IL NUOVO CIMENTO Vol. 30 C, N. 4 Luglio-Agosto 2007
DOI 10.1393/ncc/i2007-10253-7

A new formulation of the Gram-Charlier method: Performance for fitting non-normal distribution

A. Pelliccioni and M. Angeloni(*)
ISPESL-DIPIA - via Fontana Candida 1, 00040 Monteporzio Catone (Rome), Italy

(ricevuto il 30 Luglio 2007; approvato il 2 Dicembre 2007; pubblicato online il 2 Febbraio 2008)

Summary. — The Gram-Charlier expansion was derived in an attempt to express non-normal densities as infinite series involving the normal density and its derivatives, using the moments data as input terms. In classic Gram-Charlier expansion the random variable is standardized, so that the Gaussian parameters are always fixed and referred to the mean equal to zero and to the standard deviation equal to one. This assumption seems to be too strong. An improvement of Gram-Charlier expansion was obtained by an optimization process, directed to choose new values of Gaussian parameters. In order to check the performance of the new approach, an estimate of the gamma probability density function was calculated. Two probability density functions, characterized by a different degree of skewness and kurtosis, were considered. The study has shown that in comparison with the classic assumption, the new one always gives the best results in terms of probability density function reproducibility and allows the best evaluation of the input moments. Further the comparison between estimated moments of order higher than the input ones and the theoretical moments shows a good reproduction. Finally the method seems to suggest that a less restrictive condition can be considered respect to the usual convergence criterium of the Gram-Charlier expansion.

PACS `02.30.Mv` – Approximations and expansions.
PACS `02.50.Cw` – Probability theory.
PACS `02.50.Ng` – Distribution theory and Monte Carlo studies.

1. – Introduction

In many applications it is often important to find a theoretical probability density function (p.d.f.) that fits non-Gaussian distributions through input data moments. Different methods can be considered: moment-generating function based on the Fourier transformation of p.d.f., whose coefficients are associated to the moments [1]; characteristic function theory, that involves the asymptotic expansion derived from the normal distribution (Edgeworth's development [2]); Gram-Charlier method that regards an expansion in terms of Hermite's polynomials [3]; multi-Gaussian development that deter-

(*) Visiting Scientist.

mines the p.d.f. as a linear combination of p Gaussian densities (very often $p = 2$), whose parameters are found through the solution algebric equations [4].

Sometimes, the above methods can produce multimodal p.d.f. and/or negative frequencies. In practice, while multimodal probability density function can take place, negative frequencies have to be avoided, because they have not statistical meaning. With regard to this, in a very interesting work, Barton and Dennis [5] considered the (β_1, β_2) space (where β_1 is the skewness and β_2 is the kurtosis) and determined the regions of positivity and unimodality both for the Gram-Charlier and the Edgeworth expansions. Their work shows that these expansions fit distributions which are at most mildly nonnormal (in terms of skewness and kurtosis).

In the paper Jondeau and Rockinger [6] specialized the method advocated by Barton to characterize the boundary delimiting the domain in the skewness-kurtosis space over which the expansion is positive. The authors present a mapping which transforms the constrained estimation problem into an unconstrained one.

A new theoretical approach of the Gram-Charlier expansion was proposed in 1997 [7] in order to improve the original formulation [3, 8, 9].

The new formulation was applied in studies of statistical fluids properties [10] to reproduce the vertical velocity p.d.f. using experimental moments up to 3rd order as input data. In fact, generally, only the first three moments are used because the sampling variances of moments higher than the third are large. The new model worked very well to minimizing negative frequencies and reproducing input moments.

Other authors [11] applied the new formulation of Gram-Charlier in order to develop a method for blind source separation of multi-microphone signals. These authors proposed a multi-microphone model based on a nonlinear mapping system and natural phenomena. In the paper, the proposed nonlinear algorithm is a generalization of serial gradient algorithm, cross-correlations, and Gram-Charlier series, which was extended to deal with nonlinear mapping and to be able to adapt to the actual statistical distributions of the sources by the output signals.

In this work, in order to check and control the performances of the new formulation of Gram-Charlier, we confine the attention to theoretical moments deriving from a well-known distribution without data noise.

Therefore, it is possible to test the model performances in reproducing the input moments and in estimating moments of higher order than the input ones. Having a theoretical p.d.f., it is possible to compare estimated frequencies with the theoretical ones.

