# The Gram-Charlier method to evaluate the probability density function in monodimensional case (*)

A. PELLICCIONI
*ISPESL-DIPIA - Via Urbana 167, Roma, Italy*

(ricevuto il 10 Gennaio 1996; revisionato il 6 Settembre 1996; approvato l'8 Ottobre 1996)

**Summary.** — In many experimental applications, starting from a random variable, it is possible to evaluate the moments and to define the probability density function (PDF) in different ways. In this paper a new approach is shown in order to estimate the PDF by moments according to Gram-Charlier method (GCm). The approach consists of a choice of standard deviation ($\sigma_{new}$) in GCm which optimizes the values of the input moments. In particular three $\sigma_{new}$ are selected in order to minimize: 1) the sum of absolute relative deviations among theoretical and experimental moments; 2) the relative per cent of negative probabilities coming from GC expansion; 3) the product between the two previous functions. A theoretical application of the above approach is made where the input moments data set comes from the vertical velocity distribution estimated for one level of the convective mixed layer. This application consists of two different simulations. The first evaluates the moments up to 10th order, having as input data the moments up to 3rd order. The second gives the moments up to 10th order considering both the moments of the previous simulation and the 4th-order moment calculated with Gaussian closure as input data.

PACS 92.60.Fm – Boundary layer structure and processes.
PACS 47.27 – Turbulent flows, convection, and heat transfer.
PACS 01.30.Cc – Conference proceedings.

## 1. – Introduction

The purpose of this paper is to optimize the determination of the probability density function (PDF) by data moments.

In the probability theory an estimate of the characteristic function, related to Fourier transformation of PDF, comes from Mc Laurin's series development whose coefficients are related to the moments [1].

Another well-known method of the characteristic function theory is the asymptotic

(*) Paper presented at EUROMECH Colloquium 338 "Atmospheric Turbulence and Dispersion in Complex Terrain" and ERCOFTAC Workshop "Data on Turbulence and Dispersion in Complex Atmospheric Flows", Bologna, 4-7 September 1995.

expansion derived from the normal distribution. In this theory the estimate of the PDF is made in terms of the semi-invariant moments (called cumulants) and it is called Edgeworth's development [2].

The GCm [3] obtains the PDF as an orthogonal expansion derived from the normal distribution. The input data are the experimental moments.

An alternative method uses a bivariate Gaussian development to estimate the PDF by moments [4]. In this case the solution of an algebraic system gives the means and the standard deviations connected with Gaussian distribution. One observes that in [4] the PDF is always positive by definition, while this is not always verified in previous approaches [1-3].

The estimate of PDF is very important in many applications.

An example consists in the simulation of turbulent fluxes having statistical properties [5]. Other applications concern the computation of the probability density of the velocity components in the atmospheric turbulence [6]. Lagrangian models on turbulent diffusion use statistical information on the real turbulent flux to evaluate the trajectories of traces and their spatial concentration [7, 8].

Moreover, an interesting application is related to the determination of pollutant concentration by $K$-theory dynamical equations on moments [9]. The pollutant levels are obtained involving classic GCm and the moments coming from the solution equivalent $K$-equation.

The present paper illustrates the improvements of the classic GCm in order to estimate the PDF by moments data in univariate case.

## 2 – Mathematical background

A classical method to approximate a given distribution is Gram-Charlier's type-A (GC) expansion [10]. Input data are the moments up to order $k$ and the expansion gives the PDF for the continuous random variable $x$.

The general relations on univariate distribution are described below.

The PDF $F(x)$ is evaluated using a truncated expansion in terms of Hermite's polynomials $(H_n(x))$:

$$F(x) = \frac{\exp\left[-(x-m)^2/2\sigma^2\right]}{\sigma\sqrt{2\pi}} \sum_{n=1}^{k} C_n(\mu^k;\, m;\, \sigma)\cdot H_n(x;\, m;\, \sigma) = \tag{2.1}$$

$$= \alpha_{m,\sigma}(x) \sum_{n=1}^{k} C_n \cdot H_n(x)\,,$$

where $\alpha_{m,\sigma}(x)$ is the Gaussian distribution having $m$ as mean and $\sigma$ as standard deviation, and the terms $\mu^k$ are $k$-order input moments defined as

$$\mu^k = \int_{-\infty}^{+\infty} x^k F(x)\, \mathrm{d}x\,. \tag{2.2}$$

The $C_n$ coefficients are determined by the orthogonalization process applied to (2.1),

see appendix B:

$$C_n = \frac{\int F(x)\, H_n(x)\, \mathrm{d}x}{\int H_n^2(x)\, \alpha(x)\, \mathrm{d}x} = \frac{\sigma^{2n}}{n!} \int F(x)\, H_n(x)\, \mathrm{d}x \,. \tag{2.3}$$

In the classic case [10], Hermite's terms and the $C_n$ coefficients refer always to the standardized variable $Z$:

$$Z = \frac{x - m}{\sigma} \,, \tag{2.4}$$

so that in (2.1) the random variable $Z$ and the parameters $m = 0$ and $\sigma = 1$ in order to determine $C_n$ and $H_n$ are chosen (standard assumptions). If one considers $x$ as the distribution variable, the equivalent standard choice of the Gaussian parameters, $m$ and $\sigma$, is expressed by

$$\begin{cases} m = \mu^1, \\ \sigma = \sqrt{\mu^2 - (\mu^1)^2} \,. \end{cases} \tag{2.5}$$

In ref. [10] the GC calculation refers always to the standardization assumption so that the Gaussian parameters are fixed.

In order to generalize the GC expansion, other values of the Gaussian parameters are to be considered. This implies that Hermite's polynomials $H_n$ and the $C_n$ coefficients depend on $m$ and $\sigma$.

In the classic case the mean and the standard deviation are calculated from experimental moments by (2.5).

The GC expansion is convergent when the following integral is convergent [10]:

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

By Taylor's expansion of the exponential term, the convergence criterion is obtained by using the series of the experimental moments

$$\int_{-\infty}^{+\infty} \exp\left[\frac{x^2}{4}\right] F(x)\, \mathrm{d}x = \mu^0 + \frac{\mu^2}{4} + \frac{\mu^4}{32} + \frac{\mu^6}{384} + \frac{\mu^8}{6144} + \frac{\mu^{10}}{122880} + O(\mu^{11}) \,. \tag{2.7}$$

**2**˙1. *Hermite's polynomials determination.* – Hermite's polynomials are evaluated by two equivalent methods [11] (definition and theoretical background are given in appendix A). The first uses the iteration rules:

$$H_n(x) - \frac{(x - m)}{\sigma^2} H_{n-1}(x) + \frac{(n-1)}{\sigma^2} H_{n-2}(x) = 0 \,, \tag{2.8}$$

where the $H_n$ term is calculated from of $H_{n-1}$ and $H_{n-2}$ terms.

