PHYSICAL REVIEW D

PARTICLES AND FIELDS

THIRD SERIES, VOLUME 53, NUMBER 9 **1 MAY 1996**

RAPID COMMUNICATIONS

Rapid Communications are intended for important new results which deserve accelerated publication, and are therefore given priority in the editorial office and in production. A Rapid Communication in **Physical Review D** *should be no longer than five printed pages and must be accompanied by an abstract. Page proofs are sent to authors, but because of the accelerated schedule, publication is generally not delayed for receipt of corrections unless requested by the author.*

Generalized moments and cumulants for samples of fixed multiplicity

P. Lipa, H. C. Eggers,* and B. Buschbeck

Institut für Hochenergiephysik, Österreichische Akademie der Wissenschaften, Nikolsdorfergasse 18, A–1050 Vienna, Austria

(Received 11 December 1995)

Factorial moments and cumulants are usually defined with respect to the unconditioned Poisson process. Conditioning a sample by selecting events of a given overall multiplicity N necessarily introduces correlations. By means of Edgeworth expansions, we derive generalized cumulants which define correlations with respect to an arbitrary process rather than just the Poisson case. The results are applied to correlation measurements at fixed N, to redefining short-range versus long-range correlations and to normalization issues.

PACS number(s): 13.85.Hd, 02.50.−r, 05.40.+j, 25.75.Gz

In high-energy multiparticle production, the variation of correlations with overall multiplicity N is most sensitive to the underlying dynamics [1–5]. This sensitivity provides a good opportunity to test and discriminate between various models. However, with the notable exceptions of Refs. [6,7], little attention has so far been devoted to a systematic treatment of N-dependent correlation measures.

Selecting events by fixed values of N is an example of *conditioning*. One of the merits of this type of conditioning is the separation of so-called ''short-range'' from ''long-range'' correlations. The latter arise from non-Poissonian overall multiplicity distributions as well as the N dependence of the conditional single particle density $\rho(x|N)$ (where x is any phase space variable of interest).

Every conditioning, however, introduces new correlations which can be regarded as ''unphysical'' since they are merely a consequence of the selection procedure. Consequently, the very concept of a correlation must be rethought and generalized for the case of conditioned samples.

In this Rapid Communication, we seek to clarify and generalize factorial moments and cumulants for the case of fixed-multiplicity samples. This generalization suggests a natural strategy for separating nontrivial fixed-N correlations from overall multiplicity conditioning effects. Generalized cumulants will hence also be useful in the study of pion interferometry as a function of N and other situations requiring conditioning.

Multiparticle final states are best understood in terms of *point processes* [8,9]. The concept of correlation in point processes is usually based on the fundamental *Poisson process*, which is characterized by the factorial moment generating functional (FMGF)

$$Q_\gamma[\lambda(x)] = \exp\left[-\int \lambda(x)\rho(x)dx \right], \tag{1}$$

where $\rho(x) = (1/\sigma_I) d\sigma_{\rm incl}/dx$ is the differential inclusive cross section at point x. Equivalently, for the Poisson process every finite-dimensional *counting distribution*, i.e., the joint probability of the particle counts $\vec{n} = (n_1, \ldots, n_M)$ in an arbitrary finite partition of an overall phase space domain $\Omega_{\rm tot}$ into M nonoverlapping subdomains (''bins'') $\vec{\Omega} = (\Omega_1, \ldots, \Omega_M)$, factorizes into a product of M Poissonians:

$$\gamma(\vec{n};\vec{\nu}) = \prod_{m=1}^{M} \gamma(n_m;\nu_m), \tag{2}$$

*Present address: Department of Physics, University of Stellenbosch, 7600 Stellenbosch, South Africa.

where $\gamma(n_m;\nu_m)=\exp(-\nu_m)\nu_m^{n_m}/n_m!$ with mean multiplicity $\nu_m=\int_{\Omega_m}\rho(x)dx$.