Different theoretical p.d.f. could be chosen. The Barton and Dennis study [5] showed that the gamma distribution cannot be approximated by the classic Gram-Charlier expansion and that it has the worst fitting if compared with the other kind of distributions they examined (Lognormal and Pearson type V). For this reason two gamma distributions are in this study used to test the method, in order to verify improvements deriving from the new formulation.

2. – Mathematical background

All proofs of the calculations are reported in the work [7].

Gram-Charlier's type-A expansion (GCe) is one of the methods to approximate a given distribution [8]. In this expansion input data are the moments up to order (k) and the expansion gives the p.d.f. for the continuous random variable x.

In the univariate case of GCe, the p.d.f. $F(x)$ is evaluated using a truncated expansion in terms of Hermite's polynomials ($H_n(x;m;\sigma)$):

$$F_{m,\sigma}(x) = \alpha_{m,\sigma}(x) \sum_{n=1}^{k} C_n\big(\mu^k;m;\sigma\big) \cdot H_n(x;m;\sigma) \tag{2.1}$$
$$= \sum_{n=1}^{k} (-1)^n C_n\big(\mu^k;m;\sigma\big) \cdot D_x^n \alpha_{m,\sigma}(x),$$

where ${D_x}^n = \mathrm{d}^n/\mathrm{d}x^n$, $C_n(\mu^k;m,\sigma)$ are the coefficients and $\alpha_{m,\sigma}(x)$ is the Gaussian distribution having m as mean and σ as standard deviation:

$$\alpha_{m,\sigma}(x) = \frac{\exp\big[-(x-m)^2/2\sigma^2\big]}{\sigma\sqrt{2\pi}}\,.$$

μ^k is the k-order input moment defined as

$$\mu_x^k = \int_{-\infty}^{+\infty} x^k F(x)\mathrm{d}x.$$

The right side of (2.1), where the Gaussian derivatives are present, comes out from the Hermite polynomials definition.

In the classic case [8], Hermite's terms and C_n coefficents always refer to the standardized variable z, so that in (2.1) the random variable z and the parameters $m=0$ and $\sigma=1$ are chosen in order to determine C_n and H_n.

Below, are presented Hermite's polynomials and the C_n coefficients determination as calculated in the new formulation (see appendix A) and B) in Pelliccioni [7]).

– *Hermite's polynomials determination.* Hermite's polynomials can be calculated with two equivalent methods: iteration rule and Rogrigues formula. The following formula was obtained for Hermite polynomial:

$$H_{2n+\delta_d} = \frac{1}{\sigma^{2(2n+\delta_d)}} \sum_{z=0}^{n} (-1)^z (x-m)^{2n+\delta_d-2z} \frac{(2n+\delta_d)!\sigma^{2n}}{(2n+\delta_d-2z)!2^z z!}\,, \tag{2.2}$$

where $n = 0,1,2,\dots.$ and δ_d is

$$\delta_d = \begin{cases} 1 & \text{if } (2n+\delta_d) \text{ is odd,} \\ 0 & \text{if } (2n+\delta_d) \text{ is even.} \end{cases}$$

– *C_n coefficients determination.* The C_n coefficients can be determined through orthogonal rules applied to (2.1), taking into account (2.2):

$$C(2n+\delta_d) = \sum_{\alpha=0}^{n} \frac{\sigma^{2\alpha}}{2^\alpha \alpha!} \sum_{\beta=0}^{(2n+\delta_d-2\alpha)} \frac{(-1)^{2n+\delta_d-a-b}}{(2n+\delta_d-2\alpha-\beta)!\beta!} m^{2n+\delta_d-2\alpha-\beta}\mu^\beta. \tag{2.3}$$

Note that C_n depends on the mean, m, and the standard deviation, σ, so that this formula is more general than the classic one [8], that is calculated with $m = 0$, $\sigma = 1$.

– *Moment transformation rules.* As Gram-Charlier expansion refers to variable z, it is better to use moments related to $f(z)$ as input values (standardized moments).

The density of a random variable z, $F(z)$, is obtained from the density of x, $F(x)$, using the fact that

$$F(x)\mathrm{d}x = F(z)\mathrm{d}z;$$

it follows that the p.d.f. transformation from $F(x)$ to $F(z)$ is

$$F(z) = F(x) \Big/ \left|\frac{\mathrm{d}z}{\mathrm{d}x}\right|. \tag{2.4}$$

Using the definition of the k-order moment of $F(z)$, ${\mu_z}^k$, and taking into account (2.4), the following formula is obtained:

$$\mu_z^k = \frac{1}{\sigma^k}\sum_{\nu=0}^{k}(-m)^{k-\nu}\binom{k}{k-\nu}\mu_x^\nu; \tag{2.5}$$

vice versa it is possible to obtain the ${\mu_x}^k$ from the ${\mu_z}^k$ by

$$\mu_x^k = \sum_{\nu'=0}^{k}\sigma^{\nu'}m^{k-\nu'}\binom{k}{k-\nu'}\mu_z^{\nu'}.$$