The second uses the characteristic polynomials defined as:

$$H_n(x) = \frac{(-1)^n}{\alpha_{m,\sigma}(x)} D_x^n \alpha_{m,\sigma}(x) . \tag{2.9}$$

From eqs. (2.8) or (2.9), the following Hermite's polynomials terms are obtained:

$$\begin{cases} H_0(x) = 1 , \\ H_1(x) = \dfrac{(x-m)}{\sigma^2} , \\ H_2(x) = \dfrac{(x-m)^2}{\sigma^4} - \dfrac{1}{\sigma^2} , \\ H_3(x) = \dfrac{(x-m)^3}{\sigma^8} - 3\dfrac{(x-m)}{\sigma^4} , \\ H_4(x) = \dfrac{(x-m)^4}{\sigma^8} - 6\dfrac{(x-m)^2}{\sigma^6} + \dfrac{3}{\sigma^4} , \\ H_5(x) = \dfrac{(x-m)}{\sigma^2}\left( \dfrac{(x-m)^4}{\sigma^8} - 10\dfrac{(x-m)^2}{\sigma^6} + \dfrac{15}{\sigma^4} \right) . \end{cases} \tag{2.10}$$

**2**'2. *$C_n$ coefficients determination.* – The $C_n$ coefficients are calculated substituting Hermite's polynomials in (2.3) and using the definition (2.2) of the moments. Performing the above calculations the following formula is derived (all steps are shown in appendix B):

$$C_{(2n+\delta_d)} = \sum_{\alpha=0}^{n} \frac{\sigma^{2\alpha}}{2^\alpha \alpha!} \sum_{\beta=0}^{2n+\delta_d-2\alpha} \frac{(-1)^{2n+\delta_d-\alpha-\beta}}{(2n+\delta_d-2\alpha-\beta)!\beta!} m^{2n+\delta_d-2\alpha-\beta} \mu^\beta , \tag{2.11}$$

where $n = 0, 1, 2, 3, \ldots$ and $\delta_d$ is defined as

$$\delta_d = \begin{cases} +1 & \text{if } (2n+\delta_d) \text{ is odd} , \\ 0 & \text{if } (2n+\delta_d) \text{ is even} . \end{cases} \tag{2.12}$$

It is to be noted that the $C_n$ coefficients (2.11) are more generally determined than the calculation in [10], (in which $m = 0$ and $\sigma = 1$) because the latter refers to any choice of $m$ and $\sigma$.

An alternative approach to calculate the $C_n$ coefficients makes use of iteration rules (2.8) to evaluate $H_n$ terms in (2.3). With this assumption, every $C_n$ coefficient is determined by $C_{n-1}$ and $C_{n-2}$. The only $C_0$ and $C_1$ terms must be directly calculated by (2.11).

However, using one of these approaches, the GC coefficients are obtained as a

function of the Gaussian parameters. The first six coefficients are

$$
(2.13)\quad \begin{cases}
C_0 = \mu^0 , \\
C_1 = \mu^1 - m\mu^0 , \\
C_2 = \dfrac{1}{2}(\mu^2 - 2\mu^1 m + \mu^0 m^2 - \sigma^2 \mu^0) , \\
C_3 = \dfrac{1}{3!}(\mu^3 - 3\mu^2 m + 3\mu^1 m^2 - \mu^0 m^3) - \dfrac{\sigma^2 \mu^1}{2} + \dfrac{\mu^0 \sigma^2 m}{2} , \\
C_4 = \dfrac{1}{4!}[\mu^4 - 4\mu^3 m + 6\mu^2 (m^2 - \sigma^2) - \\
\qquad\qquad - 4\mu^1 (m^3 - 3\sigma^2 m) + \mu^0 (m^4 + 3\sigma^4) - 6\sigma^2 m^2] , \\
C_5 = \dfrac{1}{5!}[\mu^5 - 5\mu^4 m + 10\mu^3 (m^2 - \sigma^2) - 10\mu^2 (m^3 - 3\sigma^2 m) + \\
\qquad\qquad + 5\mu^1 (m^4 - 6\sigma^2 m^2 + 3\sigma^4) - \mu^0 (m^5 - 10\sigma^2 m^3 + 15\sigma^4 m)] .
\end{cases}
$$

## 3. – GC expansion improvement

The basic idea to improve the GC expansion is to consider the standard variable (2.5) or dispersion in the distribution (2.1) as being not necessarily related to the standard variable choice. In fact, there is no *a priori* assumption of why the standard deviation of the Gaussian of GC must be related to experimental standard deviation (2.5). For example, this statement is evident in the bimodal PDF, in which the normalized $\sigma$ is always larger than the average amplitude around every modal value, so that the $\sigma$ of the Gaussian could be lower than the standard choice. The standard deviation is the best choice when the input moments come from a Gaussian distribution. Probably the best $\sigma$ value may not coincide with the standard if the starting distribution is different from that Gaussian PDF.

According to these considerations and having performed all the calculation with $m$ and $\sigma$ variable, a new value for the Gaussian of GC is obtained. This is calculated by the determination of three functions. For each function the minimum value is extracted and is taken into account for the best standard deviation in GCm.

The chosen reference functions are described below.

**3**˙1. *Moments error function.* – The first function uses an estimate of moments derived by the GC distribution (2.1)

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

where it is shown that such moments are $\sigma$-dependent. From the above expression it is

possible to determine a $\sigma$ function defined as

$$S(\sigma) = \frac{1}{n} \sum_{k=1}^{n} \frac{|\mu_{\text{inp}}^k - \mu_0^k|}{\mu_{\text{inp}}^k} 100 , \tag{3.2}$$

which gives the deviations between the estimated (3.1) and input moments $\mu_{\text{inp}}$, and expresses the average of Absolute Relative Error (ARE) in per cent.

A new $\sigma$ value ($\sigma_s$) comes from the minimization of the above equation.

The minimization of the $S(\sigma)$ function is related to the distribution reproducibility because it regards the optimization of two moments data sets: the first set is experimental and the second is calculated from GC distribution.

**3**˙2. *Negative PDF minimization terms.* – As known, the $F(x)$ distribution coming from GC is either positive or negative. The last case happens as a consequence of some $x_{\min}$ value for which the following product is negative:

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

These $x_{\min}$ values are not to be considered in GC distribution and then the related negative probabilities are to be set to zero.

To weight the total negative probabilities coming from expansion (2.1), the following $\sigma$ function is chosen:

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

where $A(-)$ is the negative probabilities sum deriving from GC expansion and the denominator $A(+) + |A(-)|$ is the PDF total area. The function $A(\sigma)$ expresses the Relative Percent of Negative probability (RPN) of GC expansion and gives an estimate of the negative values of the PDF. It is be to noted that $A(\sigma)$ has no correlation with the optimization of input moments.

