Письма в ЭЧАЯ. А. 2011. Т. 8, № 9. С. 224–229

EDGEWORTH VERSUS GRAM–CHARLIER SERIES: x-CUMULANT AND PROBABILITY DENSITY TESTS

M. B. de Kock [a], *H. C. Eggers* [a], *J. Schmiegel* [b]

[a] Department of Physics, University of Stellenbosch, Stellenbosch, South Africa

[b] Thiele Centre for Applied Mathematics in Natural Sciences, Department of Mathematics, Åarhus University, Åarhus, Denmark

Edgeworth series are often considered the same as Gram–Charlier series in systematic expansions of non-Gaussian probability distributions. Testing direct approximations of the probability itself as well as of cumulants in coordinate space as functions of measured cumulants in momentum space, we show how the former far outperforms the latter.

PACS: 02.30.Rz

The goal of femtoscopy is to find as much information as possible on the emission function $S(\mathbf{q})$ in coordinate space, given a measured correlation function in momentum space,

$$R(\mathbf{q}) = C(\mathbf{q}) - 1 = \int d^3x\, S(\mathbf{x}) \left[|\psi(\mathbf{q},\mathbf{x})|^2 - 1\right]. \tag{1}$$

Inversion of this integral equation is in general a highly nontrivial business (see, for example, [1]). Here, we use the simplest version, that of a noninteracting final state,

$$R(\mathbf{q}) = \int d^3x\, S(\mathbf{x})\, \mathrm{e}^{i\mathbf{q}\cdot\mathbf{x}} \tag{2}$$

reduced to one dimension, $\mathbf{q} \to q$, $\mathbf{x} \to x$, to make quantitative comparisons between Edgeworth and Gram–Charlier expansions of non-Gaussian distributions such as often encountered in experimental femtoscopy. We test the accuracy of these expansions using the probability density functions (pdf's)

$$f(q) = \frac{R(q)}{\int dq\, R(q)} = f(0) \int dx\, \mathrm{e}^{iq\cdot x} g(x), \tag{3}$$

$$g(x) = \frac{S(x)}{f(0)} = \int \frac{dq\, \mathrm{e}^{-iq\cdot x}\, f(q)}{2\pi f(0)} \tag{4}$$

themselves and coordinate-space cumulants as touchstones. Cumulants are relevant because, given $f(q)$, its q-moments $\mu_r^{(q)} = \int dq\, f(q)\, q^r$ and q-cumulants $\kappa_r^{(q)}$ of lowest orders $r = 1, 2, 3, \ldots$ provide fundamental information on the properties of $f(q)$: $\mu_1^{(q)} = \kappa_1^{(q)}$ is a measure of the location of the peak of f, the variance $\kappa_2^{(q)} = \mu_2^{(q)} - (\mu_1^{(q)})^2 = \sigma^2$ measures its dispersion, the skewness $\gamma_3^{(q)} = \kappa_3^{(q)}/\sigma^3$ measures its asymmetry and the kurtosis $\gamma_4^{(q)} = \kappa_4^{(q)}/\sigma^4$ is a first description of the rate of decay of f. Higher-order «generalised kurtoses» $\gamma_r^{(q)} = \kappa_r^{(q)}/\sigma^r$ would provide successively more detail. Equivalent relations hold in coordinate space between x-moments, x-cumulants and $g(x)$, e.g., $\mu_r^{(x)} = \int dx\, g(x)\, x^r$, $\kappa_2^{(x)} = \mu_2^{(x)} - (\mu_1^{(x)})^2$, etc.

Writing $D_x^r = (d/dx)^r$ for short, q-moments are derivatives of the generating function $\Phi(x) = 2\pi f(0) g(-x) = \int dq\, \mathrm{e}^{iqx} f(q)$ and q-cumulants derivatives of its logarithm,

$$\kappa_r^{(q)} = (-i)^r D_x^r \ln \Phi(x)\big|_{x=0}. \tag{5}$$

This fixes relations between moments and cumulants to all orders. For identical particles, both $C(q)$ and $g(x)$ are symmetric, so that moments and cumulants of odd order vanish and the even-order relations in both q-space and x-space become $\kappa_2 = \mu_2$ and $\kappa_4 = \mu_4 - 3\mu_2^2$ and $\kappa_6 = \mu_6 - 15\mu_4\mu_2 + 30\mu_2^3$, etc.