– *Convergence of Gram-Charlier expansion.* The mathematical problem of the convergence of Gram-Charlier expansion was fronted by Cramér [8]. He showed that if $f(z)$ has bounded variation in $(-\infty, +\infty)$ and the following integral is convergent,

$$\int_{-\infty}^{\infty} \exp\left[\frac{z^2}{4}\right] F(z)\mathrm{d}z, \tag{2.6}$$

then the Gram-Charlier expansion converges to $F(z)$ at every continuity point.

In order to check the convergence, we express the above integral in terms of the input moments. With Taylor's expansion of the exponential term around $z = 0$, the convergence criterion is transformed by using the series of the standardized moments (2.5) as input.

$$\int_{-\infty}^{\infty} \exp\left[\frac{z^2}{4}\right] F(z)\mathrm{d}z = \sum_{n=0}^{\infty}\frac{1}{n!}\left(\frac{1}{4}\right)^n \mu_z^{2n} = \sum_{n=0}^{\infty} v_z^n, \tag{2.7}$$

where

$$v_z^n = \frac{1}{n!}\left(\frac{1}{4}\right)^n \mu_z^{2n} \tag{2.8}$$

is the n-th term.

As evident, the series has all positive terms (${\mu_z}^{2n} \geq 0$), thus convergence can be studied by using Cauchy's and D'Alambert's criterium.

It can be noted that the integral (2.6) is referred to the standardized z variable. Sometimes, referring to x random variable can be useful and the convergence criterium is obtained substituting (2.4) in (2.6):

$$\int_{-\infty}^{\infty} \exp\left[\frac{1}{4}\left(\frac{x-m}{\sigma}\right)^2\right] F(x)\mathrm{d}x. \tag{2.9}$$

With Taylor's expansion of the exponential term around $x = m$, the convergence criterium is obtained by using the series of the input non-standardized moments:

$$\int_{-\infty}^{\infty} \exp\left[\frac{1}{4}\left(\frac{x-m}{\sigma}\right)^2\right] F(x)\mathrm{d}x = \sum_{n=0}^{\infty}\frac{1}{n!}\left(\frac{1}{4}\right)^n \sum_{\nu=0}^{2n}(-m)^{2n-\nu}\binom{2n}{2n-\nu}\mu_x^\nu = \sum_{n=0}^{\infty} v_x^n. \tag{2.10}$$

The n-th term of the series is

$$v_x^n = \frac{1}{n!}\left(\frac{1}{4}\right)^n \sum_{\nu=0}^{2n}(-m)^{2n-\nu}\binom{2n}{2n-\nu}\mu_x^\nu. \tag{2.11}$$

Note that the ${\nu_x}^n$ term could be no positive.

3. – Brief description of GC improvement

The basic assumption in order to improve the GCe [7] consisted into considering the dispersion (σ) and the mean (m) in (2.1), not necessarily equal to the ones referring to the reproducing distribution. In other words, the classic expansion forces the parameters of the Gaussian one to be the same as the experimental ones. Such an assumption could be too strong, especially when experimental data come from no-Gaussian distribution. In fact, if the initial distribution is different from the Gaussian one, the best σ and m values in GCe cannot coincide with the experimental ones. According to this observation, new values of the standard deviation and the mean are to be searched. The criterium used to find them is the minimization of three guide functions [7], defined as

– *Moments error function*

The first function uses an estimate of moments derived by the Gram-Charlier distribution (2.1):

$$\mu_0^k(\sigma) = \int_{-\infty}^{+\infty} F(x;\sigma)x^k\mathrm{d}x. \tag{3.1}$$

From this expression it is possible to determine a m and σ function defined as

$$S(\sigma) = \frac{1}{n}\sum_{k=1}^{n} \frac{\left|\mu_{inp}^{k} - \mu_{0}^{k}\right|}{\mu_{inp}^{k}} 100 \tag{3.2}$$

which gives the error between the estimated ${\mu_0}^k(m,\sigma)$ and input moments ${\mu_{inp}}^k$. It also expresses the average of absolute relative error in percentage.

New m and σ values come from the minimization of the above function.

The minimization of the $S(m,\sigma)$ function is related to the distribution reproducibility, because it regards the optimization of two moments data sets: the first set is the input one and the second is calculated from the Gram-Charlier expansion.