**3**˙3. *Moments error and negative minimization area.* – The third $\sigma$ function is defined as

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

$P(\sigma)$ is related both to the input moments reproducibility and to the negative values minimization in the PDF.

The impact of $\sigma$ choices on the GC distribution are evaluated considering four values in the expansion, three coming from the minimization process and one coming from the classic case (standard choice (2.5)).

**3**˙4. *The* GC *model performance.* – In order to estimate the model performance, the following relative error function is defined:

$$\text{RE}(k) = \frac{\mu_{\text{inp}}^k - \mu_0^k}{\mu_{\text{inp}}^k} 100 . \tag{3.6}$$

This mathematical relation has the purpose to weight the goodness of the different $\sigma$ choices for the reproducibility of the input moments at any order $k$.

## 4. – Results and discussion

To evaluate the goodness of the proposed methodology, two simulations are performed.

Firstly, all moments up to 3rd order are included and secondly, the Gaussian closure is imposed on the 4th input order.

In both simulations the input moments up to the 10th order have been calculated from a given PDF.

In every simulation the output moments up to the 10th order are estimated starting from input moments of smaller orders. Moreover, the reproducibility calculation for every order is made by the relative error (3.6).

With reference to PDF, the velocity distribution is considered at the level $Z/Z_i = 0.25$ in the convective boundary layer [12]. The moments calculated by this PDF are

TABLE I. – *Moments obtained from the velocity distribution at one level in the convective boundary layer* [12].

| | | | |
|---|---|---|---|
| $\mu^0 = 1.000$ | $\mu^1 = -0.006$ | $\mu^2 = 0.396$ | $\mu^3 = 0.122$ |
| $\mu^4 = 0.387$ | $\mu^5 = 0.297$ | $\mu^6 = 0.591$ | $\mu^7 = 0.678$ |
| $\mu^8 = 1.668$ | $\mu^9 = 1.612$ | $\mu^{10} = 2.663$ | |

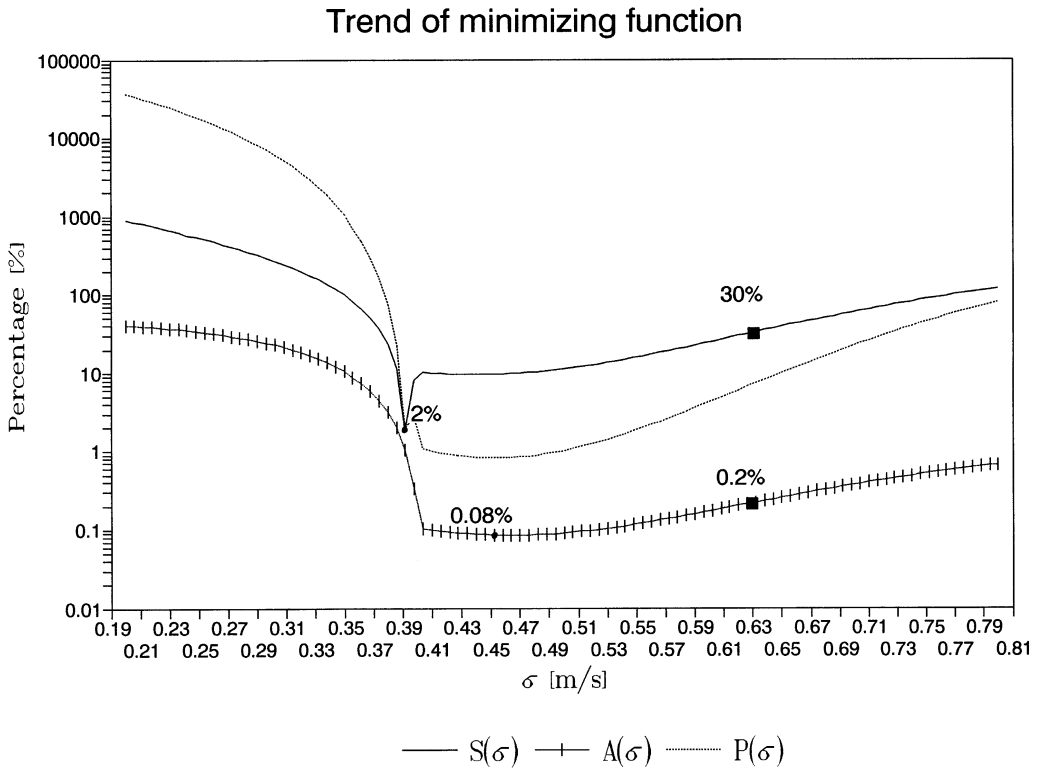


Fig. 1.

shown in table I. These values form the input data set that will be considered in the simulations.

Convergence of the series of moments must be verified before calculating the GC expansion. Substituting the theoretical input moments (table I) in eq. (2.7), the series converges as:

$$\int_{-\infty}^{+\infty} \exp[x^2/4]\, f(x)\, \mathrm{d}x = \mu^0 + \frac{\mu^2}{4} + \frac{\mu^4}{32} + \frac{\mu^6}{384} + \frac{\mu^8}{6144} + \frac{\mu^{10}}{122880} + o(\mu^{11}).$$

The standard Gaussian parameters are obtained from (2.5) with mean and variance values of table I:

$$\text{(4.1)} \qquad \begin{cases} m = \mu^1 = -0.006, \\ \sigma_{\text{Gauss}} = \sqrt{\mu^2 - (\mu^1)^2} = 0.63. \end{cases}$$

**4**·1. *First simulation: input data are all the moments up to the* 3rd *order*. – The input moments are:

$$\text{(4.2)} \qquad \mu^0 = 1.000, \quad \mu^1 = -0.006, \quad \mu^2 = 0.396, \quad \mu^3 = 0.122.$$

