International Statistical Review (2000), **68**, 2, 137–153, Printed in Mexico

The Early History of the Cumulants and the Gram–Charlier Series

Anders Hald

Department of Theoretical Statistics, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen Ø, Denmark

Summary

The early history of the Gram–Charlier series is discussed from three points of view: (1) a generalization of Laplace's central limit theorem, (2) a least squares approximation to a continuous function by means of Chebyshev–Hermite polynomials, (3) a generalization of Gauss's normal distribution to a system of skew distributions. Thiele defined the cumulants in terms of the moments, first by a recursion formula and later by an expansion of the logarithm of the moment generating function. He devised a differential operator which adjusts any cumulant to a desired value. His little known 1899 paper in Danish on the properties of the cumulants is translated into English in the Appendix.

Key words: Bessel; Bienaymé; Central limit theorem; Chebyshev; Cumulants; Gram; Gram–Charlier series; Halfinvariants; Hausdorff; Hermite; Laplace; Least squares; Moments; Orthogonal polynomials; Poisson; Thiele.

1 Introduction

The normal distribution was introduced as an error distribution by Gauss in 1809 and as a large-sample distribution of the arithmetic mean by Laplace in 1810. It did not last long before a generalization was carried out by considering the normal distribution as the first term of a series expansion, later known as the Gram–Charlier series. We shall discuss the historical development of this series from three different points of view.

In section 2 we describe how Poisson, Bessel and Bienaymé generalized the Laplacean central limit theorem by including more terms in the expansion of the logarithm of the characteristic function. By means of the inversion formula they found an expansion, the Gram–Charlier series, for the density of a sum of independently and identically distributed random variables with the normal density as the leading term. The following terms contain Hermite polynomials multiplied by complicated moment coefficients derived from the underlying distribution, which they determined up to the moments of the sixth order. These coefficients were simplified by Thiele, who expressed them in terms of the cumulants and derived a recursion formula for them.

In section 3 we discuss Chebyshev's least squares fitting of a polynomial to the observed values of a function by means of orthogonal polynomials. He pointed out the advantages of the successive determination of the coefficients and the residual sum of squares. Generalizing this method Chebyshev found the least squares approximation to an arbitrary integrable function by means of an infinite series of orthogonal polynomials, and choosing the normal density as weight function the Hermite polynomials were introduced. Applying this method to a continuous density the Gram–Charlier series follows. The generalized central limit theorem discussed in section 2 may thus be considered as the least squares approximation to a continuous density.

In section 4 we explain how Thiele and Gram introduced the Gram–Charlier series from a completely different point of view. Realizing that the normal distribution was unsatisfactory for describing economic and demographic data, they proposed to multiply the normal density by a power series and determine the coefficients by least squares, which led to the Gram–Charlier series and thus a new system of skew distributions.

Thiele pointed out the one-to-one relationship between n observations and the first n symmetric functions and stated that the cumulants are the simplest for describing the data. He stressed that the first four cumulants and the corresponding terms of the Gram–Charlier series often will give a satisfactory characterization of a distribution. As mentioned above he expressed the coefficients of the Gram–Charlier series in terms of the cumulants.

In the first instance Thiele (1889) defined the cumulants recursively in terms of the moments; we shall suggest how he may have arrived at this formula. In a little known paper (1899) he gave the modern definition of the cumulants as the coefficients in the power series for the logarithm of the moment generating function; he did not mention the similar expansion of the logarithm of the characteristic function in the derivation of the central limit theorem.

In this paper he also introduced the operator $M(-D)$, where $M(t) = E(e^{xt})$ and D denotes differentiation, and derived the Gram–Charlier series by applying this operator to the normal distribution. Finally, he showed that the operator $\exp[\alpha_r(-D)^r]$ applied to a density with cumulant κ_r increases κ_r by α_r. By means of a product of such operators he thus transformed a density with given cumulants to another with specified cumulants.

We have translated Thiele's 1899 paper from Danish into English in the Appendix.

In the early proofs the authors tacitly assumed that all moments are finite and that the moment generating function exists, only Thiele and Gram discussed problems of convergence.

In accordance with common usage in statistics we assume that the Hermite polynomials $H_r(x)$ are defined by differentiation of $\exp(-x^2/2)$ and that the reader is familiar with the properties of these polynomials.

2 The Central Limit Theorem, the Moments, and the Gram–Charlier Series

Let x be a random variable with finite moments, μ'_r and $\mu_r, r = 0, 1, 2, \ldots$, cumulants κ_r, moment generating function $M(t) = E(e^{xt})$, characteristic function $\psi(t) = E(e^{ixt})$, and cumulant generating function $\kappa(t) = \ln \psi(t)$. The following discussion is based on the expansion

$$\psi(t) = 1 + i\mu'_1 t - \mu'_2 t^2/2! - i\mu'_3 t^3/3! + \mu'_4 t^4/4! + \ldots, \tag{2.1}$$

and the corresponding power series for $\ln \psi(t)$.

Laplace uses the characteristic function and its logarithm in his proofs of the central limit theorem. For an arbitrary distribution he (1810, Art. VI; 1812, II, § 22) finds

$$\ln \psi(t) = i\mu'_1 t - (\mu'_2 - \mu_1'^2)t^2/2! + \ldots,$$

and for a symmetric distribution with zero mean he (1811, Art. VIII; 1812, II, § 20) obtains

$$\ln \psi(t) = -\mu'_2 t^2/2! + (\mu'_4 - 3\mu_2'^2)t^4/4! + \ldots,$$

where we have introduced the notation μ'_r instead of Laplace's k_r/k.

These expansions were studied in more detail by Poisson (1829, pp. 8–9; 1837, p. 269), who gives his result as

$$\ln \psi(t) = i\mu'_1 t - ht^2 - igt^3 + (l - \frac{1}{2}h^2)t^4 + \ldots,$$

where

$$h = \frac{1}{2}(\mu_2' - \mu_1'^2) = \frac{1}{2}\mu_2 = \frac{1}{2}\kappa_2,$$
$$g = \frac{1}{6}(\mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3) = \frac{1}{3!}\mu_3 = \frac{1}{3!}\kappa_3.$$
$$l = \frac{1}{24}(\mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4) = \frac{1}{4!}\mu_4,$$
$$l - \frac{1}{2}h^2 = \frac{1}{4!}(\mu_4 - 3\mu_2^2) = \frac{1}{4!}\kappa_4.$$

The expressions for h and g in terms of μ_r' are due to Poisson, he leaves the derivation of l to the reader. We have added the corresponding formulas in terms of μ_r and κ_r. Referring to Laplace, but not to Poisson, Bienaymé (1852, p. 45) gives a simpler proof of Poisson's result, including the expression for l.

