On the History of Series Expansions of Frequency Functions and Sampling Distributions, 1873-1944

By

Anders Hald

Dedicated to the memory of my teacher
Professor J.F. Steffensen (1873-1961)

Matematisk-fysiske Meddelelser **49**

Det Kongelige Danske Videnskabernes Selskab
The Royal Danish Academy of Sciences and Letters

Commission Agent: C. A. Reitzels Forlag
Copenhagen 2002

Synopsis

We consider series expansions of univariate frequency functions defined on the real axis and with the property that the function and its derivatives or differences disappear at the end of its range. The A series is a linear combination of a continuous frequency function and its derivatives, the B series a linear combination of a discontinuous frequency function and its differences, and the C series expresses the logarithm of a continuous density as a polynomial of even degree. In particular we discuss the series with the normal and the Poisson distributions as leading terms, the normal A series and the Poisson B series. The series are derived (1) as distributions of the sum of a known or an unknown number of random variables with finite moments and (2) as an expansion of a given frequency function in terms of an auxiliary frequency function. In the latter case the main problems are the choice of the leading term of the series and the determination of the coefficients by the method of moments or the method of least squares. The paper gives the historical development of these series in nearly chronological order under the headings: The Danish, German, British, and Swedish schools. The interaction between the development of expansions of arbitrary frequency functions and sampling distributions is discussed. It is shown how the same results independently are obtained by several authors using different methods of proof. Several problems of priority are resolved.

Printed in Denmark by Special-Trykkeriet Viborg a-s
ISSN 0023-3323. ISBN 87-7876-301-0

Contents

1 The A, B, and C series: definitions, notations, and prehistory, 1811–1887

We shall discuss the history of series expansions of the form

$$g(x) = \sum_{j=0}^{\infty} c_j f_j(x)\,, \quad c_0 = 1\,,\ -\infty < x < \infty\,,$$

where $g(x)$ is a given frequency function and $f(x) = f_0(x)$ is another frequency function chosen as a first approximation to $g(x)$. The series is called the A series, when the f's are continuous, and the B series when discontinuous. The terms A and B series were introduced by Charlier (1905b), who (1928) also introduced a C series in which $\log g(x)$ is represented by a polynomial of even degree. We shall mostly discuss the normal A series, also called the Gram-Charlier series, which is a linear combination of the normal density and its derivatives, and the Poisson B series, which is a linear combination of the Poisson frequency function and its differences; when it is clear from the context we leave out the qualifications normal and Poisson. The coefficients in all the series will be denoted by c_j, which thus takes on different values depending on the context.

It is supposed that $g(x)$ is uniquely determined by its moments $\mu_r = E[(x-b)^r]$, $r = 0, 1, \dots, b$ being an arbitrary number, whose value usually is chosen as zero or $E(x)$. The support of $g(x)$ may be an interval on the real line or a set of consecutive integers. It is assumed that $g(x)$ and its derivatives or differences tend to zero for $|x| \to \infty$.

The moments of $f(x)$ are $\nu_r = E[\,(x-b)^r\,]$. Similarly we define the "moments" ν_{rj} of $f_j(x)$ as $\nu_{00} = 1$, $\nu_{0j} = 0$ for $j \geqq 1$, and

$$\nu_{rj} = \int_{-\infty}^{\infty} (x-b)^r f_j(x)\,dx\,, \quad \begin{array}{l} r = 1, 2, \dots \\ j = 0, 1, \dots\,. \end{array}$$

For the discontinuous case the integral is replaced by a sum. Other kinds of symmetric functions such as factorial moments, binomial moments and cumulants will later be introduced according to the usual definitions. In the following we usually leave out the limits of integration and summation when the whole range of x is involved.

In the proofs we are going to discuss, the authors tacitly assume that all moments are finite and that the moment generating function $M(t) = E(e^{xt})$ exists.

We distinguish between expansions of sampling distributions and frequency functions.

The basic theory of the A series as an expansion of an arbitrary frequency function is due to Laplace (1811, art. V), a fact that has been overlooked until pointed out by Molina (1930). In his discussion of a diffusion problem, see Hald (1998, §17.8), Laplace writes the arbitrary initial density of the position of the particle in question, in standardized measure, in the form

$$g(x) = \pi^{-\frac{1}{2}} \exp(-x^2) \sum_{j=0}^{\infty} c_j \pi^{-\frac{1}{2}} \int (x+is)^j \exp(-s^2)\, ds\,,$$

which is obtained from the general solution for $t = 0$.