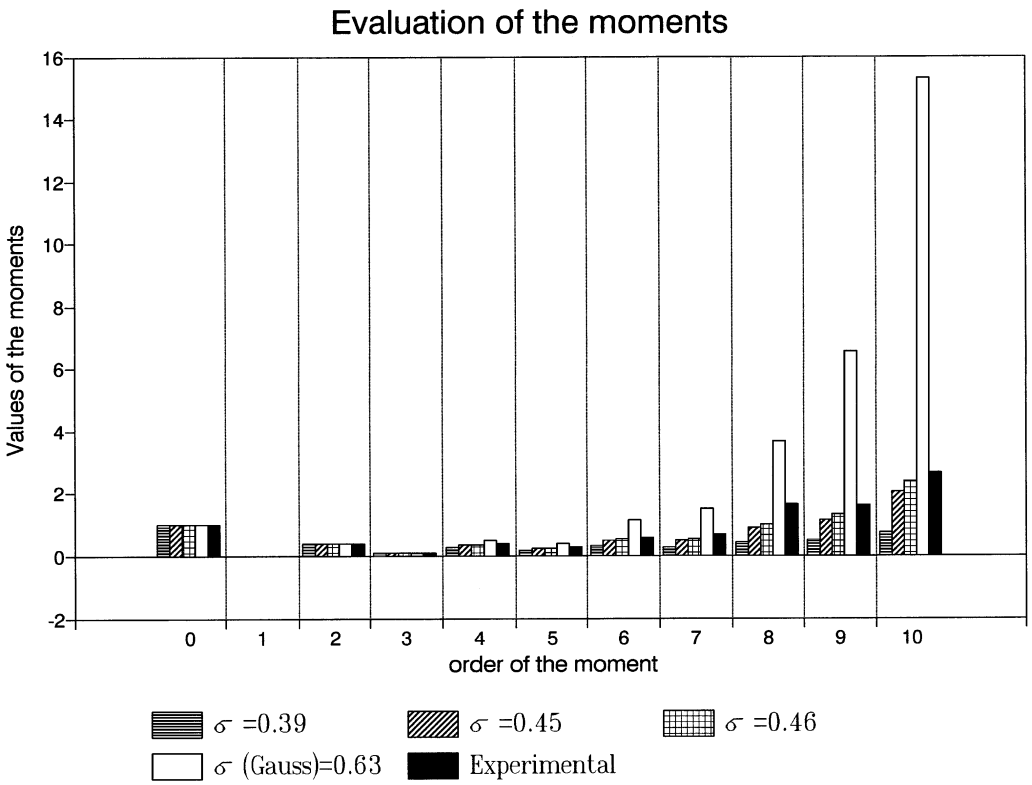


Fig. 2.

The first step is to find the three standard deviation values which minimize the reference functions $S(\sigma)$, $A(\sigma)$, $P(\sigma)$. The determination of minimum values (one for each function) is very important to obtain the best PDF associated with input moments.

The search for minimum values is not realizable by analytical solution because solvable algebric equations are not given.

The three reference functions, summarized in fig. 1, are obtained both by calculation of the GC expansion and the estimation (3.1) of the related moments. In fig. 1, there is a well-marked minimum of the $S(\sigma)$ function. This minimum is characterized by an averaged reproducibility value ($R$) of about 2% at $\sigma_S = 0.39$.

The $A(\sigma)$ function never goes to zero in the example and, therefore, some negative probabilities are always predicted from the GC expansion. The smaller contribution to negative probabilities comes at $\sigma_A = 0.45$ with the value 0.008% of the $A(\sigma)$ function.

The function product ($P(\sigma)$) has a minimum of $\sigma_P = 0.46$.

From fig. 1, another important result comes out. The choice (4.1) corresponding to classic GC expansion is not the best. In fact $\sigma_{\text{Gaus}} = 0.63$ gives both a moment reproducibility (30%) and an area value (0.2%) higher in comparison to $\sigma_S$, $\sigma_A$, $\sigma_P$ ones.

In fig. 2 the estimate (3.1) of moments up to the 10th order are given. In this figure, the moments coming from the input and from the classic choice of $\sigma$ are also shown.

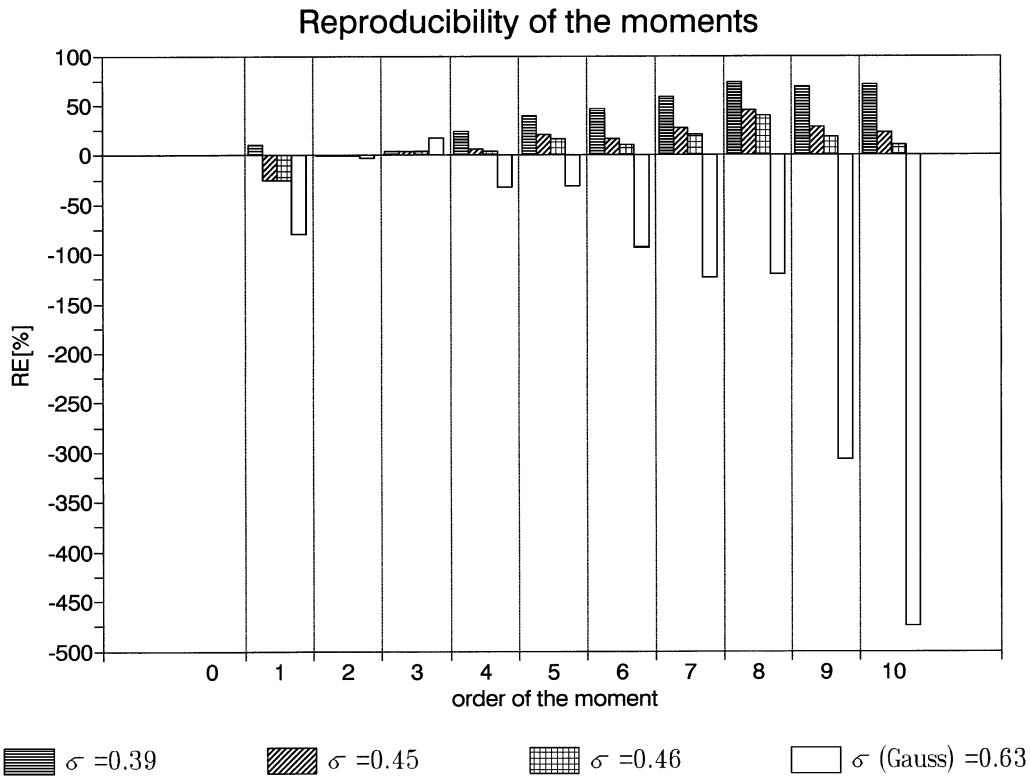


Fig. 3.

The estimate is evaluated both for the three minimizing standard deviation values ($\sigma_S$, $\sigma_A$, $\sigma_P$) and for the classic choice of ($\sigma_{\text{Gaus}}$). From the comparison between the reproduced and the experimental moments, a bad fitting results when the order moments are higher than the input ones.

This behaviour is evident in fig. 3, where the mean moments reproducibilities (RE ($k$)) up to the 10th order are given. As long as the orders of the moment are equal to the input ones, the reproducibility is not bad, but when the orders increase, bad reproducibility of the moments is evident. Higher-order moments are highly overestimated in the classic case ($\sigma_{\text{Gaus}}$), where RE ($k$) increases up to 470% to the 10th order. All of the above indicates that the classic Gram-Charlier method (referred always to standard variable) is not qualified to reproduce experimental PDF.