Particles created in a Poisson process are completely uncorrelated in the sense that all factorial cumulant densities of order $q>1$ vanish:

$$\kappa_q(x_1,\ldots,x_q)\equiv\prod_{i=1}^{q}\left(\frac{-\delta}{\delta\lambda(x_i)}\right)\ln Q_\gamma[\lambda(x)]\,\Big|_{\lambda=0}=0, \tag{3}$$

while $\kappa_1(x)=\rho(x)$.

By contrast, the *multinomial process* is defined by placing a fixed number N of particles *independently* of each other into the domain $\Omega_{\rm tot}$ according to a probability density $p(x|N)=\rho(x|N)/N$; it has a FMGF

$$Q_N^{\rm mult}[\lambda(x)]=\left[1-\int dx\lambda(x)\frac{\rho(x|N)}{N}\right]^N, \tag{4}$$

and its finite-dimensional counting distributions are multinomials.

Despite the fact that the particles are placed independently of each other, the standard factorial cumulants of the multinomial process are nevertheless nonzero, thereby indicating purely ''external'' correlations:

$$\kappa_q^{\rm mult}(x_1,\ldots,x_q|N)=(q-1)!(-N)^{1-q}\prod_{i=1}^{q}\rho(x_i|N). \tag{5}$$

In other words, standard cumulants measure not only the ''internal'' correlations between particles but also the deviation of the overall multiplicity N from the Poisson distribution $\gamma(N;\overline{N})=\exp(-\overline{N})\overline{N}^N/N!$.

It thus seems natural to ask for quantities that are sensitive not to such ''external'' deviations from the overall Poissonian but only to ''internal'' correlations, i.e., deviations from the multinomial.

The answer, we believe, lies in the use of *generalized cumulants* which successfully separate the two. To justify their existence and form, we turn to the alternative definition of cumulants as coefficient functions in an Edgeworth expansion. In most of the literature, the latter is understood as the expansion of a continuous univariate distribution $P(x)$ around the Gaussian (cf. [10]). For integer counts, it is more natural to define a multivariate *discrete Edgeworth expansion* around the Poisson process. In its most general form, the discrete Edgeworth expansion of a point process f with respect to the Poisson process γ reads, in terms of their FMGF's,

$$\begin{aligned}Q_f[\vec{\lambda}]&=\exp\left[-\int dx[\kappa_1^f(x)-\rho(x)]\lambda(x)\right.\\&\quad\left.+\sum_{q=2}^{\infty}\frac{(-1)^q}{q!}\int d\vec{x}\,\kappa_q^f(\vec{x})\vec{\lambda}(\vec{x})\right]Q_\gamma[\vec{\lambda}]\\&=\exp\left[\sum_{q=1}^{\infty}\frac{(-1)^q}{q!}\int d\vec{x}\,\kappa_q^f(\vec{x})\vec{\lambda}(\vec{x})\right],\end{aligned} \tag{6}$$

with $\vec{x}=(x_1,\ldots,x_q)$ and $\vec{\lambda}(\vec{x})=\lambda(x_1)\cdots\lambda(x_q)$. The univariate expansion of the overall multiplicity distribution $f(N)$ in terms of the Poissonian is

$$f(N)=\exp\left[(\kappa_1^f-\overline{N})(-\nabla)+\sum_{q=2}^{\infty}\frac{\kappa_q^f}{q!}(-\nabla)^q\right]\gamma(N;\overline{N}), \tag{7}$$

where ∇ is the discrete difference operator such that $\nabla\gamma(n;\nu)\equiv\gamma(n;\nu)-\gamma(n-1;\nu)$.