To simplify the notation we introduce the standardized normal distribution in the classical form

$$\vartheta(x) = \pi^{-\frac{1}{2}} \exp(-x^2)\,, \quad -\infty < x < \infty\,,$$

and the corresponding Hermite polynomials $H_r^*(x)$ defined as

$$\vartheta^{(r)}(x) = (-1)^r \vartheta(x) H_r^*(x)\,,\ r = 0, 1, \dots\,,$$

where

$$H_r^*(x) = \sum_{j=0}^{[r/2]} (-1)^j \frac{r^{(2j)}}{j!} (2x)^{r-2j}\,, \quad r^{(k)} = r(r-1)\cdots(r-k+1)\,, \tag{1.1}$$

and

$$\int H_r^*(x) H_s^*(x) \vartheta(x)\, dx = 0 \text{ for } s \neq r\,, \text{ and } = 2^r r! \text{ for } s = r\,.$$

Expanding $(x+is)^j$ by the binomial theorem and carrying out the integration it follows that

$$\pi^{-\frac{1}{2}} \int (x+is)^j \exp(-s^2)\, ds = 2^{-j} H_j^*(x)\,,$$

which leads to

$$g(x) = \vartheta(x) \sum_{j=0}^{\infty} 2^{-j} c_j H_j^*(x)\,. \tag{1.2}$$

Using integration by parts Laplace proves the orthogonality of the Hermite polynomials, and using this property he finds the coefficients

$$c_j = \frac{1}{j!} \int H_j^*(x) g(x)\, dx = \frac{1}{j!} E[H_j^*(x)]\,, \quad j = 0, 1 \dots\,, \tag{1.3}$$

by which the expansion is fully determined.

The A series as an expansion of a sampling distribution goes back to the extension of the central limit theorem proved by Laplace, Poisson and Bienaymé, see Hald (2000a). In a remarkable paper by Bienaymé (1852) he first introduces the characteristic function $\psi(t) = E(e^{itx})$ as the generating function for the moments $\mu_r = E(x^r)$, whereafter he changes to the exponential form $\exp[\ln \psi(t)]$ because this is more convenient for carrying out the integrations involved by using the inversion formula. Expanding $\ln\psi(t)$ in a power series he finds the coefficients of the first four powers of (it) as

$$\mu_1\,,\ \mu_2-\mu_1^2\,,\ \mu_3-3\mu_2\mu_1+2\mu_1^3\,,\ \mu_4-4\mu_3\mu_1-3\mu_2^2+12\mu_2\mu_1^2-6\mu_1^4\,.$$

The following coefficients are more complicated and he does not give a formula for the general case. Since $\ln\psi(t)$ is the cumulant generating function Bienaymé's coefficients equal the first four cumulants $\kappa_1,\ldots,\kappa_4$. With this notation we shall rewrite the Poisson-Bienaymé extended central limit theorem for the univariate case.

Let $s_n = x_1+\ldots+x_n$ be the sum of n independent random variables with cumulants $\kappa_r(x_i)\,,\ i=1,\ldots,n$, so that

$$\kappa_r(s_n)=\sum_{i=1}^{n}\kappa_r(x_i)=n\bar{\kappa}_r\,,\quad r=1,2,\ldots\,.$$

Introducing the standardized variable

$$t=(s_n-n\bar{\kappa}_1)(2n\bar{\kappa}_2)^{-\frac{1}{2}}\,,$$

and the standardized cumulants

$$\gamma_r^*=\bar{\kappa}_{r+2}/(2\bar{\kappa}_2)^{(r+2)/2}\,,\quad r=1,2,\ldots\,,$$

the density of s_n may be written as the A series

$$\begin{aligned}p(s_n)=(2n\bar{\kappa}_2)^{-\frac{1}{2}}\vartheta(t)\Bigg[1+\frac{\gamma_1^*H_3^*(t)}{3!\,n^{1/2}}+\frac{\gamma_2^*H_4^*(t)}{4!\,n}+\\ \frac{\gamma_3^*H_5^*(t)}{5!\,n^{3/2}}+\frac{1}{6!}\left(\frac{\gamma_4^*}{n^2}+10\frac{\gamma_1^{*2}}{n}\right)H_6^*(t)+\ldots\Bigg]\,.\end{aligned}\tag{1.4}$$

Actually, Bienaymé discusses the multivariate version of the extended central limit theorem so his result is a multivariate A series. He works out the coefficients but does not give the final form of the polynomials. However, the univariate expansion follows easily from Bienaymé's result, as shown by Meyer (1874, Appendix II).