On the contrary, all three minimum $\sigma$ have moments' reproducibility ranging from 20% to 70%.

In this simulation the best $\sigma$, that comes out from the $P(\sigma)$ function, has an average of about 15% reproducibility.

The above simulation indicates that it is possible to improve the classic Gram-Charlier method by choosing, objectively, some standard deviation values not related to the standardization.

The results of the simulation are: *a*) the standardization parameters are not the best to reproduce the input moments; *b*) the best model performances are obtained

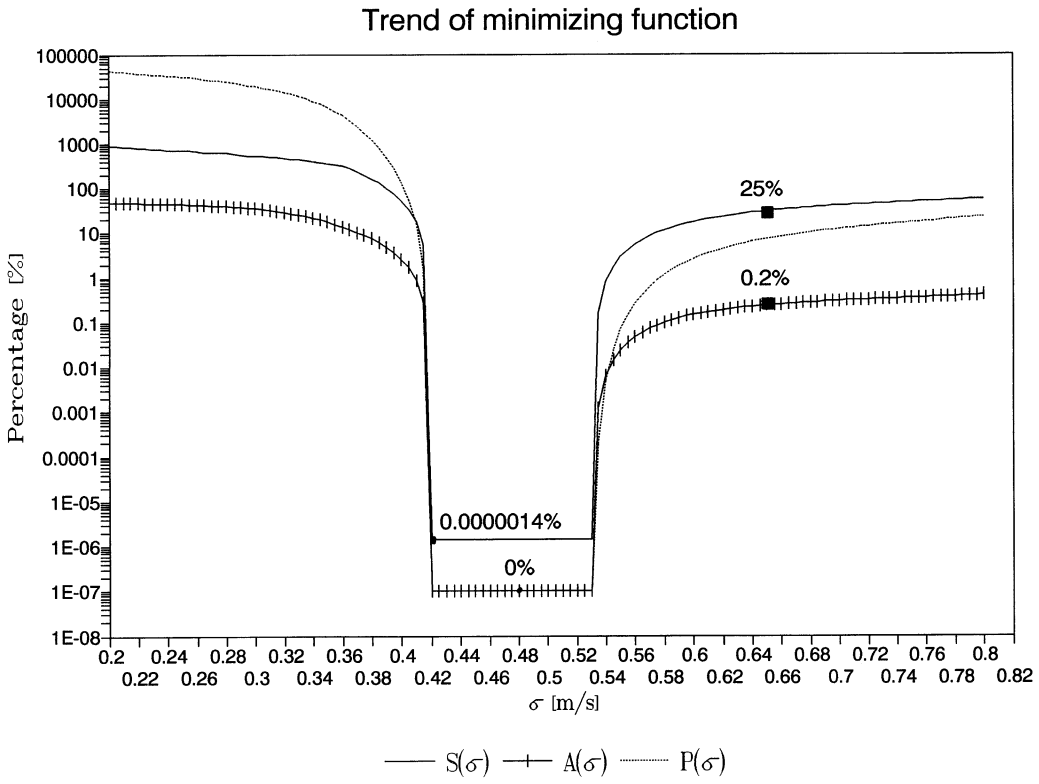


Fig. 4.

from the $\sigma$ minimization related to $S(\sigma)$ reproducibility, $A(\sigma)$ negative contribution, and $P(\sigma)$ function product.

**4**·2. *Second simulation: Gaussian closure up to* 4th *input order*. – This second simulation has both the purpose to evaluate the implications of Gaussian closure for input moments on the reproducibility of experimental data and to compare the GC model performance of the first simulation with the results. The Gaussian closure of the moments consists in the numerical estimation of higher-order moments in order to offset the experimental data scarsity [13].

This choice for the closure is very simple and was taken into account only to give an example of the possible utility of the expansion with the proposed improvement.

The moments in their even order for Gaussian closure are evaluated by the formula [14]:

$$\mu^{2n} = \frac{(2n-1)!}{2^{n-1}(n-1)!}(\mu^2)^n . \tag{4.3}$$

Applying the (4.3) estimate to the 4th-order term, a new set of input moments for the

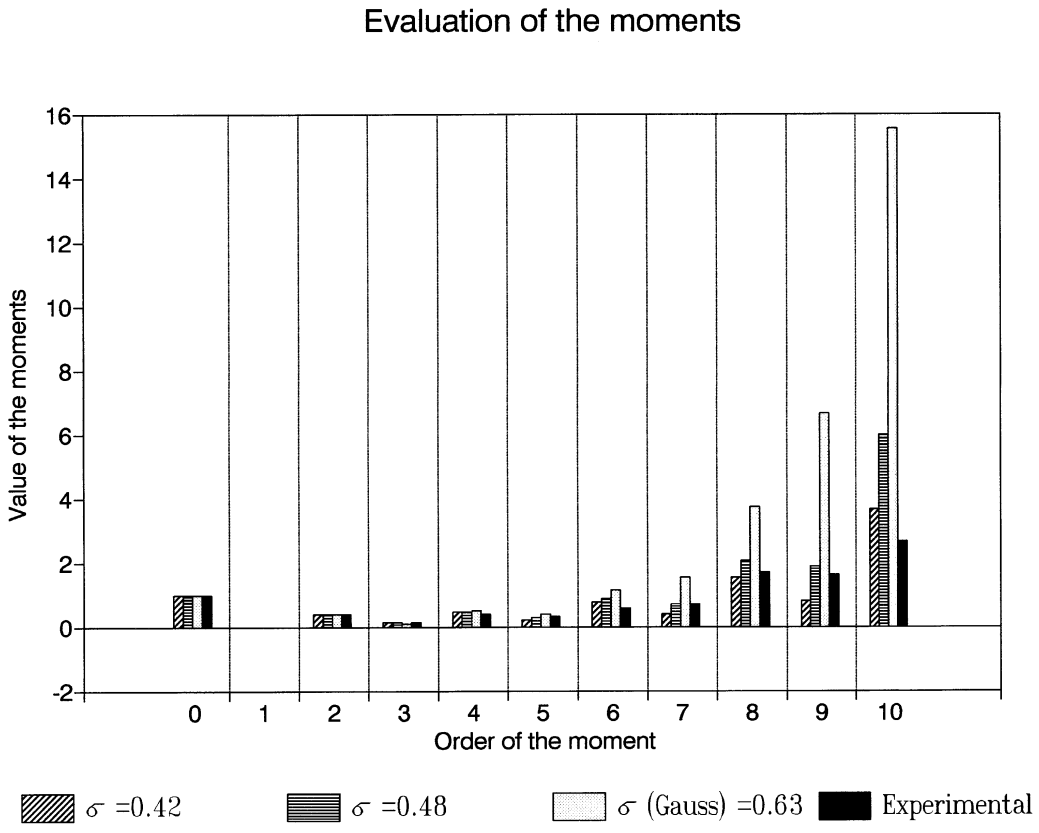


Fig. 5.