Bessel (1838) refers to Laplace and Poisson and extends Laplace's expansion for a symmetric distribution by including the term

$$-(\mu_6' - 15\mu_4'\mu_2' + 30\mu_2'^3)t^6/6! = \kappa_6 t^6/6!.$$

It follows from Laplace's definition of the characteristic function that the characteristic function for a linear combination of independent random variables equals the product of the characteristic functions for the components, and hence that the logarithm equals the sum of the logarithms. This property is used by Poisson, Bessel and Bienaymé in their proofs. As pointed out by Bru (1991) in his discussion of Bienaymé's proof this implies that

$$\kappa_r\left(\sum_1^n a_i x_i\right) = \sum_1^n \kappa_r(a_i x_i) = \sum_1^n a_i^r \kappa_r(x_i), \quad r = 1, 2, \dots. \tag{2.2}$$

Let $s_n = x_1 + \cdots + x_n$ be the sum of n independently and identically distributed variables with a continuous density. We shall sketch a proof of the central limit theorem based on a combination of the methods used by the four authors mentioned above. The basic idea and technique is due to Laplace. For brevity we shall introduce the cumulants in the expansion of $\ln \psi(t)$ instead of the coefficients expressed in terms of the moments about zero, that is, we write

$$\ln \psi(t) = \sum_{r=1}^{\infty} (it)^r \kappa_r / r!. \tag{2.3}$$

We shall like Bessel stop at the sixth term.

The density of s_n is found from the inversion formula

$$p(s_n) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-is_n t)\psi^n(t)dt,$$

where

$$\psi^n(t) = \exp(in\kappa_1 t - n\kappa_2 t^2/2!)\ \{1 + R(t)\},$$
$$R(t) = -in\kappa_3 t^3/3! + n\kappa_4 t^4/4! + in\kappa_5 t^5/5! - (n\kappa_6 + 10n\kappa_3^2)t^6/6! + \dots,$$

which follows from the expansion of $n \ln \psi(t)$.

The main terms becomes

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \exp[i(n\kappa_1 - s_n)t - n\kappa_2 t^2/2]dt = (n\kappa_2)^{-\frac{1}{2}}\phi(u), \tag{2.4}$$

where

$$u = (s_n - n\kappa_1)/\sqrt{n\kappa_2}$$

and $\phi(u)$ denotes the standardized normal density.

To evaluate the following terms the authors mentioned differentiate the two sides of (2.4) with respect to s_n with the result that

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \exp[i(n\kappa_1 - s_n)t - n\kappa_2 t^2/2](-it)^r dt = (-1)^r (n\kappa_2)^{-(r+1)/2} H_r(u)\phi(u),$$

where we have denoted the polynomials by H_r since they are equal to the Hermite polynomials. By means of this result the terms involving $R(t)$ are easily found.

Setting

$$\gamma_r = \kappa_{r+2}/\kappa_2^{(r+2)/2}, \quad r = 1, 2, \ldots, \tag{2.5}$$

the expansion may be written as

$$p(s_n) = \frac{\phi(u)}{\sqrt{n\kappa_2}}\left[1 + \frac{\gamma_1 H_3(u)}{3!n^{1/2}} + \frac{\gamma_2 H_4(u)}{4!n} + \frac{\gamma_3 H_5(u)}{5!n^{3/2}} + \frac{1}{6!}\left(\frac{\gamma_4}{n^2} + 10\frac{\gamma_1^2}{n}\right) H_6(u) + \ldots\right], \tag{2.6}$$

which today is known as the Gram–Charlier series.

The first term is due to Laplace, who also found the third term for a uniformly distributed variable. The general form of the second term is given by Poisson. For a symmetric distribution Bessel derives the third and the fifth term. Bienaymé derives the first three terms to which he adds the term $\gamma_1^2 H_6(u)/72n$ because it is of the same order in n as the third term. Ordering the terms according to powers of $n^{-\frac{1}{2}}$ the resulting series is called the Edgeworth series.

One may wonder why the synthesis above does not occur in the early literature on the central limit theorem. The reason may be that the second term, the correction for skewness, was deemed sufficient for practical applications. Laplace developed his large-sample estimation and testing theory using the asymptotic normality of the linear estimates involved. Poisson took the correction for skewness into account in some of his examples but found that the effect was negligible with the sample sizes at hand.

As noted by Molina (1930), Laplace (1811, Art. V; 1812, II, §17) derives the complete Gram–Charlier expansion in his discussion of a diffusion problem, using orthogonal polynomials proportional to the Hermite polynomials with argument $x\sqrt{2}$.

A more detailed discussion of the early history of the central limit theorem is given by Hald (1998, Chapter 17).

In the early proofs of the central limit theorem the cumulants thus implicitly occur. However, nobody thought of defining the coefficients in question as separate entities and to study their properties. This had to wait for Thiele (1889).

3 Least Squares Approximation by Orthogonal Polynomials

The problem of fitting a polynomial of a given degree, m say, to $n+1$ observations (x_0, y_0), $(x_1, y_1), \ldots, (x_n, y_n)$, $m < n$, where

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_m x_i^m + \varepsilon_i, \quad i = 0, 1, \ldots, n, \tag{3.1}$$

$E(\varepsilon_i) = 0$ and $\mathrm{var}(\varepsilon_i) = \sigma^2/w_i$, w_i being a known positive number, is a special case of the Gaussian linear minimum variance estimation by the method of least squares.

In a series of papers between 1835 and 1853 Cauchy extends the linear model by considering the case where the number of terms is unknown; in particular he discusses the polynomial model with unknown m. Rewriting (3.1) in the form

$$y_i = \alpha_0 h_0(x_i) + \alpha_1 h_1(x_i) + \alpha_2 h_2(x_i) + \cdots + \varepsilon_i, \tag{3.2}$$

where $h_r(x)$, $r = 0, 1, 2, \ldots$, is a suitably chosen polynomial of degree r, he shows how the true value of y may be estimated stepwise such that the inclusion of one more term does not change the coefficients of the previous terms. After adding a new term the residuals are calculated and used for judging the goodness of fit and thus for deciding when to stop the procedure. Cauchy does not use the method of least squares so his polynomials are nonorthogonal. Bienaymé criticizes Cauchy for not using the method of least squares but does not himself work out the corresponding theory. This was done by Chebyshev who did not know Cauchy's work when he wrote his first paper (1855). However, in his second paper (1859a) he refers to Cauchy and mentions that the regression coefficients are determined successively and independently of one another. A survey of Cauchy's work may be found in Hald (1998, Chapter 24).

Let there be given $n+1$ observations $f(x_0), f(x_1), \ldots, f(x_n)$ and let the reciprocal of the variance of $f(x_i)$ be $w(x_i) > 0$. By the method of least squares Chebyshev (1855) fits a polynomial of degree m to the $n + 1$ observations, $m < n$, using $w(x_i)$ as weight. We shall write the model as

$$f(x_i) = \alpha_0 h_0(x_i) + \alpha_1 h_1(x_i) + \cdots + \alpha_m h_m(x_i) + \varepsilon_i, \quad i = 0, 1, \ldots, n. \tag{3.3}$$

By means of the theory of continued fractions he proves that there exists a set of $m + 1$ orthogonal polynomials $\{h_r(x)\}$ satisfying the relations

$$\sum_{i=0}^{n} h_r(x_i) h_m(x_i) w(x_i) = 0, \quad r = 0, 1, \ldots, m-1, \quad h_0(x) = 1,$$

h_r being of degree r, and that the least squares approximation may be written as

$$f_m(x) = \sum_{r=0}^{m} a_r h_r(x)$$

with

$$a_r = \frac{\sum_{i=0}^{n} f(x_i) h_r(x_i) w(x_i)}{\sum_{i=0}^{n} h_r^2(x_i) w(x_i)},$$

where a_r, the estimate of α_r, depends on h_r only. The polynomials may be normed in any convenient way, in particular they may be made orthonormal.