Laplace and his followers derived the asymptotic expansion above to get an approximation to the sampling distribution of the arithmetic mean and the (regression) coefficients in the linear model, that is, s_n is a statistic calculated from a known number of observations. However, the series took on a new significance when Hagen (1837) and Bessel (1838) formulated the hypothesis of elementary errors, saying that an observation may be considered as the sum of a large number of independent elementary errors stemming from different sources and with different unknown distributions. Hence, s_n is interpreted as an observation and

$p(s_n)$ as the corresponding frequency function. A difficulty with this interpretation is that we do not know the measuring process (or other processes considered) in such detail that we can specify the number of elementary errors making up an observation, so it is only the form of $p(s_n)$ that is known. Hagen and Bessel therefore used the expansion only as an argument for considering the normal distribution as a good approximation to empirical error distributions.

Let $s_n = s$, say, be an observation and let us introduce the cumulants of s as parameters in the expansion of $p(s)$. It is easy to see that $p(s)$ becomes equal to (1.4) for $n = 1$. Hence, the expansion of the sampling distribution is a series with coefficients tending to zero for $n \to \infty$, whereas the terms of the expansion of the frequency function all are finite.

An alternative proof of (1.4) is due to Chebyshev (1887). Part of the proof is based on his (1859) approximation to a square-integrable function $F(x)$ by a linear combination of orthogonal polynomials $\Sigma c_j h_j(x)$. Choosing the normal distribution as weight function,

$$\vartheta_k(x) = (k/\pi)^{1/2} \exp(-kx^2)\,,\ h_j(x) = k^{j/2} H_j^*(x\sqrt{k})\,, \quad k > 0\,,$$

and minimizing the expected value of the squared residuals he gets

$$F(x) = \sum_{j=0}^{\infty} k^{j/2} c_j H_j^*(x\sqrt{k})\,, \tag{1.5}$$

and

$$c_j = \int F(x) H_j^*(x\sqrt{k}\,) \vartheta_k(x)\, dx/(2^j k^{j/2} j!)\,. \tag{1.6}$$

His (1887) proof is based on a new method, the method of moments. He begins by quoting an auxiliary theorem, proved in a previous paper: If the first $2m$ moments of an integrable non-negative function $p(y)$ equal the first $2m$ moments of $\vartheta_k(x)$ then

$$\left| \int_{-\infty}^{v} p(y)\, dy - \int_{-\infty}^{v} \vartheta_k(y)\, dy \right| < \varepsilon(m, v)\,, \tag{1.7}$$

where $\varepsilon(m, v)$ for any value of v tends to zero for m tending to infinity. Hence, if the infinitely many moments of $p(y)$ equal the moments of $\vartheta_k(y)$ then $p(y) = \vartheta_k(y)$.

We shall sketch Chebyshev's proof of the central limit theorem. A detailed discussion of the proof with amendments due to Markov is given by Uspensky (1937, Appendix II); see also Maistrov (1974, pp. 202–208) for some comments.

Chebyshev derives the distribution of $y = (x_1 + \ldots + x_n)n^{-1/2}$, $E(x_i) = 0$, under the assumption that the moment generating functions for the independent x's exist. The moment generating function $M_y(t)$ equals the product of the moment generating functions for $x_i n^{-1/2}$, $i = 1, \ldots , n$. Taking logarithms Chebyshev obtains the relation

$$\ln M_y(t) = \sum_{r=2}^{\infty} k_r t^r/r! = \sum_{r=2}^{\infty} \bar{\kappa}_r t^r/[r!\, n^{(r/2)-1}]\,, \tag{1.8}$$

where k_r is the rth cumulant for y and $\bar{\kappa}_r$ the average of the n cumulants of order r for the x's. For $n \to \infty$ it follows from (1.8) that

$$1 + \sum_{r=2}^{\infty} \mu_r t^r / r! = \exp(\bar{\kappa}_2 t^2 / 2) .$$

The right side is the moment generating function for a normally distributed variable with zero mean and variance $\bar{\kappa}_2$. Hence, for $n \to \infty$ the moments of y equal the moments of a normal distribution, and it then follows from (1.7) that the limit distribution of y is normal.