The extension to a multivariate expansion for the joint counting distributions in arbitrary finite partitions of $\Omega_{\rm tot}$ then reads

$$\begin{aligned}f(\vec{n})=\exp\left[\sum_{m=1}^{M}[\kappa_1^f(m)-\nu_m](-\nabla_m)\right.\\\left.+\sum_{q=2}^{\infty}\frac{(-1)^q}{q!}\sum_{m_1,\ldots,m_q}\kappa_q^f(\vec{m})\nabla_{m_1}\cdots\nabla_{m_q}\right]\gamma(\vec{n};\vec{\nu}),\end{aligned} \tag{8}$$

where $\nabla_m\gamma(\vec{n};\vec{\nu})\equiv\gamma(\vec{n};\vec{\nu})-\Pi_{m'}\gamma(n_{m'}-\delta_{m'm};\vec{\nu})$, and $\kappa_q^f(\vec{m})$ is the q-fold integral of $\kappa_q^f(\vec{x})$ over the domains $\Omega_{m_1},\ldots,\Omega_{m_q}$.

It can be verified easily that the coefficients κ_q^f in these expansions are identical with the qth order multivariate factorial cumulants of the process f. Therefore standard cumulants can be *defined* as the coefficient functions of the Edgeworth expansion around the Poisson process.

This definition can be generalized in an obvious way. From Eq. (6), it is easy to show that the generating functional Q_f of a process f can be expanded in terms of the FMGF of *any other* process[1] h:

$$Q_f[\vec{\lambda}]=\exp\left[\sum_{q=1}^{\infty}\frac{(-1)^q}{q!}\int d\vec{x}[\kappa_q^f(\vec{x})-\kappa_q^h(\vec{x})]\vec{\lambda}(\vec{x})\right]Q_h[\vec{\lambda}], \tag{9}$$

where again κ_q^f and κ_q^h represent the standard factorial cumulants of f and h.

It is quite natural, then, to define *generalized cumulants* as the coefficients in the expansion of f in terms of h,

$$\kappa_q^{f/h}\equiv\kappa_q^f-\kappa_q^h, \tag{10}$$

for $q\geqslant 2$, while for $q=1$, we set by convention $\kappa_1^{f/h}\equiv\kappa_1^f$. If h is the Poisson process, these generalized cumulants reduce to the standard cumulants.

The generalized cumulants can be found as functional derivatives of the *ratio* of generating functionals:

$$\kappa_q^{f/h}=\prod_{i=1}^{q}\left(\frac{-\delta}{\delta\lambda(x_i)}\right)\ln\left(\frac{Q_fQ_\gamma}{Q_h}\right)\,\Big|_{\lambda=0}, \tag{11}$$

[1]It is understood that such relations are valid only for processes having well-defined cumulants of all orders.

where the Poisson generating functional Q_γ has to satisfy $(\delta Q_\gamma/\delta\lambda)|_{\lambda=0}=\kappa_1^h$. This ensures that the above convention for $\kappa_1^{f/h}$ is satisfied.[2] Note that $\kappa_q^{f/h}$ is generally a standard cumulant to some well-defined process only if the process f is divisible by h in the sense that the ratio Q_f/Q_h itself again satisfies all requirements for a FMGL.

The general validity of Eq. (9) means that generalized cumulants will work for any type of conditioning, i.e., they will show up the real correlations of a process f with respect to a given conditioned reference process h.

We now apply the concept of generalized cumulants to fixed-N multiplicity samples. The discussion will range from the definition of internal cumulants for fixed-N measurements to resulting cumulant formulas for fixed-bin and correlation integral measurements, followed by a short discussion of improved short-range vs long-range correlation formulas, summing over windows in N, internal moments, and normalization.

Internal cumulants are the generalized cumulants with respect to the multinomial process (4):

$$\kappa_q^I(\vec{x}|N)\equiv\kappa_q(\vec{x}|N)-\kappa_q^{\rm mult}(\vec{x}|N), \tag{12}$$

with $\kappa_q^{\rm mult}(\vec{x}|N)$ given by Eq. (5) and $\kappa_q(\vec{x}|N)$ the standard cumulant measured for a fixed-N process. These internal cumulants will vanish everywhere, even on a differential level, if the N particles in the full domain $\Omega_{\rm tot}$ are placed independently of each other following the same probability density $p(x|N)$, i.e., behave multinomially.