In the next paper Chebyshev (1859a) proves that $h_r(x)$ may be found as the denominator of the r-th convergent in the continued fraction for the function

$$\sum_{i=0}^{n} \frac{w(x_i)}{x - x_i},$$

and finds a recursion formula for h_r in terms of h_{r-1} and h_{r-2}. He notes that the minimum sum of squares equals

$$\sum_{i=0}^{n} [f(x_i) - f_m(x_i)]^2 w(x_i) = \sum_{i=0}^{n} f^2(x_i) w(x_i) - \sum_{r=0}^{m} a_r^2 \sum_{i=0}^{n} h_r^2(x_i) w(x_i),$$

which may be used to decide when to stop.

Referring to Chebyshev (1855), Hermite (1859) gives a simpler derivation of $h_r(x)$. Neither Chebyshev nor Hermite observes that Laplace's (1816) orthogonalization of the equations of condition for the linear model provides a formula from which the orthogonal polynomials may be found by recursion.

Inspired by Fourier's representation of an arbitrary function as an infinite trigonometric series and by his own results above Chebyshev (1859b) presents a method for approximating an arbitrary integrable function $f(x)$ by an infinite series in terms of orthogonal polynomials. Briefly told, he replaces the sums above by integrals, assuming that x and $w(x)$ are continuous and that the integrals involved are finite; $f(x)$ is defined on a finite or an infinite interval, we shall leave out the limits of integration. Hence,

$$f(x) \sim \sum_{r=0}^{\infty} a_r h_r(x), \quad a_r = \int f(x)h_r(x)w(x)dx \Big/ \int h_r^2(x)w(x)dx\,, \tag{3.4}$$

and

$$\int h_r(x)h_s(x)w(x)dx = 0 \quad \text{for} \quad r \neq s. \tag{3.5}$$

The polynomial $h_r(x)$ is found as the denominator for the r-th convergent in the continued fraction for the function

$$\int \frac{w(t)}{x-t}dt.$$

By suitable choices of $w(x)$ Chebyshev obtains the series that today are named after Maclaurin, Fourier, Legendre, Laguerre, and Hermite; we shall only discuss the last one.

Choosing as weight function

$$\begin{aligned}\phi_k(t) &= \sqrt{k/\pi}\exp(-kt^2), \quad -\infty < t < \infty,\\ &= (2\pi)^{-\frac{1}{2}}\omega^{-1}\exp(-t^2/2\omega^2), \quad \omega = (2k)^{-\frac{1}{2}},\end{aligned}$$

Chebyshev finds that

$$\begin{aligned}h_r(x) &= e^{kx^2}\left(\frac{d}{dx}\right)^r e^{-kx^2}\\ &= (-1)^r\omega^{-r}H_r(u), \quad u = x/\omega.\end{aligned}$$

Hence, in our notation Chebyshev's results may be written as

$$f(x) \sim \sum_{r=0}^{\infty}\frac{1}{r!}H_r(u)\int \phi(u)H_r(u)f(\omega u)du \tag{3.6}$$

and

$$f(x) \sim \sum_{r=0}^{\infty}\frac{1}{r!}H_r(u)\omega^r\int \phi(u)f_u^{(r)}(\omega u)du.$$

Chebyshev remarks that for $\omega \to 0$ the series tends to Maclaurin's series since $H_r(u)\omega^r \to x^r$. Chebyshev does not discuss further applications of the series.

Hermite (1864) begins by defining the polynomials $U_r(x)$ by the equation

$$e^{-x^2}U_r(x) = \left(\frac{d}{dx}\right)^r e^{-x^2}, \quad r = 0, 1, \ldots .$$

He discusses the properties of these polynomials and the corresponding infinite series, remarking that it follows from Bienaymé's note to the translation of Chebyshev's 1855 paper that the series belongs to the wide category of expansions which give interpolation formulas derived by the method of least squares. Most of Hermite's paper is taken up with generalizations to two and more variables.

Gnedenko & Kolgomorov (1954) call the polynomials $H_r(x)$ Chebyshev–Hermite polynomials; presumably they did not know that the polynomials previously had been used by Laplace, as noted by Molina (1930) and Uspensky (1937, p. 72).

To see the connection with the cumulants and the Gram–Charlier series we shall consider the expansion of a frequency function written in the form $g(x) = \phi(x)f(x)$. Using Chebyshev's formula (3.6) for $\omega = 1$ we get

$$g(x) \sim \phi(x) \sum_{r=0}^{\infty} \frac{1}{r!} H_r(x)\, E[H_r(x)].$$

Since $H_r(x)$ is a polynomial of degree r, $E[H_r(x)]$ is a linear combination of the moments of $g(x)$. For a standardized variable we get

$$g(x) = \phi(x)\left(1 + \frac{\kappa_3 H_3(x)}{3!} + \frac{\kappa_4 H_4(x)}{4!} + \frac{\kappa_5 H_5(x)}{5!} + \frac{1}{6!}(\kappa_6 + 10\kappa_3^2)H_6(x) + \dots\right). \tag{3.7}$$

This is the Gram–Charlier series for $g(x)$ in the form given by Thiele (1889). It is equal to (2.6) for $n = 1$ and $\kappa_2 = 1$.

Chebyshev did not notice this coincidence in 1859. However, in 1887 he derives the main terms of the central limit theorem and gives limits for the remainder using a new method of proof, the method of moments. At the end of this paper he briefly states that $p(s_n)$ may be expanded in an infinite series by the method given in his 1859b paper and indicates the result in the form (2.6) but leaves to the reader to find the coefficients in terms of the moments. Hence, the series (2.6) may be interpreted as the least squares approximation to $p(s_n) = \phi(u)f(u)$.

4 Thiele's Halfinvariants, Their Operational Properties, and the Gram–Charlier Series

A survey of Thiele's statistical work is given by Hald (1981). Here we present some supplementary remarks on Thiele's invention and use of the cumulants and the Gram–Charlier series.

Thiele named his new symmetric functions halfinvariants. For linguistic reasons some later authors called them semi-invariants. Fisher (1929) introduced the term cumulative moment function instead of semi-invariant, and Fisher & Wishart (1931) abbreviated this to cumulant. In a personal communication S.M. Stigler has pointed out to me that Hotelling (1933) claims the priority to the term cumulant.

We shall transcribe Thiele's formulas to modern notation using h_r and κ_r for the empirical and theoretical cumulants, respectively, following Fisher's notation.

Thiele (1889) notes that there exists a one-to-one relationship between the n observations and the first n symmetric functions. He introduces the raw moments m'_r, $r = 0, 1, \dots, n$, the moments about the arithmetic mean m_r, and the reduced moments $m_r/m_2^{r/2}$, remarking that the observations without loss of information may be represented by either $(m'_1, \dots, m'_n)$, or $(m'_1, m_2, \dots, m_n)$, or $(m'_1, m_2, m_3/m_2^{3/2}, \dots, m_n/m_2^{n/2})$. He adds that the first four terms of either of the last two sets often will give a good characterization of the empirical distribution. However, he (p. 21) points out that it is a drawback that the moments of even order and also the reduced moments increase rapidly with r and that it would be advantageous to introduce a new symmetric function not having this property. We believe that this is the key to his recursive definition of the cumulants.