At the end of Chebyshev's paper he briefly states that the above result gives the main term of a series expansion that may be obtained by means of the method given in his 1859 paper. Without proof he states a theorem that implies (1.4). Following his hints we shall construct a proof.

To simplify the notation we standardize the x's by setting the variances equal to 1/2 so that the density of y for $n \to \infty$ equals $\vartheta(y)$. From (1.8) we then get

$$M_y(t) = \exp(t^2/4) \exp\left[\sum_{r=3}^{\infty} \bar{\kappa}_r t^r / [r! n^{(r/2)-1}]\right] . \tag{1.9}$$

Since

$$\exp(t^2/4) = \int e^{yt} \vartheta(y)\, dy ,$$

it is natural to write $p(y)$ as a product $\vartheta(y)F(y)$, and using (1.5) for $k = 1$ we get

$$p(y) = \vartheta(y) \sum_{j=0}^{\infty} c_j H_j^*(y) , \tag{1.10}$$

which is of the same form as (1.4) for $y = s_n n^{-1/2}$ and $\bar{\kappa}_2 = 1/2$. To prove that the coefficients are the same we derive the moment generating function of the right side of (1.10) using that

$$\int e^{yt} \vartheta^{(j)}(y)\, dy = (-t)^j \exp(t^2/4) ,$$

which may be proved by integration by parts. We then get

$$M_y(t) = \exp(t^2/4) \sum_{j=0}^{\infty} c_j t^j .$$

Comparing with (1.9) it follows that $c_0 = 1$, $c_1 = c_2 = 0$, and

$$1 + \sum_{j=3}^{\infty} c_j t^j = \exp\left[\sum_{j=3}^{\infty} \bar{\kappa}_j t^j / (j! n^{(j/2)-1})\right] , \tag{1.11}$$

which is the generating function for the c's in terms of the cumulants. It is easy to calculate $c_3, \dots, c_6$ and to check that (1.10) equals (1.4). However, Chebyshev does not do so, nor does he refer to Poisson and Bienaymé, so perhaps for these reasons this part of his paper was overlooked.

Gnedenko and Kolmogorov's (1954, pp. 191–196) exposition of Chebyshev's proof is misleading. They use characteristic functions and the inversion theorem to prove (1.4) but this method of proof is due to Poisson and Bienaymé, not to Chebyshev. Moreover, they write that "it is natural to collect terms of the same order in n. This then leads to Chebyshev's expansion", but Chebyshev did not do so, this result is due to Edgeworth (1905).

Until about 1870 the applications of statistical theory were mainly based on the binomial and normal distributions. However, the increasing wealth of data in many fields made it clear that the two classical distributions did not suffice for describing the variations encountered. From about 1870 to 1930 many attempts were made to construct systems of distributions that better fitted the variations of observations taken under the same essential conditions and thus considered as homogeneous.

The first of these systems modifies the normal, binomial and Poisson distributions by taking each of these as the main term of a series expansion, an idea that occurred indepently to several "statisticians." The many authors who worked on this problem naturally used different notations and methods of proof. We shall in some degree rewrite their contributions in a uniform notation.

In the discussion of the series $\Sigma c_j f_j(x)$ there are three problems involved: (1) the choice of $f_0(x)$, (2) the relation of $f_j(x)$, $j \geqq 1$, to $f_0(x)$, and (3) the determination of c_j.

The c's may be expressed in terms of the moments by solving the linear equations

$$\mu_r = \sum_{j=0}^{\infty} c_j \nu_{rj}\,, \quad r = 1, 2, \dots\,. \tag{1.12}$$

This fundamental formula is valid for both continuous and discontinuous distributions. The solution is commonly simplified by choosing the f's such that the matrix $\{\nu_{rj}\}$ is lower triangular, which means that c_j becomes a linear combination of $\mu_1, \dots, \mu_j$.

Another approach consists in choosing the f's as orthogonal with respect to the weight function $1/f_0(x)$ and using the method of least squares, which gives

$$c_j = \int [f_j(x)/f_0(x)] g(x)\, dx \Big/ \int [f_j^2(x)/f_0(x)]\, dx\,. \tag{1.13}$$

If $f_j(x) = f_0(x) P_j(x)$, where $P_j(x)$ is a polynomial of degree j, then c_j becomes proportional to $E[P_j(x)]$, which is a linear combination of the first j moments of $g(x)$. Hence, this special case leads to the same result as the special case of (1.12).

