{\xi}),$$

$$F_2(\delta y) = 1 + \gamma \frac{\left(1 - \exp\left(-\frac{\delta y}{2\xi}\right)\right)}{\delta y/2\xi}. \tag{4.11}$$

In Fig. 7 we have compared (4.10)-(4.11); clearly either is a good approximation to the other. One can also show that if a box is covered by a small number of strips, very accurate numerical agreement is obtained. (For many purposes the strip definition is preferred since fluctuations due to binning are reduced.) Since the shapes of (4.10)-(4.11) are similar, one is easily mapped on the other by tuning the parameters γ and ξ. Fig. 5 shows how extremely good fits (for the UA1 experiment) can be obtained from such a primitive formula as (4.11).

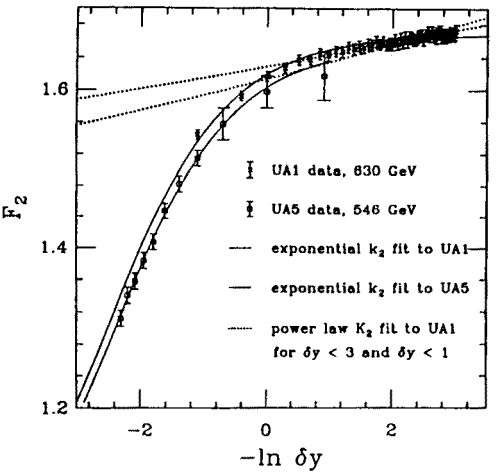

Fig. 7. Best fits for F_2^n moments for exponential correlation $k_2 = \gamma \exp(-|y_1 - y_2|/\xi)$ for UA1 [$\gamma = 0.669$, $\xi = 3.24$] and UA5 [$\gamma = 0.656$, $\xi = 2.95$] (solid lines). Dashed lines show best fits using a power-law correlation $k_2 = c(y_1 - y_2)^{-\nu}$. For the latter, fits were done on all points $\delta y < 3$ and $\delta y < 1$, giving [$c = 0.61$, $p = 0.033$] and [$c = 0.63$, $p = 0.022$] respectively.

In order to extend the foregoing analysis to higher orders of F_p moments and of the corresponding correlations it is important to recognize that the density correlations typically [32] contain lower order background correlations that are best handled by the introduction of cumulant correlation

functions C_p:

$$C_2(y_1,y_2) = \rho_2(y_1,y_2) - \rho_1(y_1)\rho_1(y_2),$$
$$C_3(y_1,y_2,y_3) = \rho_3(y_1,y_2,y_3) - \sum_{\text{perms}} \rho_2(y_1,y_2)\rho_1(y_3) + 2\rho_1(y_1)\rho_1(y_2)\rho_1(y_3). \quad (4.12)$$

Statistical independence of any y_i results in factorization of the ρ_p correlations and vanishing of the cumulant. Correspondingly we can invert (4.12) to get

$$r_2 \equiv \rho_2(y_1,y_2)/\rho_1(y_1)\rho_1(y_2) = 1 + k_2(y_1,y_2),$$
$$r_3 \equiv \rho_3(y_1,y_2,y_3)/\rho_1(y_1)\rho_1(y_2)\rho_1(y_3) = 1 + k_2(y_1,y_2) + k_2(y_2,y_3) + k_2(y_3,y_1) + k_3(y_1,y_2,y_3), \quad (4.13)$$

with k_2 given by (4.8) and $k_3 = C_3/\rho_1(y_1)\rho_1(y_2)\rho_1(y_3)$. Comparison with Eqs (2.12) should convince the reader that Eqs (4.13) are the correlation counterparts of the moment identities. Either (4.12) or (4.13) expands the F moment (or the density correlation) in a series of terms exhibiting increasingly complex correlations. The context of Eqs (2.12) and (4.13) is similar to carrying out a phase shift analysis of a scattering amplitude: here we are interested in the breakdown with respect to increasing orders of (cumulant) correlations.

We note several uses of Eqs (2.12) (refined as in the sequence of Eqs (4.1)-(4.7)). First of all we can systematically [33] extract $K_2, K_3, K_4, \ldots$ from the measured factorial moments, and study their dependence on energy, rapidity and bin width. The K_p are the key quantities to be produced by theoretical models. As it turns out, a large fraction of the numerical value of the F_p moments is produced by the K_2 contributions alone, until CERN collider energies at which the higher moments are fairly sizable (and interesting). The related data analysis is too lengthy to be given here, but can be found in Refs [33] and [37].

What about the higher order reduced cumulant correlations K_p appearing in Eq. (4.13)? How are they related to the negative binomial moments of Eq. (2.17)?