In his discussion of the relation between the moments he derives the formula

$$m'_r = \sum_{j=0}^{r} \binom{r}{j} m_1'^{\,r-j} m_j,\ m_0 = 1, m_1 = 0, r = 0, 1, \dots.$$

To get a symmetric function of even order which is smaller than m_j it is necessary to replace $m_1'^{\,r-j}$ by a larger number, a natural choice being m'_{r-j}. This may have induced Thiele to define the cumulants by the recursion

$$m'_{r+1} = \sum_{j=0}^{r} \binom{r}{j} m'_{r-j} h_{j+1}, \quad r = 0, 1, \ldots, n-1, \tag{4.1}$$

and $h_0 = 1$. Solving for the h's Thiele finds the first six empirical cumulants in terms of m' and m, the first four being $h_1 = m'_1$, $h_2 = m_2$, $h_3 = m_3$, and $h_4 = m_4 - 3m_2^2$, the last one showing the desired effect.

Since the theoretical distribution is the limit of the empirical for $n \to \infty$ the theoretical cumulants are defined by the recursion

$$\mu'_{r+1} = \sum_{j=0}^{r} \binom{r}{j} \mu'_{r-j} \kappa_{j+1}, \quad r = 0, 1, \ldots, \tag{4.2}$$

and $\kappa_0 = 1$. By means of this formula Thiele derives the cumulants for several distributions; we shall only give his results for the normal distribution and the Gram–Charlier series.

For the standardized normal distribution he finds that $\mu_{2r+1} = 0$ and $\mu_{2r} = 1 \cdot 3 \cdots (2r-1)$. It follows from (4.2) that the cumulants of uneven order equal zero and those of even order satisfy the relation

$$1 \cdot 3 \cdots (2r-1) = \binom{2r-1}{1} 1 \cdot 3 \cdots (2r-3) + \sum_{j=1}^{r-2} \binom{2r-1}{2j+1} 1 \cdot 3 \cdots (2r-2j-3) \kappa_{2j+2} + \kappa_{2r},$$

which shows that $\kappa_{2r} = 0$ for $r = 2, 3, \ldots$.

It must have been a great satisfaction for Thiele to find that the mean and the variance of a normally distributed variable equals κ_1 and κ_2, respectively, and that all the cumulants of higher order equal zero. Thiele (1903, p. 25) writes: "This remarkable proposition [$\kappa_r = 0$ for $r \geq 3$] has originally led me to prefer the half-invariants to every other system of symmetrical functions."

Obviously, the normal distribution is insufficient for describing distributions of economic and demographic data as encountered by actuaries. It seems that the many-talented Danish philologist, politician, forester, statistician, and actuary L.H.F. Oppermann is the first to propose a system of skew frequency functions obtained by multiplying the normal density by a power series. Moreover, he proposed to estimate the coefficients by the method of moments. Oppermann did not himself publish these ideas; they are reported by his younger actuarial colleagues Thiele (1873) and Gram (1879, pp. 93–94).

Thiele (1873) writes the density in the form

$$g(x) = k_0\phi(x) + k_1\phi'(x) + k_2\phi''(x) + \ldots, \quad \phi(x) = \exp(-\pi x^2).$$

In the textbook he (1889, p. 14) uses

$$\begin{aligned} g(x) &= \sum_{r=0}^{\infty} (-1)^r c_r \phi^{(r)}(x)/r! \\ &= \phi(x) \sum_{r=0}^{\infty} c_r H_r(x)/r!, \quad \phi(x) = (2\pi)^{-\frac{1}{2}} \exp(-x^2/2). \end{aligned} \tag{4.3}$$

Regarding the two forms of ϕ he (1903, p. 17) remarks: "In the majority of its purely mathematical applications $e^{-\pi u^2}$ is preferable, unless (as in the whole theory of observations) the factor $\frac{1}{2}$ in the index is to be preferred on account of the resulting simplifications of most of the derived formulæ." He disapproves of the then commonly used form $\pi^{-\frac{1}{2}} e^{-u^2}$, which goes back to Gauss. Independently of Thiele, Stigler (1983) has proposed the form $e^{-\pi u^2}$.

To estimate the coefficients in the expansion for a grouped frequency distribution Thiele (1889,

p. 100) introduces the standardized variable $z = (x - \bar{x})/s$, where $\bar{x}$ and s denote the sample mean and standard deviation for the n observations. Taking the length of the constant class interval as unit and denoting the i-th relative frequency by f_i, $\sum f_i = 1$, Thiele writes

$$f_i \cong f(z_i) = \phi(z_i) \sum_{r=0}^{\infty} c_r H_r(z_i)/r!,$$

where $f(z_i)$ denotes the density at the midpoint of the interval. Using $1/\phi(z_i)$ as weight the least squares estimate of c_r is obtained by minimizing

$$\sum [f_i - f(z_i)]^2/\phi(z_i),$$

which gives the normal equations

$$\sum_i H_r(z_i) f_i = \sum_i H_r(z_i) f(z_i) \cong c_r, \; r = 1, 2, \ldots, \tag{4.4}$$

if the sum on the right side can be approximated by the integral. The estimate is thus obtained by equating the empirical and the theoretical means of $H_r(x)$. This is Gram's explanation of Oppermann's method of estimation. The justification for the choice of weight function and the replacement of the sum by the integral was provided by Thiele.

Thiele (1873) investigates the difference between the sum and the integral in (4.4) and concludes (wrongly) that it is insignificant for small values of r if the length of the class interval is less than the standard deviation. He missed the Sheppard corrections for grouping. In (1189, p. 28) he states that for $r = 3$ and 4 the class interval should be at most one-fourth of the standard deviation.

Thiele (1889, p. 100) considers $g(x)$ as a special case of the linear model. He notes that $\mathrm{var}(f_i) = f_i(1 - f_i)/n \cong f_i/n$, if the number of classes is large. The weight is thus proportional to $1/f_i$ which he approximates by $1/\phi(x_i)$, and the least squares solution then follows as shown above. However, Thiele overlooks that f_i and f_j, $i \neq j$, are correlated with covariance $-f_i f_j/n$ so that the assumptions for applying the method of least squares are not satisfied. Nevertheless the method leads to simple estimates of high efficiency for moderately skew distributions. Thiele should of course have minimized the quadratic form obtained by taking the covariance into account, which would have led him to the minimum χ^2 estimate.

To find the coefficients of the Gram–Charlier series in terms of the cumulants Thiele starts from the relation $c_r = E[H_r(x)]$, which gives c_r in terms of $\mu'_1, \ldots, \mu'_r$. He then introduces the cumulants instead of the moments by means of the recursion formula. We shall give a slightly simplified version of Thiele's proof using the symbol μ'^r for μ'_r and similarly for κ so that (4.2) becomes

$$\mu'^{r+1} = \kappa(\mu' + \kappa)^r,$$

and $E[H_r(x)] = H_r(\mu')$.