While for purely Gaussian sources, the second-order cumulants are related by $\kappa_2^{(x)} = 1/\kappa_2^{(q)}$ and all higher-order cumulants are identically zero, neither of these statements is true in general. We consider both the modification of $\kappa_2^{(x)}$ resulting from nonzero $\gamma_r^{(q)}$ and of the x-kurtosis $\gamma_4^{(x)} = \kappa_4^{(x)}/(\kappa_2^{(x)})^2$. Since x-moments are found from the generating function $\Phi(q) = f(q)\,/\,f(0)$ through

$$\mu_r^{(x)} = (-i)^r D_q^r \Phi(q)\big|_{q=0}, \tag{6}$$

we can obtain x-cumulants as combinations of measured q-cumulants.

Gauss Gram–Charlier (GGC) and Edgeworth (GEW) series expansions systematically compare the measured $f(q)$ with a Gaussian reference pdf $f_0(q)$,

$$f_0(q') = \frac{\mathrm{e}^{-q'^2/2}}{\sqrt{2\pi}} \tag{7}$$

with $q' = q/\sigma$. As shown elsewhere [2, 3], the GGC series results from expanding the generating function for the non-Gaussian $f(q')$ in powers of x',

$$\Phi(x') = \mathrm{e}^{-x'^2/2} \exp\left[\sum_{j=3}^{\infty} \zeta_j (ix')^j\right] = \mathrm{e}^{-x'^2/2} \sum_{m=0}^{\infty} \frac{c_m(\boldsymbol{\zeta})}{m!}\,(ix')^m, \tag{8}$$

where each $c_m(\boldsymbol{\zeta})$ is a polynomial in the set of q-kurtoses $\boldsymbol{\zeta} = \{\zeta_r = \gamma_r^{(q)}/r!\}_{r=4}^m$. Taking the inverse Fourier transform term by term, one obtains an expansion in terms of Chebychev–Hermite polynomials $H_r(q')$,

$$f(q') = f_0(q')\left[1 + \sum_{j=2}^{\infty} \frac{c_{2j}(\boldsymbol{\zeta})}{(2j)!} H_{2j}(q')\right], \tag{9}$$

$$H_r(q') = f_0^{-1}(q')\,(-D_{q'})^r f_0(q'), \tag{10}$$

with lowest-order terms (writing $H_r(q') = H_r$ for short)

$$f(q') = f_0(q')\Big[1 + \zeta_4 H_4 + \zeta_6 H_6 + \left(\zeta_8 + \frac{1}{2}\zeta_4^2\right) H_8 + (\zeta_{10} + \zeta_6\,\zeta_4)\,H_{10} + {} \\ {} + \left(\zeta_{12} + \zeta_8\,\zeta_4 + \frac{1}{2}\zeta_6^2 + \frac{1}{6}\zeta_4^3\right) H_{12} + \ldots\Big]. \tag{11}$$

With the help of the recursion relation $(-D_{q'})^r f_0(q')H_{2j}(q') = f_0(q')H_{2j+r}(q')$, the rth derivative of the x-moment generating function can, for even r, be expressed as

$$\Phi^{(r)}(q') = \mathrm{e}^{-q'^2/2}\left[H_r(q') + \zeta_4 H_{4+r}(q') + \zeta_6 H_{6+r}(q') + \ldots\right],$$

from which the x-cumulants follow as ratios of generating functions at $q' = 0$ in terms of generalised q-kurtoses $\gamma_r = \kappa_r^{(q)}/\sigma^r$; for example, the second-order x-cumulant is

$$\kappa_2^{(x)} = \frac{(-i)^2}{\kappa_2^{(q)}}\frac{\Phi^{(2)}(q')}{\Phi^{(0)}(q')}\bigg|_{q'=0} = \frac{1}{\kappa_2^{(q)}}\left[\frac{1+\frac{5}{8}\gamma_4 - \frac{7}{48}\gamma_6 + \frac{3}{128}(\gamma_8 + 35\gamma_4^2) + \ldots}{1+\frac{1}{8}\gamma_4 - \frac{1}{48}\gamma_6 + \frac{1}{384}(\gamma_8 + 35\gamma_4^2) + \ldots}\right], \tag{12}$$

while the x-kurtosis in fourth order is (omitting the arguments of Φ)

$$\gamma_4^{(x)} = \frac{\Phi^{(4)}\,\Phi^{(0)} - 3\Phi^{(2)}\,\Phi^{(2)}}{\Phi^{(2)}\,\Phi^{(2)}}\bigg|_{q'=0} = \left[\frac{\gamma_4 - \frac{1}{2}\gamma_6 + \frac{1}{8}\gamma_8 + \frac{15}{4}\gamma_4^2 + \ldots}{1+\frac{5}{4}\gamma_4 - \frac{7}{24}\gamma_6 + \frac{3}{64}\gamma_8 + \frac{65}{32}\gamma_4^2 + \ldots}\right], \tag{13}$$

with a similar expression for $\kappa_4^{(x)}$.