The following linked pair ansatz [32] provides the connection:

$$k_3(1,2,3) = \tfrac{2}{3}[k_2(1,2)k_2(2,3) + 2 \text{ perms}], \quad (4.14)$$

$$k_4(1,2,3,4) = \tfrac{6}{12}[k_2(1,2)k_2(2,3)k_2(3,4) + 11\text{perms}]. \quad (4.15)$$

For k_p the numerator is $(p-1)!$ and the denominator $^{p!}/_2$ counts the number of symmetrized terms in the bracket. The connectivity is required for

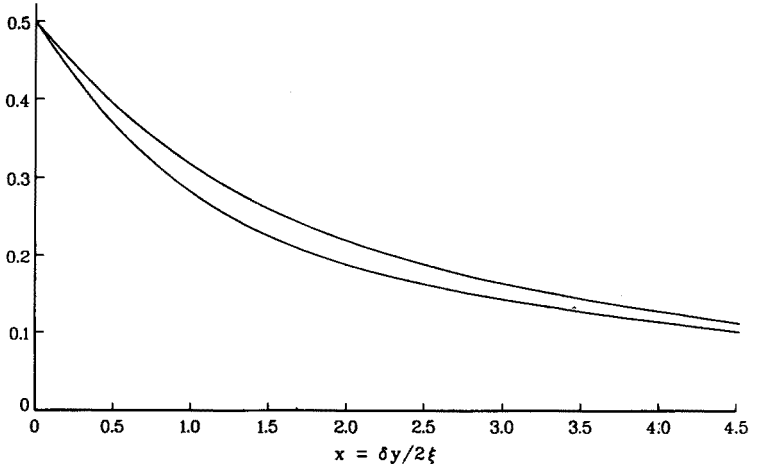

Fig. 8. Here the box integral for $(F_2-1)/\gamma$ (4.10) is compared with $(F_2-1)/\gamma$ for the strip integral, in the case of the exponential cumulant correlation function $k_2 = \gamma \exp(-|y_1-y_2|/\xi)$. Note that $(F_2-1)/\gamma = K_2/\gamma = 1/k\gamma$, k the negative binomial parameter when that is appropriate. Note $(K_2/\gamma)_{\rm box} = [x-2+\exp(-2x)]/x^2$ while $(K_2/\gamma)_{\rm strip} = (1-\exp(-x))/x$. Numerically these functions are close, as could be surmised from Fig. 7.

statistical dependence of the cumulant. In Fig. 7(a) we show the "linked pairs" of Eqs (4.14)-(4.15). Note that other linked graphs allowed by graph theory are not included (Fig. 8(b)).

In our original conjecture [32] we introduced coefficients A_p instead of $(p-1)!$: the data fit produced values close to negative binomial ones [38]. Therefore it seems that the linked pair approximation (LPA) merits further study both for phenomenological purposes and for derivation from a dynamical-stochastic framework. Of course one is accustomed to express correlations, Green's functions, *etc.* in terms of two-body correlations of Feynman propagators. But that is only possible when one knows the free energy or S-matrix in terms of observables, unlike the situation in QCD.

As mentioned, the (integrated) moments are not sensitive to details of the correlation function. On the other hand it should be possible to check the structure of k_3 for selected rapidity values.

Curiously, the same correlation structures (4.14)-(4.15) describe galaxy correlations, [39] although the magnitude of the coefficients is not as well determined as for multihadron production. And for cluster-cluster correlations, [40] the coefficient $^2/_3$ in k_3 comes directly from data.

This unlikely coincidence of correlation structures is cause for some meditation. To be sure the two particle cumulant k_2 is cutoff for one-dimensional

rapidity distributions and scaling ($\propto r^{-1.8}$) for galaxies. Yet both theories are basically self-similar and probably involve dissipation via some cascade mechanism. Hopefully the answer to this puzzle will be available by next year.

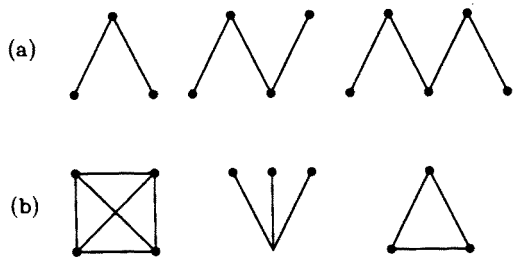

Fig. 9. The linked pair structure of Eqs (4.14)-(4.15) is indicated graphically in (a). Other topological structures allowed by graph theory are not needed to describe moment data. If present they would modify negative binomial counts.

A further insight into this situation is provided by the intermittency (factorial moment) analysis recently carried out to fifth order by Chmaj, Doroba and Słominski [41]. Excellent scaling is seen (in this case the highest cumulants dominate) and the magnitudes seem in reasonable agreement with linking with negative binomial coefficients.

This work was supported in part by the United States Department of Energy, and the Alexander von Humboldt Foundation. The author is indebted to many colleagues for enlightening conversations and especially to his collaborators C. C. Shih, Minh Duong- Van, Ina Sarcevic, Hans Eggers, Tugrul Hakioglu and Peter Lipa for their participation in this work.