Thiele writes the Hermite polynomials in the form

$$H_{2r}(x) = \sum_{i=0}^{r} (-1)^i \frac{(2r)^{(2i)}}{2^i i!} x^{2r-2i}, \quad r = 0, 1, \ldots,$$

and

$$H_{2r+1}(x) = \sum_{i=0}^{r} (-1)^i \frac{(2r+1)^{(2i)}}{2^i i!} x^{2r-2i+1}, \quad r = 0, 1, \ldots.$$

From the recursion formula

$$H_r(x) - xH_{r-1}(x) + (r-1)H_{r-2}(x) = 0,$$

he obtains

$$c_r + (r-1)c_{r-2} = E[xH_{r-1}(x)].$$

Hence,

$$\begin{aligned}
c_{2r} + (2r-1)c_{2r-2} &= \mu'_{2r} - \frac{(2r-1)^{(2)}}{2\cdot 1!}\mu'_{2r-2} + \frac{(2r-1)^{(4)}}{2^2\cdot 2!}\mu'_{2r-4} - \dots \\
&= \kappa(\mu'+\kappa)^{2r-1} - \frac{(2r-1)^{(2)}}{2\cdot 1!}\kappa(\mu'+\kappa)^{2r-3} + \dots \\
&= \kappa_1\left[\mu'_{2r-1} - \frac{(2r-1)^{(2)}}{2\cdot 1!}\mu'_{2r-3} + \dots\right] \\
&+ \kappa_2\left[\binom{2r-1}{1}\mu'_{2r-2} - \frac{(2r-1)^{(2)}}{2\cdot 1!}\binom{2r-3}{1}\mu'_{2r-4} + \dots\right] + \dots \\
&= \kappa_1 H_{2r-1}(\mu') + \binom{2r-1}{1}\kappa_2 H_{2r-2}(\mu') + \dots \\
&= \kappa_1 c_{2r-1} + \binom{2r-1}{1}\kappa_2 c_{2r-2} + \dots .
\end{aligned}$$

An analogous result holds for odd subscripts. The proof above corresponds to Thiele's (p. 27) more elementary procedure consisting of successive elimination of the moments by which he finds the first six expressions for $c_r + (r-1)c_{r-2}$ in terms of the cumulants. From the pattern obtained he concludes that

$$c_r + (r-1)c_{r-2} = \sum_{j=0}^{r-1}\binom{r-1}{j} c_{r-1-j}\,\kappa_{j+1}, \tag{4.5}$$

or

$$c_r = c_{r-1}\kappa_1 + (r-1)c_{r-2}(\kappa_2 - 1) + \sum_{j=2}^{r-1}\binom{r-1}{j} c_{r-1-j}\kappa_{j+1}, \quad r = 1, 2, \dots,$$

and $c_0 = 1$. This is the first time that a simple formula is given for the calculation of the coefficients in the Gram–Charlier series.

Setting $\kappa_1 = 0$ and $\kappa_2 = 1$ Thiele finds $c_1 = c_2 = 0$, $c_3 = \kappa_3$, $c_4 = \kappa_4$, $c_5 = \kappa_5$, and

$$c_r = \sum_{j=2}^{r-4}\binom{r-1}{j} c_{r-1-j}\,\kappa_{j+1} + \kappa_r, \quad r = 6, 7, \dots,$$

which gives

$$c_6 = 10c_3\kappa_3 + \kappa_6 = 10\kappa_3^2 + \kappa_6.$$

We have already used these results by writing $g(x)$ in the form (3.7). Thiele remarks that the right side of (4.5) is of the same form as (4.2).

Noting (p. 30) that $\kappa_1(ax+b) = a\kappa_1 + b$ and $\kappa_r(ax+b) = a^r\kappa_r$, $r = 2, 3, \dots$, Thiele (p. 60) states the fundamental formula

$$\kappa_r(a_1x_1 + \dots + a_nx_n) = a_1^r\kappa_r(x_1) + \dots + a_n^r\kappa_r(x_n), \quad r = 1, 2, \dots .$$

For the arithmetic mean he finds $\kappa_r(h_1) = \kappa_r/n^{r-1}$ and $\gamma_r(h_1) = \gamma_r/n^{r/2}$ and thus the asymptotic normality of h_1. Moreover, he proposes to test the significance of h_r in relation to a hypothetical distribution by using the asymptotic normality of $(h_r - \kappa_r)/\sqrt{\kappa_2(h_r)}$ and if necessary to take the higher cumulants of h_r into account.

Thiele realized that his proofs are cumbersome because of the recursive definition of the cumulants. Looking for a more direct definition he (1899) found the formula

$$\exp(\kappa_1 t + \kappa_2 t^2/2! + \dots) = 1 + \mu_1' t + \mu_2' t^2/2! + \dots = \int e^{tx} g(x) dx, \tag{4.6}$$

which is the definition used today. Replacing t by it the moment generating function $M(t)$ on the right side becomes the characteristic function $\psi(t)$ which leads to the alternative definition (2.3).

By means of (4.6) Thiele derives the direct expression for κ in terms of μ' and *vice versa*. He also proves the recursion formula. Setting $t = -D$, where D denotes differentiation, and operating on the normal distribution with mean ξ and variance σ^2 he obtains

$$g(x) = \exp[-(\kappa_1 - \xi)D + (\kappa_2 - \sigma^2)D^2/2! - \kappa_3 D^3/3! + \dots](2\pi\sigma^2)^{-\frac{1}{2}} \exp[-(x-\xi)^2/2\sigma^2],$$

which is the Gram–Charlier series in symbolic form. Finally, he proves that the operator $\exp[\alpha_r(-D)^r]$ applied to a density with cumulant κ_r increases κ_r by α_r so that "any law of error may be derived from anyone else with the same mean and standard deviation by the operation"

$$g^*(x) = \exp[-(\kappa_3^* - \kappa_3)D^3/3! + (\kappa_4^* - \kappa_r)D^4/4! - \dots]g(x).$$

Instead of proving these results we shall let Thiele speak for himself through the translation of his paper given in the Appendix.

In the English version of his textbook Thiele (1903) starts from the definition (4.6) which he uses for simplifying the proofs from 1889. He does not mention the operator $M(-D)$.

The results in Thiele's 1899 paper were rediscovered by Cornish & Fisher (1937); they do not refer to Thiele. In the fourth edition of *Statistical Methods* (1932) and in later editions Fisher gives a misleading evaluation of Thiele's theory of the cumulants. An even greater distortion may be found in Fisher's letter to Hotelling of 1 May 1931, published by Bennett (1990, pp. 320–321).

Independently of Thiele, Hausdorff (1901) defines the cumulants, which he calls "canonical parameters", by (4.6) and derives the same results as Thiele, except for the operational properties depending on $M(-D)$. A survey of Hausdorff's contributions to probability theory is due to Girlich (1996).

The contributions of Charlier and Edgeworth have been discussed in the fundamental paper by Cramér (1928), see also Särndal (1971) and Cramér (1972).

Acknowledgement

I thank Professor E. Arjas for suggestions that have improved the presentation of the paper.

References

Bennett, J.H. (Ed.) (1990). *Statistical Inference and Analysis. Selected Correspondence of R.A. Fisher*. Oxford: Clarendon Press.

Bessel, F.W. (1838). Untersuchungen über die Wahrscheinlichkeit der Beobachtungsfehler. *Astron. Nachrichten*, **15**, No. 358–359, 369–404.