GC expansion is obtained

(4.4) $\mu^0 = 1.000\,, \quad \mu^1 = -0.006\,, \quad \mu^2 = 0.396\,, \quad \mu^3 = 0.122\,, \quad \mu^4 = 0.470\,.$

By these input moments the GC expansion is computed for any $\sigma$ in a range (0.20–0.80). In fig. 4 the $S(\sigma)$, $A(\sigma)$, $P(\sigma)$ functions are shown. The $S(\sigma)$ function (fig. 4) presents a broad minimum of $10^{-6}$% at $\sigma_S = 0.42$.

This small minimum value means that the Gaussian closure to the 4th order has improved the average reproducibility in comparison with the results in the 1st simulation (in which the minimum of $S(\sigma)$ was 2%).

From the $A(\sigma)$ trend, it can be noted that the negative contribution to probability is zero in a range of $\sigma$ (0.42–0.54). In this case the minimum is chosen to be the average of extreme terms, which is $\sigma_A = \sigma_P = 0.48$. Worthy of note is that the introduction of the 4th-order term in the input moments has deleted the small negative contribution found in the first simulation.

Moreover in this simulation the classic choice $\sigma_{\text{Gaus}}$ produces a very high value both for mean reproducibility (30%) and for negative area (0.2%).

In fig. 5 the reproduced moments up to the 10th order and the experimental ones are shown. The moments of the same order as in entry are well fitted with the input moments for all three $\sigma$ choices. The behaviour of higher-order moments is similar

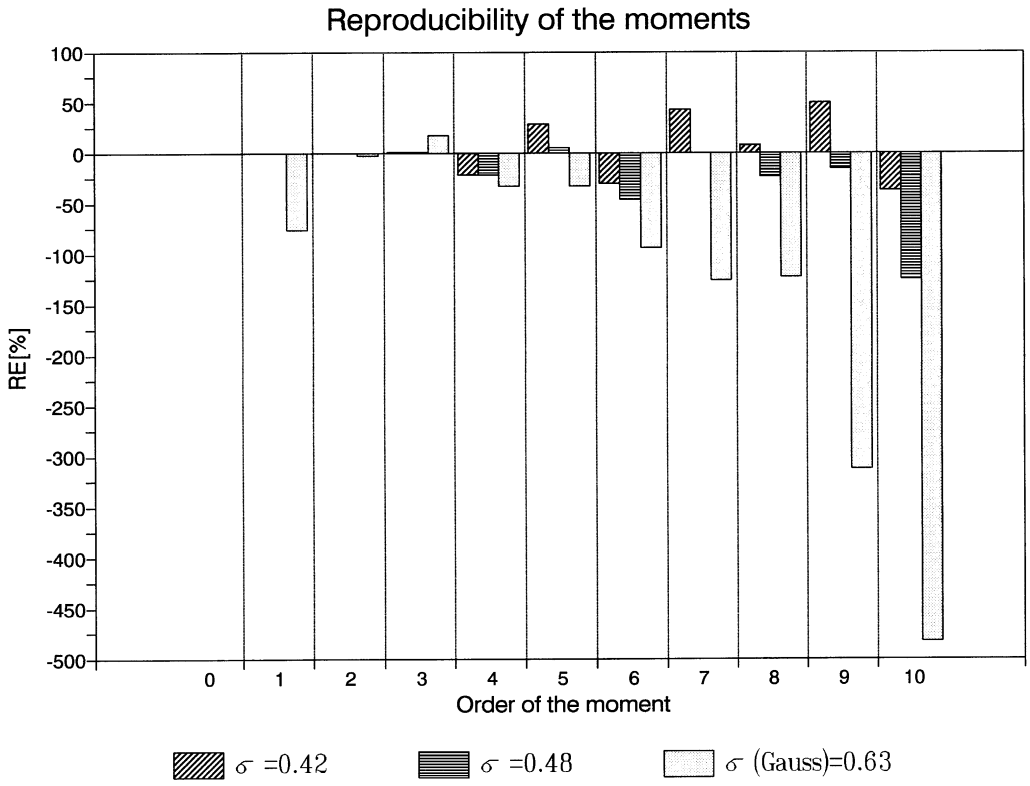


Fig. 6.

TABLE II. – *Reproducibility* (RE ($k$) *in* %) *of moments values up to* 10th *order. The comparison is made with input data set in table* I.

| | 1st simulation | | | | 2nd simulation | | |
|---|---|---|---|---|---|---|---|
| | $\sigma_S$ 0.39 | $\sigma_A$ 0.45 | $\sigma_P$ 0.46 | $\sigma_{\text{Gaus}}$ 0.63 | $\sigma_S$ 0.42 | $\sigma_A = \sigma_P$ 0.45 | $\sigma_{\text{Gaus}}$ 0.63 |
| $\mu^1$ | 10.74 | −24.5 | −24.5 | −78.3 | −0.0000013 | −0.001252 | −75.3 |
| $\mu^2$ | −0.7 | −0.6 | −0.6 | −2.4 | 0.0000025 | 0.000129 | −2.3 |
| $\mu^3$ | 3.1 | 3.5 | 3.7 | 17.6 | 0.0000026 | 0.000576 | 17.0 |
| $\mu^4$ | 23.7 | 6.2 | 3.4 | −32.7 | −21.5 | −21.6 | −32.9 |
| $\mu^5$ | 39.9 | 20.1 | 16.7 | −31.4 | 28.1 | 4.7 | 32.7 |
| $\mu^6$ | 46.8 | 16.0 | 10.2 | 92.9 | −30.7 | −45.9 | −94.2 |
| $\mu^7$ | 58.9 | 27.3 | 20.8 | −123.3 | 42.6 | 0.71 | −125.9 |
| $\mu^8$ | 62.5 | 22.5 | 13.7 | 215.3 | 33.0 | −76.6 | −218.8 |
| $\mu^9$ | 69.4 | 28.1 | 18.3 | −307.3 | 49.8 | −16.3 | −313.2 |
| $\mu^{10}$ | 71.7 | 22.8 | 10.3 | −474.6 | −37.7 | −124.5 | −483.2 |
| Average of absolute reproducibility values | | | | | | | |
| MD | 39 | 18 | 12 | 137 | 24 | 29 | 137 |
| SD | 26 | 9 | 7 | 144 | 17 | 39 | 144 |

to previous simulations where the $\sigma_{\text{Gaus}}$ corresponding to the classic choice yields a very high overestimate of the theoretical values.