REFERENCES

[1] P. Carruthers, *Int. J. Mod. Phys.* **A4**, 5587 (1989).
[2] P.J.E. Peebles, *The Large-Scale Structure of the Universe*, Princeton Univ. Press, Princeton 1980.
[3] A. Białas, R. Peschanski, *Nucl. Phys.* **B273**, 703 (1986); **B308**, 857 (1988).
[4] H.G.E. Hentschel, I. Procaccia, *Physica* **D8**, 435 (1983).
[5] P. Carruthers, C.C. Shih, *Phys. Rev. Lett.* **62**, 2073 (1989).
[6] F. Takagi, *Phys. Rev. Lett.* **53**, 427 (1984); *Phys. Rev.* **C32**, 1799 (1985).
[7] P. Carruthers, T. Hakioglu, to be published.
[8] T.C. Halsey, M.H. Jensen, L.P. Kadanoff, I. Procaccia, B.I. Schraiman, *Phys. Rev.* **A33**, 1141 (1986).
[9] R. Hwa, *Phys. Rev.* **D41**, 1456 (1990).

[10] J.M. Combes, A. Grossman, Ph. Tchamitchian, eds., *Wavelets* Springer-Verlag, Berlin 1989.
[11] M. Kendall, A. Stuart, *The Advanced Theory of Statistics*, Charles Griffin, London 1971, vol.1.
[12] P. Carruthers, C.C. Shih, *Int. J. Mod. Phys.* **A2**, 1447 (1987).
[13] M. Planck, *Sitzungsber. Deutsch. Akad. Wie Berlin* **33**, 365 (1932).
[14] L. Mandel, *Proc. Phys. Soc.* **74**, 233 (1959).
[15] A. Giovannini, *Nuovo Cimento* **10A**, 713 (1972); **15A**, 543 (1973); **24A**, 421 (1974); **34A**, 547 (1976).
[16] P. Carruthers, C.C. Shih, *Phys. Lett.* **127B**, 242 (1983).
[17] UA5 Collaboration, G.J. Alner *et al., Phys. Rept.* **154**, 247 (1987).
[18] Z. Koba, H.B. Nielsen, P. Olesen, *Nucl. Phys.* **B40**, 319 (1991).
[19] B. Buschbeck, P. Lipa, R. Peschanski, *Phys. Lett.* **215B**, 788 (1988).
[20] J. Klauder, E.C.G. Sudarshan, *Quantum Optics*, W.A. Benjamin, New York 1968.
[21] R.J. Glauber, *Phys. Rev.* **131**, 109 (1963).
[22] P. Carruthers, M.M. Nieto, *Am. J. Phys.* **33**, 537 (1966).
[23] M. Gaździcki, R. Szwed, G. Wrochna, A. Wrobewski, *Mod. Phys. Lett.* **A6**, 981 (1991).
[24] P. Carruthers, C.C. Shih, *Phys. Lett.* **165B**, 209 (1985); *Mod. Phys. Lett.* **A2**, 89 (1987).
[25] E. Hubble, *Ap. J.* **79**, 8 (1934).
[26] P. Carruthers, Minh Duong-Van, *Phys. Lett.* **131B**, 116 (1983).
[27] F. Zwicky, E. Herzog, P. Wild, M. Karpowicz, C. J. Mowal, *Catalog of Galaxies and Clusters of Galaxies*, California Institute of Technology, 1961-68, vols 1-6.
[28] P. Crane, W. C. Saslaw, *Ap. J.* **301**, 1 (1986).
[29] J.N. Fry, R. Giovanelli, M.P. Haynes, A. M. Melott, R.J. Scherrer, *Ap. J.* **340**, 11 (1989).
[30] E. DeWolf; S. Hegyi: private communications.
[31] P. Lipa, H.C. Eggers, F. Botterweck, M. Charlet, University of Arizona preprint AZPH-TH/91-14.
[32] P. Carruthers, I. Sarcevic, *Phys. Rev. Lett.* **63**, 1562 (1989).
[33] P. Carruthers, H.C. Eggers, I. Sarcevic, *Phys. Lett.* **B254**, 258 (1991).
[34] NA22 Collaboration: I. V. Ajinenko, *et al., Phys. Lett.* **B222**, 30 (1989); **B235**, 373 (1990).
[35] UA1 Collaboration, C. Albajar, *et al., Nucl. Phys.* **B345**, 1 (1990).
[36] P. Carruthers, *Phys. Rev.* **A43**, 2632 (1991).
[37] P. Carruthers, P. Lipa, I. Sarcevic, *Phys. Rev.* D,in press.
[38] E. DeWolf, *Acta Phys. Pol.* **B21**, 611 (1990).
[39] P. Carruthers, *Ap. J.*, in press.
[40] G. Toth, J. Jollos, A. Szalay, *Ap. J.* **344**, 75 (1989).
[41] T. Chmaj, W. Doroba, W. Słominski, *Z. Phys.* **C50**, 333 (1991).