Univariate internal cumulants are obtained from the multivariate ones by integrating all q variables $x_1,\ldots,x_q$ over the same domain Ω_m, $\kappa_q^I(m|N)\equiv\int_{\Omega_m}d\vec{x}\,\kappa_q^I(\vec{x}|N)$, yielding

$$\kappa_2^I(m|N)=\langle n_m^{[2]}\rangle_N-(1-N^{-1})\langle n_m\rangle_N^2,$$

$$\kappa_3^I(m|N)=\langle n_m^{[3]}\rangle_N-3\langle n_m^{[2]}\rangle_N\langle n_m\rangle_N+2(1-N^{-2})\langle n_m\rangle_N^3,$$

etc., with $n^{[q]}=n!/(n-q)!$. Event averages $\langle\ \rangle_N$ are taken over subsamples of fixed N. These integrated internal cumulants differ from the inclusive ones [11] by a correction factor $(1-N^{1-q})$ in the last term.

Integrated internal cumulants of order $q\geqslant 2$ vanish *for any fixed-N distribution* if the integration domain Ω_m is enlarged to the full domain $\Omega_{\rm tot}$, in contrast to the standard cumulants which integrate to $(-1)^{q-1}(q-1)!N$.

The same procedure can be used in the correlation integral method. In the notation of Ref. [9], if $a\equiv\Sigma_j\Theta(\epsilon-X_{ij})$ is the ''sphere count'' around particle i in event a and $\langle b\rangle_N\equiv\langle\Sigma_j\Theta(\epsilon-X_{ij}^{ab})\rangle_N$ is the average of counts within other events b, the lowest-order internal cumulants are correspondingly given by

$$\kappa_2^I(\epsilon|N)=\left\langle\sum_i\ [a-(1-N^{-1})\langle b\rangle_N]\right\rangle_N, \tag{13}$$

$$\kappa_3^I(\epsilon|N)=\left\langle\sum_i\ [a^{[2]}-\langle b^{[2]}\rangle_N-2a\langle b\rangle_N+2(1-N^{-2})\langle b\rangle_N^2]\right\rangle_N, \tag{14}$$

and so on; the same $(1-N^{1-q})$ factors appear as coefficients of the last terms $\langle b\rangle_N^{q-1}$ for higher-order internal cumulants. Again, these internal correlation integrals vanish for distances ϵ large enough to encompass the full domain $\Omega_{\rm tot}$.

Clearly, the multinomial correction will be most important for small N and lowest order. The correction may or may not be substantial, depending on the numerical size of the uncorrected κ_q.

When experimental samples are not very large, fixed-N correlation measurements become problematic due to large sampling errors. Therefore one traditionally sums up all fixed-N cumulants [12,2,4] to yield what are (inaccurately) called ''short-range correlations'' (SRC's); in second order

$$\kappa_2^{\rm SRC}(x_1,x_2)\equiv\{\kappa_2(x_1,x_2|N)\}=\sum_N\ P_N\kappa_2(x_1,x_2|N). \tag{15}$$

Inclusive cumulants are then split up into the SRC contribution plus ''long-range correlations'' (LRC's):

$$\kappa_2(x_1,x_2)=\{\kappa_2(x_1,x_2|N)\}+\{\Delta\rho_N(x_1)\Delta\rho_N(x_2)\}, \tag{16}$$

where $\Delta\rho_N(x)=\rho_1(x|N)-\{\rho_1(x|N)\}$. The SRC sum over fixed-N cumulants is purported to represent an average over the standard fixed-N correlations, while the second term represents the LRC contributions resulting from the strong variation of $\rho(x|N)$ with overall multiplicity N.

As we have shown, it is preferable to replace the $\kappa_q(\vec{x}|N)$ with internal cumulants $\kappa_q^I(\vec{x}|N)$ to eliminate the fixed-N multinomial contributions from the short-range part. We therefore propose that the traditional SRC-LRC formula should be modified in favor of splitting the inclusive cumulants into N averages over ''internal'' and ''external'' correlations. In second order, this would be

$$\kappa_2(x_1,x_2)=\{\kappa_2^I(x_1,x_2|N)\}+\{\kappa_2^{\rm mult}(x_1,x_2|N)+\Delta\rho_N(x_1)\Delta\rho_N(x_2)\}. \tag{17}$$

LRC or SRC formulas up to fourth order were catalogued in Ref. [7]. In all cases, ''internal'' correlations would be given by $\{\kappa_q(\vec{x}|N)-\kappa_q^{\rm mult}(\vec{x}|N)\}$, while the ''external'' correlations correspond to the formulas in [7] plus the appropriate multinomial cumulant.