To infinite order, the above would be exact, at least formally, but for realistic applications, these series must of course be truncated at some finite order. To quantify the difference between such truncated versions of the GGC series and the exact expressions, we make use of the Symmetric Normal Inverse Gaussian (SNIG),

$$f(q\,|\,\alpha,\delta) = \frac{\alpha\delta\,\mathrm{e}^{\alpha\delta}\,K_1(\alpha\sqrt{\delta^2+q^2})}{\pi\sqrt{\delta^2+q^2}}, \tag{14}$$

a special case of the Normal Inverse Gaussian density [4], as a solvable toy model for $f(q')$ which yields exact expressions for both coordinate- and momentum-space cumulants (K_1 is the modified Bessel function). The SNIG reverts to a Gaussian in the limit $\alpha \to \infty$ and has q-moment generating function $\Phi(x\,|\,\alpha,\delta) = \exp\left[\delta\alpha - \delta\sqrt{\alpha^2+x^2}\right]$. Measuring $\kappa_2^{(q)}$ and $\gamma_4^{(q)}$ fixes the parameters α and δ, so that higher-order q-cumulants and kurtoses of the SNIG can be expressed as closed functions in terms of these quantities [2]. Similarly, using the SNIG pdf as x-moment generating function, we obtain exact expressions for x-cumulants. Omitting below the arguments of the Bessel functions, which are $\alpha\delta = 3/\gamma_4^{(q)}$ in every case, these «exact» x-cumulants are

$$\kappa_{2,\mathrm{SNIG}}^{(x)} = \frac{1}{\kappa_2^{(q)}}\frac{K_2}{K_1}, \tag{15}$$

$$\kappa_{4,\mathrm{SNIG}}^{(x)} = \frac{1}{\kappa_2^{(q)\,2}}\frac{3K_3K_1 - 3K_2^2}{K_1^2}, \tag{16}$$

$$\gamma_{4,\mathrm{SNIG}}^{(x)} = \frac{3K_3K_1 - 3K_2^2}{K_2^2}. \tag{17}$$

With these exact x-cumulants as reference, we can test the accuracy of various truncations of Eqs. (12), (13) as a function of the Gram–Charlier order $m = 2j$ of Eq. (9). The results are disastrous: truncated-GGC versions of $\kappa_2^{(x)}$ differ from the exact SNIG expression by up to 40% for $\gamma_4^{(q)} \simeq 1$, while the truncated $\gamma_4^{(x)}$ deviates from the exact expression by factors 2 or more. Furthermore, the accuracy of the approximations deteriorates with increasing order m. GGC series fail completely to approximate the exact x-cumulants.

By contrast, the Gauss–Edgeworth (GEW) series is derived by considering the random variable q' to be a pro-forma convolution of n identical independent random variables q_i each with pdf $f_1(q_i)$, a corresponding generating function $\Phi_1(x_i)$ and second-order cumulant $\kappa_2^{(q)}(n{=}1) = \sigma_1^2$, in terms of which the generating function for x', the dual to standardised variable q', is

$$\Phi(x') = \left[\Phi_1\left(\frac{x_i}{\sigma_1\sqrt{n}}\right)\right]^n . \tag{18}$$

Expanding the exponential in powers of $\ell = 1/\sqrt{n}$ rather than x' results in

$$\Phi(x') = \mathrm{e}^{-x'^2/2} \exp\left[\sum_{j=3}^{\infty} \zeta_j\,(ix')^j\,\ell^{j-2}\right] = \mathrm{e}^{-x'^2/2} \sum_{w=0}^{\infty} p_w(\boldsymbol{\zeta}, ix')\,\ell^w, \tag{19}$$

and again inverting term by term, one obtains the Gauss–Edgeworth series (again writing $H_r = H_r(q')$ for short),

$$\begin{aligned} f(q') = f_0(q') \Bigg[1 + \ell^2\zeta_4 H_4 + \ell^4\left(\frac{\zeta_4^2}{2!}H_8 + \zeta_6 H_6\right) + \ell^6\left(\frac{\zeta_4^3}{3!}H_{12} + \zeta_4\zeta_6 H_{10} + \zeta_8 H_8\right) + \\ + \ell^8\left(\frac{\zeta_4^4}{4!}H_{16} + \frac{\zeta_4^2\zeta_6}{2!}H_{14} + \frac{\zeta_6^2}{2!}H_{12} + \zeta_4\zeta_8 H_{12} + \zeta_{10}H_{10}\right) + \ldots \Bigg] . \end{aligned} \tag{20}$$

Unlike the equivalent GGC expansion of Eq. (11), in which the order of the expansion was identical with the order of H_m, a given term of order w in the GEW series (20) contains linear combinations of Hermite polynomials.