– *Minimization of negative frequencies*

The $F(x)$ distribution may be positive or negative. The negative case happens in consequence of some of the x values, for which the following product is negative:

$$C_n\left(\mu^k; m; \sigma\right) \cdot H_n(x; m; \sigma) < 0. \tag{3.3}$$

Generally these x values are not considered in the Gram-Charlier distribution and the relative negative probabilities are settled to zero.

To weigh the total negative contributions coming from the expansion (2.1), the following function of m and σ is defined:

$$A(\sigma) = \frac{|A(-)|}{A(+) + |A(-)|} 100, \tag{3.4}$$

where $A(-)$ is the sum of negative frequencies deriving from GCe, and the denominator $A(+) + |A(-)|$ is the p.d.f. total area. The function $A(m,\sigma)$ gives the relative negative probability of GCe in percentage and is linked to an estimate of the negative values existing in the p.d.f.

– *Moments error and negative frequencies minimization*

The third m and σ function is defined as the product of the above functions:

$$P(\sigma) = S(\sigma) \cdot A(\sigma). \tag{3.5}$$

$P(m,\sigma)$ is related both to the input moments reproducibility and to the negative values minimization in the p.d.f. When the $A(m,\sigma)$ function is zero, we think it reasonable to consider the second minimum for $P(m,\sigma)$.

The study of the above functions is an instrument to find a pair of m and σ. Other criteria could be considered (for example, the minimization of the distance between the theoretical and reproduced p.d.f., the minimization of the maximum distances between cumulative distributions, and so on), but we believe that the above criteria are the most important in the context of probability theory, because they reproduce the input moments in the best way and, at the same time, reduce the weight of negative frequencies.

The search for minimum values is not realizable by analytical solutions because solvable algebric equations are not given. Finally four values are found: three are from the minimization process and one is from the classic case ($m = 0$, $\sigma = 1$).

4. – Application

To evaluate the performance of the new formulation, it is to be dealt with no-Gaussian distribution. Gamma function is chosen in consequence of considerations given in the introduction (see Barton *et al.* [5]). A random variable, x, following a gamma distribution, has a probability density given by

$$G(x;p,\lambda) = \frac{(\lambda)^p}{\Gamma(p)} x^{p-1} \exp[-\lambda x] \qquad \begin{cases} p > 0, \\ \lambda > 0, \\ 0 < x < \infty, \end{cases} \tag{4.1}$$

where $\Gamma(p)$ is the gamma function; p and λ are the shape and the scale parameters, respectively.

The k-order moment for the $G(x;p,\lambda)$ is given by the iteration rule

$$_{\Gamma}\mu_x^k = {}_{\Gamma}\mu_x^{k-1} \frac{(p+k-1)}{\lambda} \tag{4.2}$$

for $k = 1, 2, 3, \ldots$, with $_{\Gamma}\mu_x^0 = 1$.

From (4.2) it follows that the mean of the distribution is p/λ and the variance is $p/\lambda 2$.

The k-order standardized moment for the $G(x,p,\lambda)$ is obtained by the transformation rule (2.5).

According to the value of p and λ the distribution is more or less non-normal (in terms of skewness and kurtosis). In order to analyze the efficiency of the proposed methodology to fitting non-normal theoretical p.d.f., two different values of gamma parameters (p and λ) are taken into consideration. In both cases measures of skewness, $\sqrt{\beta_1} = m_3/m_2^{3/2}$, and kurtosis, $\beta_2 = m_4/m_2^2$ (where m_k is the k-th central moment) are computed (table I). Considering that the normal values of $\sqrt{\beta_1}$ and β_2 are 0 and 3, respectively, we can say that case I is related to a mildly non-normal distribution and case II to a much more non-normal one.

In order to verify the model performance in the above cases we have to lead the two distributions to the same mean ($m = 0$) and variance ($\sigma = 1$). In this way moments up to the 10th order have been calculated according to (2.5) and (4.2). For each case table II shows both standardized and non-standardized moments. The first ones are the data set given as input in the expansion.

Convergence of the series of standardized moments has to be studied before calculating GCe. Table III shows the first 10 terms of the series and the convergence coefficents (2.8), obtained by applying the Cauchy and d'Alambert criteria.

The convergence is linked to the integral (2.6) and, although the two series do not seem to converge, the GCe is calculated as well.