For practical reasons, N averages of internal correlation measures over limited multiplicity ranges $[A,B]$ are of importance:

$$\kappa_q^I(\vec{x}|A,B)=\frac{\Sigma_{N=A}^B P_N\kappa_q^I(\vec{x}|N)}{\Sigma_{N=A}^B P_N}. \tag{18}$$

Fixed-N correlations are but one example of generalized cumulants. Equation (9) shows that subtraction of cumulants is appropriate when *any* kind of reference distribution is

[2]One can view the ratio Q_h/Q_γ as the *central moment* generating functional of the process h.

given. Applications which come to mind immediately are subensembles characterized by a given number of jets and any Monte Carlo–generated simulation. Differences between real data and the dynamics contained in a given Monte Carlo code would be quantifiable again by a generalized cumulant $\kappa \equiv \kappa_{\rm data} - \kappa_{\rm MC}$.

Generalized moments $\rho_q^{f/h}$ are defined as functional derivatives of the ratio of generating functionals,

$$\rho_q^{f/h}(\vec{x}) = \prod_{i=1}^{q} \left(\frac{-\delta}{\delta \lambda(x_i)} \right) \left(\frac{Q_f Q_\gamma}{Q_h} \right) \Bigg|_{\lambda=0}; \tag{19}$$

for the fixed-N case, the corresponding internal moments are related to the standard moments by

$$\rho_1^I(x|N) = \kappa_1^I(x|N) = \rho_1(x|N), \tag{20}$$

$$\begin{aligned} \rho_2^I(x_1,x_2|N) &= \kappa_2^I(x_1,x_2|N) + \kappa_1^I(x_1|N)\kappa_1^I(x_2|N), \\ &= \rho_2(x_1,x_2|N) + (1/N)\rho_1(x_1|N)\rho_1(x_2|N), \end{aligned} \tag{21}$$

and further, omitting the arguments for brevity,

$$\rho_3^I = \rho_3 + \rho_1\rho_1\rho_1\left(\frac{3}{N} - \frac{2}{N^2}\right), \tag{22}$$

$$\rho_4^I = \rho_4 + \frac{1}{N}\sum_{(6)} \rho_1\rho_1\rho_2 - \rho_1\rho_1\rho_1\rho_1\left(\frac{5}{N^2} - \frac{6}{N^3}\right), \tag{23}$$

where the figure in parentheses under the sum indicates the number of permutations. Internal moments have the property that when the *measured* moment $\rho_q(\vec{x}|N)$ behaves multinomially, $\rho_q(\vec{x}|N) \overset{\rm mult}{\to} (N^{[q]}/N^q)\Pi_{i=1}^q \rho_1(x_i|N)$, they factorize without a prefactor:

$$\rho_q^I(\vec{x}|N) \overset{\rm mult}{\to} \prod_{i=1}^{q} \rho_1(x_i|N). \tag{24}$$

For a subsample of fixed N, moments and cumulants at fixed multiplicity can be normalized in two ways. The most natural normalization procedure is to use in the denominator exactly that quantity which the numerator would default to if the process were fully independent. Hence the normalization for fixed-N factorial moments over some domain Ω_m would be

$$\begin{aligned} F_q(\Omega_m|N) &= \frac{\int_{\Omega_m} d\vec{x}\,\rho_q(\vec{x}|N)}{\int_{\Omega_m} d\vec{x}\,\rho_q^{\rm mult}(\vec{x}|N)}, \\ &= \frac{N^q}{N^{[q]}} \frac{\int_{\Omega_m} d\vec{x}\,\rho_q(\vec{x}|N)}{\int_{\Omega_m} d\vec{x}\,\rho_1(x_1|N)\ldots\rho_1(x_q|N)}, \end{aligned} \tag{25}$$

while the internal moment would be normalized according to

$$F_q^I(\Omega_m|N) = \frac{\int_{\Omega_m} d\vec{x}\,\rho_q^I(\vec{x}|N)}{\int_{\Omega_m} d\vec{x}\,\rho_1(x_1|N)\ldots\rho_1(x_q|N)}. \tag{26}$$