For an appropriate choice of the f's the first few terms of the series will often give a good approximation to $g(x)$. However, the partial sum

$$g_m(x) = \sum_{j=0}^{m} c_j f_j(x)\,, \quad m = 1, 2, \dots\,,$$

will not necessarily be a frequency function, $g_m(x)$ may for example take on negative values.

Authors beginning their investigations of frequency functions by deriving the normal A series naturally remark that the extension of the central limit theorem follows by interpreting the variable in question as a sum of n independent random variables.

The French and German authors express the normal density in the form $\vartheta(x)$, whereas the Danish, British and Swedish authors after Thiele (1889, p. 26) and Pearson (1894) use

$$\phi(x) = (2\pi)^{-\frac{1}{2}} \exp(-x^2/2)\,,$$

and define the corresponding Hermite polynomials $H_r(x)$ as

$$\phi^{(r)}(x) = (-1)^r \phi(x) H_r(x)\,, \quad r = 0, 1, \dots\,,$$

where

$$H_r(x) = \sum_{j=0}^{[r/2]} (-1)^j \frac{r^{(2j)}}{2^j j!} x^{r-2j}\,, \tag{1.14}$$

and

$$\int H_r(x) H_s(x) \phi(x)\, dx = 0 \ \text{ for } s \neq r\,, \ \text{ and } \ = r! \text{ for } s = r\,.$$

Turning to the statistical applications of the series it is clear that only a finite number of terms is necessary for describing an empirical distribution consisting of m relative frequencies. We assume that a sample of n observations from the population with frequency function $g(x)$ is distributed on the values $x_1, \dots, x_m$, $m < n$, where the x's are consecutive integers in the discontinuous case and midpoints of class-intervals of unit length in the continuous case. The relative frequency of x_i is denoted by g_i, $i = 1, \dots, m$, $\Sigma g_i = 1$, and the empirical moments by $m_r, r = 0, 1, \dots$. In the continuous case m_r is a consistent estimate of the corresponding moment of the grouped theoretical distribution, so to obtain an estimate of μ_r a correction for grouping is needed. Sheppard (1898) derived the main term of the correction as a function of the length of the class-interval, and independently Bruns (1906a, pp. 174–190) gave the complete solution taking both the position and the length of the class-interval into account, see Hald (2001). The estimate of c_j is obtained by replacing μ_r by the corrected value of m_r. However, many authors only remarked that small class-intervals should be used.

In the following we shall sketch the historical development in chronological order under the headings: The Danish, German, British and Swedish schools. We limit the exposition to the univariate case since no new principles are involved in the extension to multivariate distributions.

We shall not discuss the convergence and asymptotic properties of the series; this has been done by many authors, see Cramér (1928, 1937) and Boas (1949a, 1949b) for results and further references.

The present paper is a continuation, with some overlapping and amendments, of Hald (2000a).

2 The Danish school

Oppermann and Thiele on the normal A series, 1873.

The geodesist G. K. C. Zachariae (1835–1907) gave in his textbook on the method of least squares (1871, pp. 71–92) an account of Bessel's hypothesis of elementary errors and Bessel's proof of (1.4) for symmetric distributions of the elementary errors.

L. H. F. Oppermann (1817–1883), Professor of German and besides working as an actuary, suggested to multiply the normal density function by a power series to obtain a system of skew distributions, see Gram (1879, p. 94).

T. N. Thiele (1838–1910), at the time working as an actuary, later becoming Professor of astronomy, followed this suggestion by presenting (1873) the first (after Laplace) version of the A series

$$\begin{aligned} g(x) &= \sum_{j=0}^{\infty} c_j D^j[\exp(-\pi x^2)]\,,\ D = d/dx\,, \\ &= \exp(-\pi x^2)\sum_{j=0}^{\infty} c_j T_j(x)\,, \end{aligned} \tag{2.1}$$

where

$$T_j(x) = (-1)^j \pi^{j/2} H_j^*(x\sqrt{\pi}\,)\,.$$

Gram's orthogonalization of the linear model, 1879, 1883.