TABLE I. – *Gamma parameters, skewness and kurtosis.*

	p	λ	Skewness ($\sqrt{\beta_1}$)	Kurtosis (β_2)
Case I	14.1	9.2	0.533	3.426
Case II	7.2	1	0.745	3.833

TABLE II. – *Standardized and non-standardized moments of $G(x;p,\lambda)$.*

k	Case I		Case II	
	${}_{\Gamma}\mu_z^k$	${}_{\Gamma}\mu_x^k$	${}_{\Gamma}\mu_z^k$	${}_{\Gamma}\mu_x^k$
0	1	1	1	1
1	0	7.20E+00	0	1.53E+00
2	1	5.90E+01	1	2.52E+00
3	7.45E-01	5.43E+02	5.33E-01	4.40E+00
4	3.83E+00	5.54E+03	3.43E+00	8.18E+00
5	8.70E+00	6.21E+04	5.78E+00	1.61E+01
6	3.54E+01	7.57E+05	2.48E+01	3.34E+01
7	1.31E+02	9.99E+06	7.43E+01	7.30E+01
8	5.90E+02	1.42E+08	3.12E+02	1.67E+02
9	2.81E+03	2.16E+09	1.26E+03	4.02E+02
10	1.47E+04	3.49E+10	5.83E+03	1.01E+03

The first step is to find three mean and standard deviation values that minimize the guide functions $S(m,\sigma)$, $A(m,\sigma)$, $P(m,\sigma)$. As both distributions to be reproduced are standardized, the new m and σ values are searched around 0 and 1, respectively. In particular, see table IV for the ranges considered.

TABLE III. – *Study of convergence for case I and case II.*

	Case I			Case II		
n	${u_z}^n$	Cauchy ($\sqrt[n]{u_z^n}$)	D'Alambert (u_z^{n+1}/u_z^n)	${u_z}^n$	Cauchy ($\sqrt[n]{u_z^n}$)	D'Alambert (u_z^{n+1}/u_z^n)
1	0.250	0.250	-	0.250	0.250	-
2	0.107	0.327	0.428	0.120	0.346	0.479
3	0.065	0.401	0.604	0.092	0.452	0.769
4	0.051	0.475	0.786	0.096	0.557	1.043
5	0.047	0.544	0.934	0.120	0.654	1.248
6	0.050	0.606	1.047	0.170	0.745	1.421
7	0.057	0.664	1.144	0.269	0.829	1.581
8	0.070	0.717	1.234	0.468	0.909	1.737
9	0.093	0.768	1.320	0.884	0.986	1.889
10	0.130	0.816	1.404	1.803	1.061	2.040

TABLE IV. – *Typical m and σ ranges used for the simulation.*

	m			σ		
	min	max	step	min	max	step
Case I	−0.5	1.0	0.1	0.1	1.2	0.5
Case II	−0.5	1.0	0.1	0.5	2.5	0.5

In this way the local minimum for every guide function and for the two cases is found with 440 (40×11) Gram-Charlier calculations. As an example fig. 1 shows the trends of the three guide functions *vs.* σ values, when $m = 0.3$ in case II.

The results of minimization process of the three guide functions $S(m, \sigma)$, $A(m, \sigma)$, $P(m, \sigma)$ are shown in table IV.

Referring to $A(m, \sigma)$ function, when the theoretical p.d.f. is as like as the Gaussian one (case I), negative frequencies are not present and the best choice coincides with the classic choice ($m = 0$, $\sigma = 1$).

As soon as the theoretical p.d.f. is different from the normal one (case II), the $A(m, \sigma)$ function never goes to zero, but the minimum contribution to negative probabilities comes at $m = 0.5$ and $\sigma = 1.05$. Furthermore in both cases $S(m, \sigma)$ and $P(m, \sigma)$ functions have a minimum for $m \neq 0$ and $\sigma \neq 1$.

This means that the choice corresponding to classic GCe is the worst one, since even if the $G(x; p, \lambda)$ functions are not so much different from the Gaussian shape, the optimization shows that the best m and σ values are not equal to the theoretical distributions one.