Both these definitions yield $F_q = F_q^I \equiv 1$ for any Ω_m when the measured $\rho_q(\vec{x}|N)$ behaves multinomially within the total window $\Omega_{\rm tot}$. For internal cumulants, the same normalization (24) as for the internal moments would be appropriate.

Internal moments integrate under the total phase space domain $\Omega_{\rm tot}$ to $\int_\Omega d\vec{x}\,\rho_q^I(\vec{x}) = N^q$, while $\rho_q(\vec{x}|N)$ integrates to $N^{[q]}$, so that both normalized moments become unity when integrated over $\Omega_{\rm tot}$ for *any* distribution:

$$F_q^I(\Omega_{\rm tot}|N) = F_q(\Omega_{\rm tot}|N) = 1. \tag{27}$$

The integral over $\Omega_{\rm tot}$ of normalized internal cumulants will, of course, be zero just as the unnormalized ones.

In summary, we have shown that generalized cumulants are an improved measure of correlations in samples of fixed multiplicity. The latter are more sensitive in discriminating between dynamical mechanisms and models than inclusive quantities. The procedure outlined here can be applied to a number of other conditioning problems within and beyond high energy physics. Detailed applications to correlation measurements such as pion interferometry are in progress.

We wish to thank Peter Carruthers for encouraging discussions. P.L. gratefully acknowledges support by the Austrian Academy of Sciences through APART (Austrian Programme for Advanced Research and Technology). H.C.E. was kindly supported by the Austrian Fonds zur Förderung der Wissenschaftlichen Forschung (FWF).

[1] R. Singer *et al.*, Phys. Lett. **49B**, 481 (1974); K. Eggert *et al.*, Nucl. Phys. **B86**, 201 (1975); B. Y. Oh *et al.*, Phys. Lett. **56B**, 400 (1975).

[2] L. Foà, Phys. Rep. **22**, 1 (1975); J. Whitmore, *ibid.* **27**, 187 (1976).

[3] W. Bell *et al.*, Z. Phys. C **22**, 109 (1984); **32**, 335 (1986).

[4] UA5 Collaboration, R. E. Ansorge *et al.*, Z. Phys. C **37**, 191 (1988).

[5] UA1 Collaboration, C. Albajar *et al.*, Nucl. Phys. **B345**, 1 (1990); UA1 Collaboration, Y. F. Wu *et al.*, Act. Phys. Slov. **44**, 141 (1994); P. Lipa, B. Buschbeck, and H. C. Eggers (unpublished).

[6] E. L. Berger, Nucl. Phys. **B85**, 61 (1975).

[7] P. Carruthers, Phys. Rev. A **43**, 2632 (1991).

[8] S. K. Srinivasan, *Stochastic Point Processes*, Griffin's Statistical Monographs and Courses Vol. 34 (Griffin, London, 1974).

[9] H. C. Eggers, P. Lipa, P. Carruthers, and B. Buschbeck, Phys. Rev. D **48**, 2040 (1993).

[10] A. Stuart and J. K. Ord, *Kendall's Advanced Theory of Statistics*, 5th ed. (Oxford University Press, New York, 1987), Vol. 1.

[11] P. Carruthers, H. C. Eggers, and I. Sarcevic, Phys. Lett. B **254**, 258 (1991).

[12] A. Białas, in *4th International Symposium on Multiparticle Dynamics*, Pavia, Italy, 1973, edited by F. Duimio, A. Giovannini, and S. Ratti (World Scientific, Singapore, 1974).