The reproducibility RE ($k$) for the moments is shown in fig. 6. When the moments order is the same as in the input, the best reproducibility is within 22% for the $\sigma$ values related to the minimization process and within 75% for the $\sigma_{\text{Gaus}}$ choice.

Similarly to the previous simulation, higher-order moments are also overestimated in the classic case.

**4**·3. *Discussion.* – The comparison of the results points out that the Gaussian closure gives slightly worse results than in the first simulation.

In table II a summary of the moments reproducibility which comes from the two simulations is given.

It is evident that in the second simulation, the 4th moment is reproduced with about 21.5%, while in the first case it was at 9.6% on average. This means that the value of the 4th experimental moment is not well reproduced and the Gaussian closure can also be considered a very strong hypothesis.

For every $\sigma$ choice, table II also gives the mean (MD) and the standard deviation (SD) of absolute reproducibility values so defined:

$$\text{(4.5)} \qquad \begin{cases} \text{MD} = \dfrac{1}{10} \sum_{k=1,\,10} |\text{RE}(K)| \,, \\ \text{SD} = \sqrt{\dfrac{1}{9} \sum_{k=1,\,10} (|\text{RE}(k)| - \text{MD})^2} \,. \end{cases}$$

The best result comes from the first simulation $\sigma_P = 0.46$ with the average of the absolute reproducibility of 12% and 7% for MD and SD, respectively.

From table II it is clear that it is possible to improve not only the reproducibility of all the input moments (up to 3rd and up to 4th order), but also the moments up to 10th order.

Starting from $\mu^3$ and for the $\sigma_P = 0.46$ choice, the higher-order moments have an average reproducibility of 13.3%.

It is worth noting that in the table the bad result is given by a classic choice $\sigma\,(\sigma_{\text{Gaus}})$ in both simulations (in which the MD values are about 137%).

Sometimes Gaussian closure can improve the moments reproducibility but makes it compulsory to introduce a further hypothesis in the original PDF. Each time, it is possible to measure the quality of this hypothesis through the moments reproducibility for the same input order.

## 5. – Conclusion

The GC expansion calculates the probability density function using as input terms the moments data of any order.

The classic method uses a standardized random variable $Z$ and is related to the Gaussian distribution with mean equal to zero and standard deviation equal to one. This assumption is too strong and it is possible to regulate the standard deviation of the Gaussian distribution which may not necessarily relate to the classic choice.

The proposed approach improves the performance of the GC expansion.

The new formulation of the expansion optimizes the standard deviation of the Gaussian by the minimization of three functions related to the input moments. For every function a minimal value can be found.

Two simulations are performed using a given data set of input moments.

The resulting output moments, obtained from the minimization of the reproducibility and of the negative contributions, reproduce both the input moments in the best way possible and minimize the negative probabilities coming from GC expansion. Further, the new $\sigma$ choices give a good approximation of the moments of higher orders.

The $\sigma$ related to the standardization choice of the RV produces bad results in both simulations.

## Appendix A

### Determination and characterization of Hermite polynomials

The Hermite polynomials are a particular class of classic orthogonal polynomials (COP) in the range of $(-\infty\,;\,+\infty)$, with respect to weight function of the Gaussian type [3]. As a particular weight function the non-standard Gaussian function chosen

is defined as follows:

$$\alpha_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[x-m]^2/2\sigma^2 , \tag{A.1}$$

where $m$ and $\sigma$ are the mean and standard deviation of the Gaussian.

The Hermite polynomials are defined through a particular Rodriguez formula [11] and are computable as follows:

$$H_n(x) = \frac{(-1)^n}{\alpha_\sigma(x)} D_x^n \alpha_\sigma(x), \tag{A.2}$$

where the operator is defined as

$$D_x^n = \frac{\mathrm{d}^n}{\mathrm{d}x^n} . \tag{A.3}$$

All orthogonal polynomials take into account the "generating functions" $F(x)$ so that every polynomial $P_{n,x}$ is substantially the coefficient of $t^n$ in the series expansion of the $t$ function $F(x, t)$. In the specific case of Hermite polynomials, the generating function is defined as

$$\exp\left[\frac{t(x-m)}{\sigma^2} - \frac{t^2}{2\sigma^2}\right] = \sum_{j=0}^{\infty} \frac{t^j}{j!} H_j(x). \tag{A.4}$$

From (A.4) it can be deduced that the polynomials are defined as:

$$H_j(x) = D_t^j \left( \exp\left[\frac{t(x-m)}{\sigma^2} - \frac{t^2}{2\sigma^2}\right]\right)_{t=0} . \tag{A.5}$$

The above formulae (A.4) and (A.5) are fundamental for the deduction of the analytical properties of Hermite polynomials. Deriving (A.4) with respect to $x$ and equalling the coefficients of $t^j$, the derivation rules for $H_j$ are obtained as

$$D_x H_j(x) = \frac{j!}{(j-1)!} \frac{H_{j-1}(x)}{\sigma^2} = j \frac{H_{j-1}(x)}{\sigma^2} . \tag{A.6}$$

Generalizing the procedure, that is deriving (A.4) with respect to $x$ for $n$ times, the $n$-th derivative of $H_n$ is calculated as

$$D_x^n H_j(x) = \frac{j!}{(j-1)!} \frac{H_{j-n}(x)}{\sigma^2} . \tag{A.7}$$

The iteration rules for $H_j$ are very important because they are pratically used in the calculations of polynomials of $n$-order as a function of polynomials to the $(n-1)$ and $(n-2)$ order; deriving (A.4) with respect to $t$ and equalling the coefficients, the following formula is finally obtained:

$$H_j(x) - \frac{(x-m)}{\sigma^2} H_{j-1}(x) + \frac{(j-1)}{\sigma^2} H_{j-2}(x) = 0 . \tag{A.8}$$

The differential equation for Hermite polynomials is deduced from the $H_{j-1}$.