– *The GC model performance: moments reproducibility*

In order to estimate the model performance, the following relative error function was defined for moments of each order:

$$\mathrm{RE}(k) = \left| \frac{\mu^k_{\mathrm{th}} - \mu^k_{\mathrm{rep}}}{\mu^k_{\mathrm{th}}} \right| 100, \tag{4.3}$$

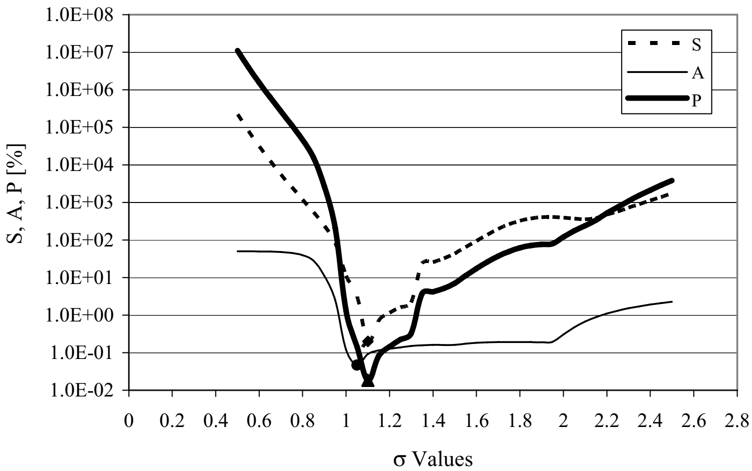

Fig. 1. – Trend of the three guide functions against σ values ($m = 0.3$—case II).

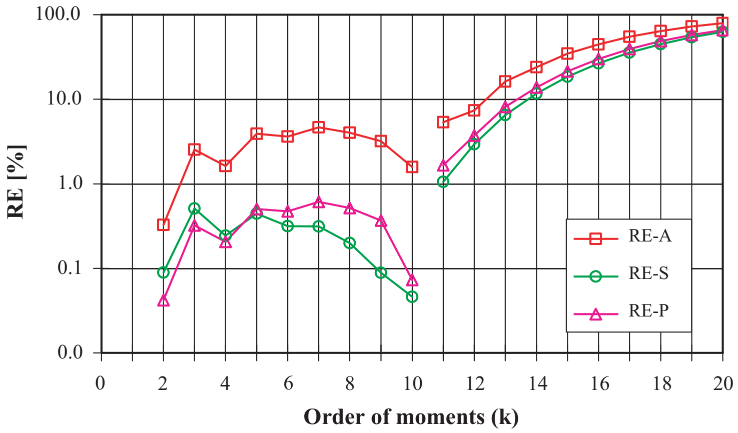

Fig. 2. – Reproducibility of moments: case I.

where ${\mu_{\mathrm{th}}}^k$ is the k-order gamma moment and ${\mu_{\mathrm{rep}}}^k$ is the reproduced one. As we deal with a theoretical p.d.f., the moments of every order are available. According to that it is possible to evaluate the goodness of the different m and σ choices to reproduce both the input moments and higher ones. In particular gamma moments up to the 20th order are computed and the only first 10 moments are used as input data. Moments from 11th order up to 20th order are considered to check the model performance in reproducing moments of order higher than the input ones. In practise all available moments are used as input data, so that the study of moments reproducibility separately for input moments and moments of higher order than the input ones, has no meaning. Our aim is to check the performance of the new GC formulation in estimating moments of higher order than the available ones, in order to apply the method in all cases in which the use of the available moments (for example the ones greater than the 3rd order) is not suitable because the sampling variance is large.

In figs. 2 and 3 the RE(k) up to the 10th order (input moments) and from 11th order up to 20th order are given. The estimate is evaluated for the m and σ values resulting from the minimization of the guide functions (RE$_S$, RE$_P$, RE$_A$) and for the Gaussian choice (RE$_G$).

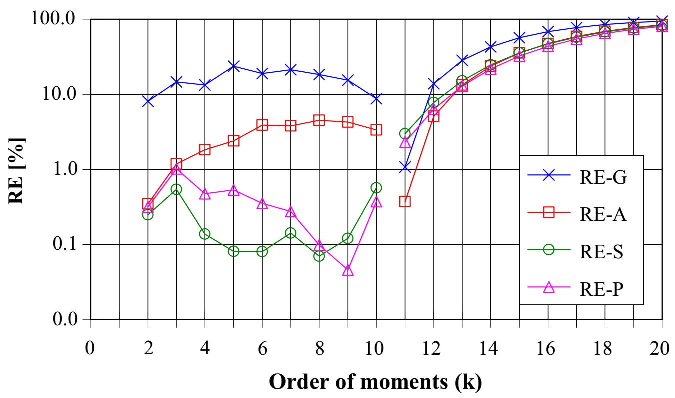

Fig. 3. – Reproducibility of moments: case II.