Bienaymé, I.J. (1852). Sur la probabilité des erreurs d'aprés la méthode des moindres carrés. *J. Math. Pures Appl.*, **17**(1), 33–78.

Bru, B. (1991). A la recherche de la démonstration perdue de Bienaymé. *Math. Inf. Sci. hum.*, **29**, 5–17.

Chebyshev, P.L. (1855). Sur les fractions continues [in Russian]. *J. Math. Pures Appl.*, (1858) **3**, 289–323 [in French].

Chebyshev, P.L. (1859a). Sur l'interpolation par la méthode des moindres carrés. *Mém. Acad. Sci. St. Pétersbourg*, **1**(7), No. 15, 1–24.

Chebyshev, P.L. (1859b). Sur le développement des fonctions à une seule variable. *Bull. phys.-math. Acad. Sci. St.Pétersbourg*, **1**, 193–200.

Chebyshev, P.L. (1887). Sur deux théorèmes relatifs aux probabilités. *Bull. phys.-math. Acad. Sci. St. Pétersbourg*, **55** [in Russian]. *Acta Math.* (1890–1891), **14**, 305–315 [in French].

Cornish, E.A. & Fisher, R.A. (1937). Moments and cumulants in the specification of distributions. *Rev. Inst. Intern. Statist.*, **5**, 307–320.
Cramér, H. (1928). On the composition of elementary errors. *Skand. Aktuarietidskr.*, **11**, 13–74, 141–180.
Cramér, H. (1972). On the history of certain expansions used in mathematical statistics. *Biometrika*, **59**, 205–207.
Fisher, R.A. (1929). Moments and product moments of sampling distributions. *Proc. London Math. Soc.*, Ser. 2, **30**, 199–238.
Fisher, R.A. (1932). *Statistical Methods for Research Workers*. Fourth edition. Edinburgh: Oliver and Boyd.
Fisher, R.A. & Wishart, J. (1931). The derivation of the pattern formulae of two-way partitions from those of simpler patterns. *Proc. London Math. Soc.*, Ser. 2, **33**, 195–208.
Girlich, H.-J. (1996). Hausdorffs Beiträge zur Wahrscheinlichkeitstheorie. In *Felix Hausdorff zum Gedächtnis*, Ed. E. Brieskorn, Band I, 31–70. Opladen: Westdeutscher Verlag.
Gnedenko, B.V. & Kolmogorov, A.N. (1954). *Limit Distributions for Sums of Independent Variables*. Cambridge, Mass.: Addison–Wesley.
Gram, J.P. (1879). *Om Rækkeudviklinger, bestemte ved Hjælp af de mindste Kvadraters Methode*. Kjøbenhavn: Høst.
Hald, A. (1981). T.N. Thiele's contributions to statistics. *Intern. Statist. Rev.*, **49**, 1–20.
Hald, A. (1998). *A History of Mathematical Statistics from 1750 to 1930*. New York: Wiley.
Hausdorff, F. (1901). Beiträge zur Wahrscheinlichkeitsrechnung. *Ber. Verh. Kgl. Sächs. Ges. Wiss. Leipzig*, **53**, 152–178.
Hermite, C. (1859). Sur l'interpolation. *C. R. Acad. Sci. Paris*, **48**, 62–67.
Hermite, C. (1864). Sur un nouveau développement en série des fonctions. *C. R. Acad. Sci. Paris*, **58**, 93–100, 266–273.
Hotelling, H. (1933). Review of *Statistical Methods for Research Workers*, by R.A. Fisher. Fourth Edition. Edinburgh: Oliver and Boyd. 1932, xiii, 307 pp. *J. Amer. Statist. Assoc.*, **28**, 374–375.
Laplace, P.S. (1810). Mémoire sur les approximations des formules qui sont fonctions de très grands nombres et sur leur application aux probabilités. *Mém. Acad. Sci. Paris*, 1809, 353–415.
Laplace, P.S. (1811). Mémoire sur les intégrales définies et leur application aux probabilités, et spécialement à la recherche du milieu qu'il faut choisir entre les résultats des observations. *Mém. Acad. Sci. Paris*, 1810, 279–347.
Laplace, P.S. (1812). *Théorie analytique des probabilités*. Paris: Courcier, 2nd ed. 1814, 3rd. ed. 1820.
Laplace, P.S. (1816). Sur l'application du calcul des probabilités à la philosophie naturelle. *Théorie analytique. Premier supplément.*
Molina, E.C. (1930). The theory of probability: Some comments on Laplace's *Théorie analytique*. *Bull. Amer. Math. Soc.*, **36**, 369–392.
Poisson, S.D. (1829). Suite du Mémoire sur la probabilité du résultat moyen des observations, inséré dans la Connaissance des Temps de l'année 1827. *Conn. des Temps pour 1832*, 3–22.
Poisson S.D. (1837). *Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précedées des règles générales du calcul des probabilités*. Paris: Bachelier.
Särndal, C.-E. (1971). The hypothesis of elementary errors and the Scandinavian school in statistical theory. *Biometrika*, **58**, 375–392.
Stigler, S.M. (1983). A modest proposal: a new standard for the normal. *Amer. Statistician*, **36**, 137–138.
Thiele, T.N. (1873). Om en Tilnærmelsesformel. *Tidsskrift for Math.*, Tredie Række, **3**, 22–31.
Thiele, T.N. (1889). *Almindelig Iagttagelseslære: Sandsynlighedsregning og mindste Kvadraters Methode*. Kjøbenhavn: Reitzel.
Thiele, T.N. (1899). Om Iagttagelseslærens Halvinvarianter. *Kgl. danske Videnskabernes Selskabs Forhandlinger*, 1899, Nr. 3, 135–141.
Thiele, T.N. (1903). *Theory of Observations*. London: Layton.
Uspensky, J.V. (1937). *Introduction to Mathematical Probability*. New York: McGraw-Hill.

Résumé

L'histoire ancienne de la série de Gram et Charlier est discutée selon trois points de vue: (1) une généralisation du théorème central-limite de Laplace, (2) une approximation par la méthode des moindres carrés d'une fonction continue au moyen de polynômes de Chebyshev et Hermite, (3) une généralisation de la distribution normale de Gauss à une système de distributions asymétriques. Thiele a défini les cumulants en fonction des moments, d'abord par une formule de récurrence, puis par un développement du logarithme de la fonction génératrice des moments. Il a construit un opérateur différentiel qui ajuste un cumulant quelconque à une valeur desirée. Son article peu connu de 1899, portant sur les propriétes des cumulants, est traduit du danois en anglais en annexe.

Mots clés: Bessel; Bienaymé; Théorème central-limite; Chebyshev; Cumulants; Gram; Série de Gram et Charlier; Semi-invariants; Hausdorff; Hermite; Laplace; Moindres carrés; Moments; Polynômes orthogonaux; Poisson; Thiele.

Appendix

T.N. Thiele: Om Iagttagelseslærens Halvinvarianter

Oversigt over Det kgl. danske Videnskabernes Selskabs Forhandlinger,

1899, Nr. 3

Translated into English by A. Hald

On the Halfinvariants in the Theory of Observations

By T.N. Thiele

Read at the Meeting 10 March 1899.