In fact from the $H_{j-1}$ and the $H_{j-2}$ terms, calculating by (A.6) and (A.7) and substituting this expression in eq. (A.8), the equation below is derived:

$$jH_j(x) - (x-m)\frac{\mathrm{d}}{\mathrm{d}x}H_j(x) + \sigma^2\frac{\mathrm{d}^2}{\mathrm{d}x^2}H_j(x) = 0\,. \tag{A.9}$$

As a last step, the orthogonalization polynomials rules must be obtained and the following integral must be considered:

$$\int_{-\infty}^{+\infty} H_m(x)\,H_n(x)\,\alpha_\sigma(x)\,\mathrm{d}x = (-1)^n \int_{-\infty}^{+\infty} H_m(x)\,D_x^n\,\alpha_\sigma(x)\,\mathrm{d}x =$$

$$= (-1)^n H_m(x)\,D_x^{n-1}\alpha_\sigma(x)\,\Big|_{-\infty}^{+\infty} + (-1)^{n-1}\int_{-\infty}^{+\infty}\frac{\mathrm{d}}{\mathrm{d}x}H_m(x)\,D_x^{n-1}\alpha_\sigma(x)\,\mathrm{d}x\,,$$

where the partial integral rules are evidently used. The first term shall be

$$(-1)^n H_m(x)\,H_n(x)\,\alpha_\sigma(x)]_\infty = 0\,, \tag{A.10}$$

because, obviously,

$$\alpha_\sigma(\pm\infty) = 0\,.$$

The second term is similar to the integral of the first step. Performing the calculations, the ortogonalization rules are

$$\int_{-\infty}^{+\infty} H_m(x)\,H_n(x)\,\alpha_\sigma(x)\,\mathrm{d}x = \delta_n^m\,\frac{n!}{\sigma^{2n}}\,. \tag{A.11}$$

## Appendix B

### Determination of Gram-Charlier coefficients

The development of GC consists in determining the PDF as

$$f(x) = \sum_{j=0}^{\infty} C_j\,H_j(x)\,\alpha_\sigma(x) = \alpha_\sigma(x)\sum_{j=0}^{\infty} C_j\,H_j(x)\,. \tag{B.1}$$

From (B.1), multiplying by $H_n$ and integrating by (A.11), the following formula results:

$$\int_{-\infty}^{+\infty} f(x)\,H_n(x)\,\mathrm{d}x = C_n\,\frac{n!}{\sigma^{2n}}\,, \tag{B.2}$$

from which the coefficients are

$$C_n = \frac{\sigma^{2n}}{n!} \int_{-\infty}^{+\infty} f(x)\, H_n(x)\, \mathrm{d}x \,. \tag{B.3}$$

From (A.4) it is possible to explicity deduce the Hermitian polynomial for different values of $n$:

$$H_{2n+\delta_d} = \frac{1}{\sigma^{2(2n+\delta_n)}} \sum_{z=0}^{n} (-1)^z (x-m)^{2n+\delta_d-2z} \frac{(2n+\delta_d)!\sigma^{2z}}{(2n+\delta_d-2z)!\,2^z z!} \,, \tag{B.4}$$

where $n = 0, 1, 2, \ldots$ and $\delta_d$ is

$$\delta_d = \begin{cases} +1 & \text{if } (2n+\delta_d) \text{ is odd} \,, \\ 0 & \text{if } (2n+\delta_d) \text{ is even} \,. \end{cases} \tag{B.5}$$

Replacing (B.4) in (B.3), the following formula is obtained:

$$C_{(2n+\delta_d)} = \sum_{z=0}^{n} \frac{(-1)^z \sigma^{2z}}{(2n+\delta_d-2z)!\,2^z z!} \int_{-\infty}^{+\infty} f(x)(x-m)^{2n-\delta_d-2z}\, \mathrm{d}x \,. \tag{B.6}$$

Substituing the $(x-m)^k$ in (B.6) and taking into account the decomposition:

$$(x-m)^k = \sum_{\nu=0}^{k} (-1)^{k-\nu} x^\nu m^{k-\nu} \binom{k}{k-\nu}, \tag{B.7}$$

the calculation of the coefficients $C_n$ is

$$C_{(2n+\delta_d)} = \sum_{z=0}^{n} \frac{(-1)^z \sigma^{2z}}{(2n+\delta_d-2z)!\,2^z z!} \cdot \tag{B.8}$$

$$\cdot \sum_{\nu=0}^{2n+\delta_d-2z} = \binom{2n+\delta_d-2z}{2n+\delta_d-2z-\nu} (-1)^{2n+\delta_d-2z-\nu} m^{2n+\delta_d-2z-\nu} \mu^\nu \,,$$

where the $\mu^k$ moments are explicitly introduced.

Replacing the binomial coefficient in (B.7) the final $C_n$ determination is

$$C_{(2n+\delta_d)} = \sum_{z=0}^{n} \frac{\sigma^{2z}}{2^z z!} \sum_{\nu=0}^{2n+\delta_d-2z} \frac{(-1)^{2n+\delta_d-z-\nu}}{(2n+\delta_d-2z-\nu)!\,\nu!} m^{2n+\delta_d-2z-\nu} \mu^\nu \,. \tag{B.9}$$

This formula is the most general possible form with which coefficents are represented in the development of GC because it refers to any choice of $\sigma$ and $m$.

## REFERENCES

[1] WIENER N., *The Fourier Integral and Certain of its Applications* (Cambridge) 1933.
[2] EDGEWORTH F. Y., *The Law of Error,* PCPS, **20** (1905) 36.

[3] CHARLIER C. V. L., *Researches into the theory of probability, K. Fysiogr. Sällsk. Handl., B* 16 (1906).
[4] SCARANI C., *Valutazione di distribuzioni in un problema di dispersione turbolenta, Statistica*, anno LI, 4 (1991).
[5] BURESTI G., MORDACCI A., PETAGNA P., TALAMELLI A. and TANZINI G., *LDA Characterization of Coaxial Jets and Comparison with other Techniques*, *Proceedings of the VI International Symposium on Applications of Laser Techniques to Fluid Mechanics, Lisbon*, paper 17-2, 1992.
[6] MENGALI G. and MICHELI M., *Simulazione di segnali rappresentativi di flussi turbolenti con caratteristiche statistiche assegnate*, *ADIA 93, 5 Giugno 1993.*
[7] THOMSON D. J., *Random walk modelling of diffusion in inhomogeneous turbulence*, *Q. J. R. Meteorol. Soc.*, **110** (1984) 1107.
[8] TAMPIERI F., SCARANI C., GIOSTRA U., BRUSASCA G., TINARELLI G., ANFOSSI D. and FERRERO E., *On the application of random flight dispersion models in inhomogeneous turbulent flows, Ann. Geophys.*, **10** (1992) 749.
[9] LUPINI R. and TIRABASSI T., *Solution of the advection-diffusion equation by the moments method*, *Atmos. Environ.*, **17** (1983) 965.
[10] CRAMER H., *Mathematical Methods of Statistics* (Princeton University Press) 1961.
[11] ROSSETTI C., *Metodi matematici per la fisica* (Libreria Editrice Universitaria Levrotto & Bella) 1982.
[12] NIEUWSTADT F. T. and VAN DOP H. (Editors), *Atmospheric Turbulence and Air Pollution Modeling* (D. Reidel, Dordrecht) 1982.
[13] CANCELLI C., CENEDESE A. and LEUZZI G., *Modelli Lagrangiani per la valutazione della dispersione di inquinanti, Ingegneria Sanitaria*, **4** (1989) 8.
[14] PAPOULIS A., *Probability, Random Variable, and Stochastic Processes*, third edition (McGraw-Hill) 1991.