TABLE V. – *Guide functions values (A, S, P) for minimizating mean and standard deviation.*

Case I				
m	σ	$A(m,\sigma)$ (%)	$S(m,\sigma)$ (%)	$P(m,\sigma)$
0	1	0	0.255	0
0.2	1.05			0.311$^{(*)}$
0.3	1.05		0.225	
Case II				
m	σ	$A(m,\sigma)$ (%)	$S(m,\sigma)$ (%)	$P(m,\sigma)$
0	1	2.620	14.230	37.283
0.3	1.1		0.199	
0.4	1.1			0.016
0.5	1.05	0.013		

(*) Second minimum for $P(m,\sigma)$.

From the comparison between the reproduced and the theoretical moments, a bad fitting results when the order of moment is higher than the input one. As long as the order of the moment is equal to the input one, the reproducibility is very good (about 1% both for case I and II). When the orders increase, the reproducibility of the moments is worst (about 31.9% for case I and 40.7% for case II). However both input and higher-order moments are much more overestimated in the classic case, and the error is as much greater as the p.d.f. is different from the Gaussian one (table V).

– *The GC model performance: p.d.f. reproducibility*

In order to evaluate the accuracy of approximation to gamma distributions for each m and σ choice, the follow index based on a function of the difference between the gamma distribution function $G(x;p,\lambda)$ and the reproduced, $\hat{G}(x;p,\lambda)$ is computed:

$$M = \max_x \left|\hat{G}(x;p,\lambda) - G(x;p,\lambda)\right|. \tag{4.4}$$

For case I the interpolation coming from the classic choice, that coincides with the one coming from the minimization of $A(m,\sigma)$ function, is slightly worse than the fitting deriving from the minimization of $S(m,\sigma)$ and $P(m,\sigma)$ functions. While for case II the interpolation coming from the classic choice is quite worst than the fitting resulting from the minimization of $A(m,\sigma)$, $S(m,\sigma)$ and $P(m,\sigma)$ functions (tables VI and VII).

Figures 4 and 5 show the gamma functions and the interpolations both in case I and II (capital letters A, P, S stand for interpolation coming from the minimization of $A(m,\sigma)$, $S(m,\sigma)$ and $P(m,\sigma)$ functions, while G stands for interpolation coming from the classic choice).

TABLE VI. – *Reproducibility of the moments.*

	Case I		Case II	
	RE(k) up to the 10th order			
	mean	st. dev.	mean	st. dev.
RE_S	0.2	0.2	0.2	0.2
RE_A	2.5	1.5	2.6	1.6
RE_P	0.3	0.2	0.3	0.3
RE_G	-	-	14.2	6.7
	RE(k) from 11th order up to 20th order			
	mean	st. dev.	mean	st. dev.
RE_S	26.5	20.9	41.7	27.4
RE_A	40.3	25.5	41.4	28.9
RE_P	29.0	21.9	39.2	27.0
RE_G	-	-	55.7	31.2

TABLE VII. – *M values for each m and σ choices.*

	Case I	Case II
M_G	-	0.0643
M_A	0.0091	0.0117
M_S	0.0023	0.0070
M_P	0.0024	0.0065

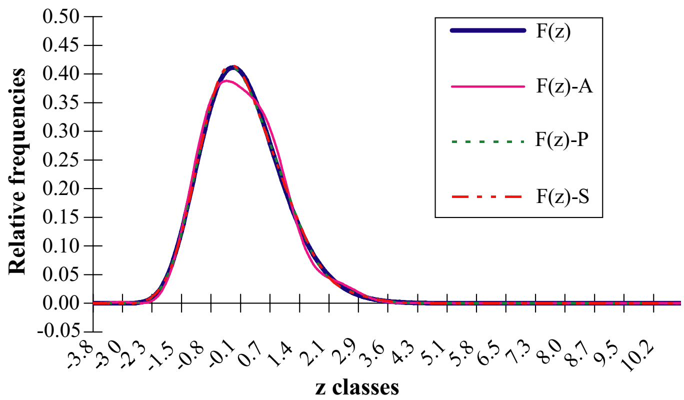

Fig. 4. – Theoretical and reproduced gamma p.d.f.: case I.

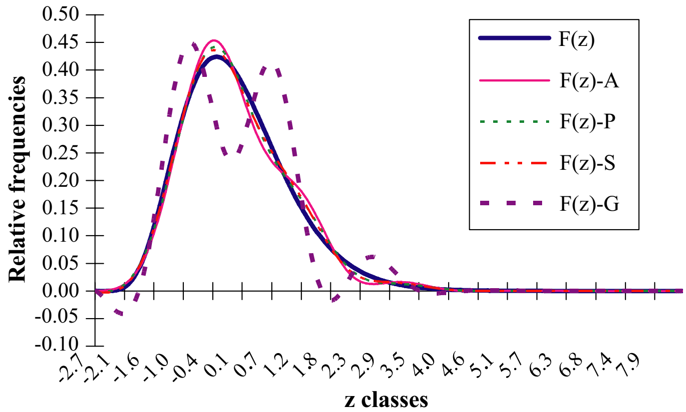

Fig. 5. – Theoretical and reproduced gamma p.d.f.: case II.