Edgeworth re-ordering of terms leads to expressions for the x-cumulants as ratios of power series in ℓ. For the SNIG test case, these series simplify to

$$\kappa_2^{(x)} = \frac{1}{\kappa_2^{(q)}}\left[\frac{1 + \frac{5}{8}\gamma_4\ell^2 + \frac{35}{384}\gamma_4^2\ell^4 - \frac{35}{3072}\gamma_4^3\ell^6 + \frac{385}{98304}\gamma_4^4\ell^8 + \ldots}{1 + \frac{1}{8}\gamma_4\ell^2 - \frac{5}{384}\gamma_4^2\ell^4 + \frac{35}{9216}\gamma_4^3\ell^6 - \frac{175}{98304}\gamma_4^4\ell^8 + \ldots}\right], \tag{21}$$

$$\gamma_4^{(x)} = \left[\frac{\gamma_4\ell^2 + \frac{5}{4}\gamma_4^2\ell^4 + \frac{35}{96}\gamma_4^3\ell^6 - \frac{35}{1152}\gamma_4^4\ell^8 + \ldots}{1 + \frac{5}{4}\gamma_4\ell^2 + \frac{55}{96}\gamma_4^2\ell^4 + \frac{35}{384}\gamma_4^3\ell^6 + \frac{35}{18432}\gamma_4^4\ell^8 + \ldots}\right]. \tag{22}$$

The disappearance of $\gamma_r^{(q)}$ with $r \geqslant 6$ from the above expression as compared to the equivalent GGC relations (12), (13) is due to the fact that the SNIG fixes higher-order kurtoses in terms of $\gamma_4^{(q)}$,

$$\gamma_r^{(q)} = F_r\,[\gamma_4^{(q)}]^{\frac{r}{2}-1}, \tag{23}$$

where for the SNIG the constants $\{F_4, F_6, F_8, \ldots\}$ are $\left\{1, 5, \frac{175}{3}, \ldots\right\}$. While true for the SNIG case at hand, relation (23) is true for all n-divisible distributions as shown below.

In Fig. 1, we show the percentage deviations of the Edgeworth-truncated approximations (21), (22) from their respective exact SNIG values (15) and (17) as a function of $\gamma_4^{(q)}$. We find a dramatic improvement in accuracy over the GGC ordering. The GEW approximation also continues improving as terms of higher order in w are included.

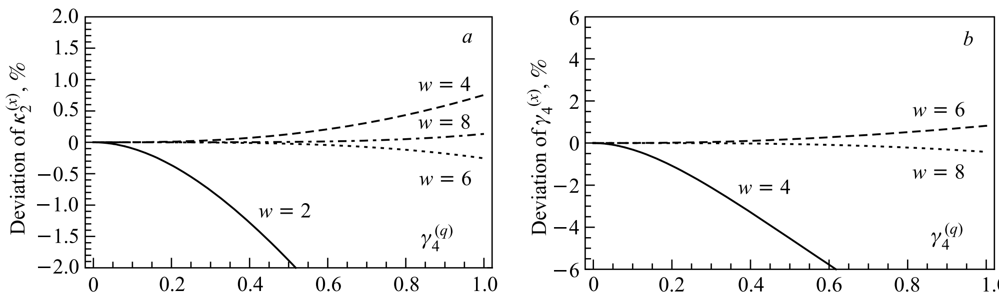

Fig. 1. Percentage deviations of GEW approximations of $\kappa_2^{(x)}$ and of $\gamma_4^{(x)}$ as a function of the measured q-kurtosis $\gamma_4^{(q)}$ for various Edgeworth orders w

The GEW truncated series also yields approximations very close to the original SNIG $f(q')$. Figure 2 shows that the first $O(\ell^2)$ correction term of Eq. (20) suffices for $\gamma_4^{(q)} = 1.0$, while for $\gamma_4^{(q)} = 1.5$, inclusion of the second $O(\ell^4)$ term is enough [5].