I have in several works on the theory of observations shown that many and great advantages are obtained by expressing the laws of error by certain symmetric functions—the halfinvariants—of the repeated observations. But I have not failed to notice that my use of these functions also had its drawbacks.

In particular, the definition I had given of the halfinvariants was imperfect because it was both indirect and complicated; indirect because it demanded another system of symmetric functions, the sums of powers, inserted as intermediate between the observations and the halfinvariants, and complicated because it only in the form of recurrent equations was simple enough to be remembered and used in proofs.

It was a no lesser defect that a relation between the halfinvariants and the frequency functions was lacking. It must be recognized that the frequency functions are the most direct and in certain cases the most advantageous expression for the law of error, and previously they have nearly always been employed as if the frequency function was the only possible mathematical expression for laws of error.

These defects I can now remedy. The relation between the halfinvariants μ_r and the sums of power s_r can be written as the following identical equation in the variable z

$$s_0 e^{\frac{\mu_1}{\underline{|1}}z+\frac{\mu_2}{\underline{|2}}z^2+\frac{\mu_3}{\underline{|3}}z^3+\dots} = s_0 + \frac{s_1}{\underline{|1}}z + \frac{s_2}{\underline{|2}}z^2 + \frac{s_3}{\underline{|3}}z^3 + \dots \tag{1}$$

which means that the μ-series equals the logarithm of the s-series. In this identity the right-hand side may be resolved into a sum in which every term depends only on one of the repeated observations $o_1, o_2, \dots, o_{s_0}$; hence we have

$$s_0 e^{\frac{\mu_1}{\underline{|1}}z+\frac{\mu_2}{\underline{|2}}z^2+\frac{\mu_3}{\underline{|3}}z^3+\dots} = e^{o_1 z} + e^{o_2 z} + \dots e^{o_{s_0} z}, \tag{2}$$

which is excellently suited as a definition of the halfinvariants μ_r.

From the identity (1) my previous systems of equations may be derived by the method of undetermined coefficients. Comparing directly the coefficients of z^i the explicit equations for s_i will result:

$$\frac{1}{s_0}\frac{s_i}{\underline{|i}} = \sum_{r=1}^{r=i}\sum \frac{1}{\underline{|a}}\left(\frac{\mu_\alpha}{\underline{|\alpha}}\right)^a \frac{1}{\underline{|b}}\left(\frac{\mu_\beta}{\underline{|\beta}}\right)^b \cdots \frac{1}{\underline{|d}}\left(\frac{\mu_\delta}{\underline{|\delta}}\right)^d,$$

$$\text{where } i = a\alpha + b\beta + \cdots + d\delta$$
$$r = a + b + \cdots + d.$$

Taking the logarithms of both sides of the identity (1) the explicit equations for μ_i will result:

$$\frac{\mu_i}{\underline{|i}} = \sum_{r=1}^{r=i} (-1)^{r-1} \underline{|r-1} \sum \frac{1}{\underline{|a}} \left(\frac{s_\alpha}{s_0 \underline{|\alpha}} \right)^a \cdot \frac{1}{\underline{|b}} \left(\frac{s_\beta}{s_0 \underline{|\beta}} \right)^b \cdots \frac{1}{\underline{|d}} \left(\frac{s_\delta}{s_0 \underline{|\delta}} \right)^d,$$
$$i = a\alpha + b\beta + \cdots + d\delta$$
$$r = a + b + \cdots + d.$$

Finally, differentiating the identity (1) with respect to z before comparing coefficients of z^i we find

$$\left(\frac{s_0}{\underline{|0}} + \frac{s_1}{\underline{|1}} z + \ldots \right) \left(\frac{\mu_1}{\underline{|0}} + \frac{\mu_2}{\underline{|1}} z + \ldots \right) = \frac{s_1}{\underline{|0}} + \frac{s_2}{\underline{|1}} z + \ldots$$

resulting in the recursive system of equations

$$s_1 = s_0 \mu_1$$
$$s_2 = s_1 \mu_1 + s_0 \mu_2$$
$$s_3 = s_2 \mu_1 + 2 s_1 \mu_2 + s_0 \mu_3$$
$$\text{etc.}$$

whose relative simplicity and analogy to the binomial formulae allowed me in my previous lack of anything better to use it as the definition of the halfinvariants.

Since both s_r and μ_r are constants for a law of error it is tempting occasionally to use the complete arbitrariness of z to interpret this variable as a symbol of operation; the only danger is that it is impossible *a priori* to secure the convergence of an infinite series in powers of an operator. But otherwise the symbol of differentiation D is to be recommended for operations on laws of error because of the definition (2), since as is well-known $e^{aD} * f(x) = f(x+a)$; but for not to produce sums but differences between x and the observed numbers o_i it will be practical to set $z = -D$ instead of $z = D$ in our definition.

In particular we shall use the identical symbols according to equation (2) to operations on frequency functions trying to derive one function from another. As a first example we shall naturally choose the general typical or exponential frequency function $\exp\left[-\frac{1}{2}\left(\frac{x-m}{n}\right)^2\right]$ with mean m and standard error n. Since this law of error is continuous we have to assume about the values o_i in (2) that their number s_0 is infinite and that the frequency of a value between o and $o + do$ is $\Phi(o)do$, where also the frequency function $\Phi(o)$ must be assumed continuous.

From (2) follows the symbolic equation

$$e^{-\frac{\mu_1}{\underline{|1}} D + \frac{\mu_2}{\underline{|2}} D^2 - \frac{\mu_3}{\underline{|3}} D^3 + \ldots} * e^{-\frac{1}{2}\left(\frac{x-m}{n}\right)^2} = \int_{-\infty}^{+\infty} \Phi(o) e^{-oD} * e^{-\frac{1}{2}\left(\frac{x-m}{n}\right)^2} do$$
$$= \int_{-\infty}^{+\infty} \Phi(o) e^{-\frac{1}{2}\left(\frac{x-m-o}{n}\right)^2} do.$$

Furthermore we shall assume that $\Phi(o) = \Phi(x - m + (o + m - x))$ can be developed in powers of $o + m - x$ according to Taylor's series. Inserting this in the right-hand side of the equation, from which the symbols temporarily had disappeared, we again introduce the differentiation symbol D, which is identical with the previous one because m is constant. Hence

$$\int_{-\infty}^{+\infty} \Phi(o) e^{-\frac{1}{2}\left(\frac{x-m-o}{n}\right)^2} do$$
$$= \int_{-\infty}^{+\infty} \left(\Phi(x-m) + \frac{o+m-x}{\underline{|1}} D * \Phi(x-m) + \right.$$
$$\left. + \frac{(o+m-x)^2}{\underline{|2}} D^2 * \Phi(x-m) + \ldots \right) e^{-\frac{1}{2}\left(\frac{x-m-o}{n}\right)^2} d(o+m-x)$$
$$= n\sqrt{2\pi}\left(\Phi(x-m) + \frac{1\cdot n^2}{\underline{|2}} D^2 * \Phi(x-m) + \frac{1\cdot 3\cdot n^4}{\underline{|4}} D^4 * \Phi(x-m) + \ldots\right)$$
$$= n\sqrt{2\pi} e^{-mD+\frac{n^2}{2}D^2} * \Phi(x).$$

so

$$e^{-\frac{\mu_1}{\underline{|1}}D + \frac{\mu_2}{\underline{|2}}D^2 - \frac{\mu_3}{\underline{|3}}D^3 + \ldots} * e^{-\frac{1}{2}\left(\frac{x-m}{n}\right)^2} = n\sqrt{2\pi} e^{-mD+\frac{n^2}{2}D^2} * \Phi(.x). \tag{3}$$