It is of interest to note that the convergence criterium given in the 2nd section seems to be too restrictive, when optimizing m and σ values are considered, in fact although the convergence criterium is not verified, the Gram-Charlier expansion converges to $f(z)$ in both cases.

5. – Conclusion

The Gram-Charlier expansion calculates the probability density function starting from the moments of any order. It concerns an infinite series involving the normal density and its derivatives. In practice, a finite number of terms of the series is taken to fit the p.d.f.

Classic approach considers the parameters of the Gaussian p.d.f. as fixed and in according to it no modulation is allowed.

Sometimes this assumption can be too strong and can bias the performance of the models in a bad way. Consequently the truncated series may be negative over certain intervals or may exhibit multimodality. In the past some authors [5] considered the (β_1, β_2) space (where β_1 is the skewness and β_2 is the kurtosis) and determined typical domains of positivity and unimodality.

The new formulation of the Gram-Charlier expansion allows to adapt the shape of the Gaussian distribution to the experimental one. In fact, when parameters are fixed, we force the Gaussian distribution to have the same mean and standard deviation of the experimental one. This assumption demonstrated to be too restrictive to fit distributions, expecially when they have no Gaussian shape. From this observation it derives the idea to choice mean and standard deviation values that are different from the classic choice. The new parameters are chosen to best reproduce the input moments and to minimize negative frequencies.

The Gram-Charlier performance clearly improves with the modulation of the Gaussian distribution.

The proposed approach has been tested by estimating a theoretical p.d.f., the gamma one.

To check better the reproducibility of the input distribution and the relative moments, two gamma p.d.f.'s, characterized by a different degree of skewness and kurtosis, were taken into consideration.

Having theoretical p.d.f. and, consequentely, theoretical moments, different checks can be performed. First, it is possible to verify exactly the capability to reproduce the distribution and the relative input moments of any order.

The results show that when the input distribution is as like as the Gaussian one, classic expansion gives good performance too. On the contrary, when skewed distributions are considered, results coming from the classic approach are generally worse than the ones coming from the new approach. The improvements rising from the new formulation seems to suggest that in this case, a less restrictive criterium than the one considered in the classic approach, could be considered to front the mathematical problem of the convergence of the Gram-Charlier expansion.

A second application rising from the knowledge of theoretical moments concerns the possibility to verify the accuracy of the estimated moments of higher order than that of input ones. Usually this aspect cannot be tested, because everybody uses all available experimental moments as input. In this case too, classic expansion always gives a worse performance than the new one.

REFERENCES

[1] Wiener N., *The Fourier Integral and Certain of Its Applications* (Cambridge) 1993.
[2] Edgeworth F. Y., *Proc. Cambridge Philos. Soc.*, **20** (1905) 36.
[3] Charlier C. V. L., *K. Fysiogr. Sällsk. Handl. B*, **16** (1906).
[4] Scarani C., *Statistica*, anno LI (1991) 4.
[5] Barton D. E. and Dennis K. E., *Biometrika*, **39** (1952) 425 (D).
[6] Jondeau E. and Rockinger M., *Estimating Gram-Charlier Expansions with Positivity Constraints*, Papers 56, Banque de France - Direction Generale des Etudes (1999).
[7] Pelliccioni A., *Nuovo Cimento C*, **20** (1997) 435.
[8] Cramer H., *Mathematical Methods of Statistics* (Princeton University Press) 1961.
[9] Kendall M. and Stuart A., *The Advanced Theory of Statistics* (MacMillan, New York) 1977.
[10] Pelliccioni A., Agnello P. and Angeloni M., *Optimization of the Probability Density Function with a New Formulation of the Gram-Charlier Method*, *5th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, 18-21 May, Rhodes, Greece*, preprints, edited by Bartzis J. G. and Konte K. (INTRP/Environmental Research Laboratory) 1998.
[11] Leong W. Y. and Homer J., *Implementing Nonlinear Algorithm in Multimicrophone Signal Processing*, *IEEE Workshop on Machine Learning for Signal Processing, 28-30 Sept. 2005* (2005), pp. 33-39, ISBN: 0-7803-9517-4.