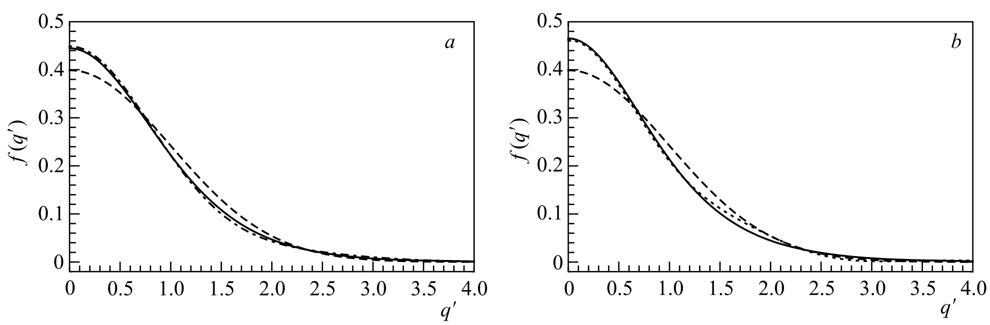

Fig. 2. Comparison of non-Gaussian $f(q')$ of (14) with Edgeworth approximations (20). Solid line — non-Gaussian SNIG $f(q')$ with $\gamma_4^{(q)} = 1.0$ (*a*) and 1.5 (*b*); dashed line — Gaussian reference pdf $f_0(q')$; dash-dotted line in panel *a* — $O(\ell^2)$ approximation; dotted line in panel *b* — $O(\ell^4)$ approximation

Clearly, the convolution with parameter $n = \ell^{-2}$ is playing a crucial role in the success of the GEW. This convolution is not, however, a physical effect but rather a statistical improvement: In statistics terms, any deviation from Gaussian is captured in the *rate of approach* of higher-order cumulants to zero as n increases: many proofs of the Central Limit Theorem rely on the fact that cumulants of the sum of n independent random variables obey $\kappa_r(n) = n\kappa_r(n{=}1)$ and that the rate of approach to zero of generalised kurtoses is therefore

$\gamma_r(n) = \gamma_r(1)/n^{\frac{r}{2}-1}$, from which it follows immediately that

$$F_r = \frac{\gamma_r(n)}{(\gamma_4(n))^{\frac{r}{2}-1}} = \frac{\gamma_r(1)}{(\gamma_4(1))^{\frac{r}{2}-1}} \tag{24}$$

is indeed constant in n for any n-divisible $f(q')$. The SNIG happens to be n-divisible, which explains *post factum* the simplicity of formulae (21)–(23), but it is by no means unique in satisfying this property.

The success of GEW ordering is therefore based on the fact that all contributing terms in a given order of $\ell = n^{-1/2}$ have the same rate of convergence to the Gaussian limit. Furthermore, due to the alternating sign of $H_{2r}(0) = (-1)^r(2r-1)!!$, the sum of contributions within a given $O(\ell^w)$ term tends to be substantially smaller than the individual contributions; for example, the $O(\ell^4)$ term for the SNIG test case is made up of $(1/2)\zeta_4^2 H_8(0) = 0.091\gamma_4^2$ and $\zeta_6 F_6 H_6(0) = -0.104\gamma_4^2$, adding up to $-0.013\gamma_4^2$.

Equation (24) shows that it is not necessary to know the value of n to make use of the GEW ordering: once the ordering has been established, we can set $n = 1$ and use the experimental q-cumulants in their GEW ordering independently of n. It is not even necessary to require n-divisibility as such: For the GEW ordering to work, we require only that $f(q')$ is reasonably close to a Gaussian as quantified by the errors shown in Fig. 2. The derivation does not rely on a particular form of $f_1(q_i)$ other than requiring existence of its cumulants, or on the size or even existence of a convolution.

The present one-dimensional calculation cannot, of course, be applied immediately to experimental data, but is meant to show that even on the fundamental level of expansions, there are important questions which must be addressed first. In sorting out the fundamental issue of re-ordering, the present results represent an important step towards a consistent statistical framework for shape description.

Application to experimental data will require generalisation to three dimensions using the existing 3D machinery of [3]. Furthermore, sampling fluctuations of experimental cumulants will have to be taken into account. Fortunately, experimental sample sizes are now large enough to warrant some optimism in this regard. In this connection, we also note that the GEW ordering has the additional advantage of placing terms with higher powers of $\gamma_4^{(q)}$ into low orders of w, making it unnecessary to measure higher-order kurtoses.

This work was supported by the National Research Foundation of South Africa.

REFERENCES

1. *Danielewicz P., Pratt S.* // Phys. Rev. C. 2007. V. 75. P. 034907.
2. *Eggers H. C., de Kock M. B., Schmiegel J.* arXiv:1011.3950[nucl-th]. 2010.
3. *Eggers H. C., Lipa P.* // Braz. J. Phys. 2007. V. 37(3a). P. 877.
4. *Barndorff-Nielsen O. E.* // Scand. J. Statist. 1997. V. 24. P. 1.
5. *De Kock M. B.* MSc Thesis. Univ. of Stellenbosch, 2009.