It is thus possible to get such a general law of error as that given by the halfinvariants μ_r symbolically derived from the general typical. In particular, if we dispose of the constants such that $m = \mu_1$, $n^2 = \mu_2$, that is, neither the mean nor the standard error is changed, then we have

$$\Phi(x) = \frac{1}{\sqrt{2\pi\mu_2}} e^{-\frac{\mu_3}{\underline{|3}}D^3 + \frac{\mu_4}{\underline{|4}}D^4 + \ldots} * e^{-\frac{1}{2}\frac{(x-\mu_1)^2}{\mu_2}}. \tag{4}$$

Accordingly it should be possible to derive any law of error from anyone else with the same mean and standard error by the operation

$$\Phi^1(x) = e^{\frac{\mu_3-\mu_3^1}{\underline{|3}}D^3 - \frac{\mu_4-\mu_4^1}{\underline{|4}}D^4 + \ldots} * \Phi(x), \tag{5}$$

of course under the assumption that $\Phi(x)$ is not a function such that any of the necessary expansions will not be convergent.

To investigate whether the derivation also may be extended to changes of the mean and the standard error we shall treat the case for the typical laws of error. From (3) it follows, setting $\mu_3 = \mu_4 = \cdots = \mu_r = 0$, that

$$\begin{aligned}
\Phi(x) &= \frac{1}{n\sqrt{2\pi}} e^{(m-\mu_1)D + \frac{1}{2}(\mu_2 - n^2)D^2} * e^{-\frac{1}{2}\left(\frac{x-m}{n}\right)^2} \\
&= \frac{1}{n\sqrt{2\pi}} e^{\frac{1}{2}(\mu_2-n^2)D^2} * e^{-\frac{1}{2}\left(\frac{x-\mu_1}{n}\right)^2} \\
&= \frac{1}{\sqrt{2\pi\mu_2}} e^{-\frac{1}{2}\frac{(x-\mu_1)^2}{\mu_2}}
\end{aligned} \tag{6}$$

To prove this theorem we have to study the effect of the symbol $\exp(\frac{1}{2}bD^2)$. For an arbitrary function $f(y)$ we obtain

$$e^{\frac{1}{2}bD^2} * f(y) = f(y) + \frac{1}{\underline{|2}} bD^2 * f(y) + \frac{1\cdot 3}{\underline{|4}} b^2 D^4 * f(y) + \ldots,$$

but since

$$\int_{-\infty}^{+\infty} u^{2r+1}e^{-\frac{1}{2}u^2}du = 0 \quad \text{and} \quad \int_{-\infty}^{+\infty} u^{2r}e^{-\frac{1}{2}u^2}du = \sqrt{2\pi}\cdot 1\cdot 3\ldots(2r-1),$$

we get

$$e^{\frac{1}{2}bD^2} * f(y) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}u^2}\left(f(y) + \frac{u\sqrt{b}D * f(y)}{\underline{|1}} + \frac{u^2bD^2 * f(y)}{\underline{|2}} + \ldots\right)du$$
$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}u^2} f(y + u\sqrt{b})du,$$

if the Taylor expansion can be used on $f(x)$. Choosing particularly

$$f(y) = \frac{1}{\sqrt{2\pi a}}e^{-\frac{1}{2}\frac{y^2}{a}},$$

we find

$$e^{\frac{1}{2}bD^2} * \frac{e^{-\frac{1}{2}\frac{y^2}{a}}}{\sqrt{2\pi a}} = \frac{1}{2\pi\sqrt{a}}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(u^2+\frac{(y+u\sqrt{b})^2}{a}\right)}du$$
$$= \frac{1}{\sqrt{2\pi(a+b)}}e^{-\frac{1}{2}\frac{y^2}{a+b}}.$$

The operation given by the symbol $\exp(\frac{1}{2}bD^2)$ applied to the frequency function thus results in an addition of b to the mean square error. This corroborates (6). It has to be remembered, however, that it is essential that the standard error is real for the typical law of error. Hence, if the addition of b to the mean square error should cause a change of sign the validity of the formula ceases. Also the limiting case that the standard error may become equal to zero ought to be excluded. Observations without error do not have a proper law of error.

Closer investigations cause greater doubts regarding the general formulae (3), (4), and (5). Even if they may be useful in special cases it is easy to show that they not always will lead to convergent series. Particularly one could consider to use them for elucidating the importance of individual halfinvariants in relation to the law of error by deriving frequency functions for such laws of error that deviate from the typical form by having only one of the higher halfinvariants different from zero. A law of error where $\mu_1 = m = 0$, $\mu_2 = n^2 = 1, \ldots, \mu_{r-1} = 0$, $\mu_{r+1} = 0, \ldots$, would according to (4) have the frequency function

$$\Phi_r(.x) = \frac{1}{\sqrt{2\pi}}e^{(-1)^r\frac{\mu_r}{\underline{|r}}D^r} * e^{-\frac{1}{2}x^2}$$
$$= \frac{1}{\sqrt{2\pi}}\left(e^{-\frac{1}{2}x^2} + (-1)^r\frac{\mu_r}{\underline{|r}}D^r * e^{-\frac{1}{2}x^2} + \frac{1}{\underline{|2}}\left(\frac{\mu_r}{\underline{|r}}\right)^2 D^{2r} * e^{-\frac{1}{2}x^2} + \ldots\right)$$

but this series is not convergent for $r > 2$, which is particularly apparent when after the differentiations x is put equal to zero. For example

$$\Phi_3(0) = \frac{1}{\sqrt{2\pi}}\left(1 - \frac{1}{\underline{|2}}\left(\frac{\mu_3}{6}\right)^2\cdot 1\cdot 3\cdot 5 + \frac{1}{\underline{|4}}\left(\frac{\mu_3}{6}\right)^4\cdot 1\cdot 3\cdot 5\cdot 7\cdot 9\cdot 11 - \ldots\right)$$

has

$$\frac{u_{r+1}}{u_r} = -\frac{\mu_3^2}{36}\cdot\frac{(6r+1)(6r+3)(6r+5)}{(2r+1)(2r+2)} = \infty \quad \text{for } r = \infty$$

and

$$\Phi_4(0) = \frac{1}{\sqrt{2\pi}}\left(1 + \frac{1}{\underline{|1}}\frac{\mu_4}{24}\cdot 1\cdot 3 + \frac{1}{\underline{|2}}\left(\frac{\mu_4}{24}\right)^2\cdot 1\cdot 3\cdot 5\cdot 7 + \dots\right)$$

has

$$\frac{u_{r+1}}{u_r} = \frac{\mu_4}{24}\cdot\frac{(4r+1)(4r+3)}{r+1} = \infty \quad \text{for } r = \infty.$$

It will be seen that one cannot freely combine arbitrarily chosen values of the halfinvariants $\mu_3, \mu_4, \dots$ to get a usable law of error.

[*Received September 1998, accepted December 1999*]