J. P. Gram (1850–1916), an actuary working together with Thiele, considered the A series as a special case of the linear model. Beginning with a model with m independent variables he writes the adjusted value of $y = (y_1, \dots, y_n)'$ in three ways

$$\begin{aligned} \hat{y}^{(m)} &= b_{m1}x_1 + \dots + b_{mm}x_m \\ &= \hat{y}^{(1)} + (\hat{y}^{(2)} - \hat{y}^{(1)}) + \dots + (\hat{y}^{(m)} - \hat{y}^{(m-1)}) \\ &= c_1 h_1 + \dots + c_m h_m\,, \end{aligned} \tag{2.2}$$

where $(x_1, \dots, x_m)$ are linearly independent n-dimensional vectors, $m \leqq n$, the b's are the least squares regression coefficients, and $(h_1, \dots, h_m)$ are orthogonal

n-dimensional vectors, h_r being a linear combination of $x_1, \dots, x_r$. He uses the orthogonality to prove that

$$c_j = h_j' y / h_j' h_j \,, \quad j = 1, 2, \dots ,$$

so c_j does not depend on the number of terms in the model in contradistinction to b_{mj}.

Since

$$c_m h_m = \hat{y}^{(m)} - \hat{y}^{(m-1)} = d_{m1} x_1 + \dots + d_{mm} x_m \,, \quad m = 1, 2, \dots , n \,,$$

say, the problem is to prove the orthogonality of the successive differences and to determine the coefficients of the linear combination.

Gram's algebraic proof is discussed by Hald (1998, §25.4). Setting $h_1 = x_1$ he finds

$$h_r = \sum_{s=1}^{r} A_{rs}^{(r)} x_s = \begin{vmatrix} a_{11} & \dots & a_{1r} \\ \vdots & & \vdots \\ a_{r-1,1} & \dots & a_{r-1,r} \\ x_1 & \dots & x_r \end{vmatrix} \quad r = 2, 3, \dots , \tag{2.3}$$

where $a_{rs} = x_r' x_s$ and $A_{rs}^{(r)}$ denotes the cofactor of a_{rs} in the determinant $A^{(m)} = |a_{rs}|$ of the normal equations. The residual sum of squares equals

$$R_m = (y - \hat{y}^{(m)})'(y - \hat{y}^{(m)}) = y'y - \sum_{r=1}^{m} c_r^2 h_r' h_r \,.$$

To explain the orthogonality in geometrical terms we note that $\hat{y}^{(m)}$ is the projection of y on the space spanned by $(x_1, \dots , x_m)$ so that the residual $y - \hat{y}^{(m)}$ is perpendicular to each of these vectors, that is,

$$x_r'(y - \hat{y}^{(m)}) = 0 \,, \quad r = 1, \dots , m \,.$$

Since

$$\hat{y}^{(m)} - \hat{y}^{(m-1)} = (y - \hat{y}^{(m-1)}) - (y - \hat{y}^{(m)})$$

it follows that

$$x_r'(\hat{y}^{(m)} - \hat{y}^{(m-1)}) = 0 \,, \quad r = 1, \dots , m-1 \,,$$

from which the orthogonality of the successive differences follows as each difference before the last one is a linear combination of $x_1, \dots , x_r$ for $r < m$.

Gram's decomposition of $\hat{y}^{(m)}$ expresses the fact that the explanatory variable x_1 leads to the adjusted value $\hat{y}^{(1)}$, the two explanatory variables (x_1, x_2) lead to $\hat{y}^{(2)}$ so the net effect of taking x_2 into account is $\hat{y}^{(2)} - \hat{y}^{(1)}$, which is orthogonal to $\hat{y}^{(1)}$, and so on.

Gram assumes that var $(y_i) = \sigma^2/w_i$, where $w_i > 0$ is a known number. This means that the sums of squares and products in the formulas above should be read as, for example, $\Sigma x_{ri} y_i w_i$ instead of $\Sigma x_{ri} y_i$.

Gram considers the special case of (2.2) in which the components of the vectors involved are functions of an independent variable, t say, so that the true value of $y(t)$ is represented by a linear combination of $x_1(t), \dots, x_m(t)$. Generalizing this set-up he seeks a representation of a given square-integrable function $y(t)$ as an infinite series

$$y(t) \sim \sum_{r=1}^{\infty} b_r x_r(t) = \sum_{r=1}^{\infty} c_r h_r(t)\,,$$

where the $x(t)$'s are linearly independent known functions and the coefficients are to be determined by the method of least squares using the known function $w(t) > 0$ as weight. The function $h_r(t)$ is determined from (2.3) with

$$a_{rs} = \int x_r(t) x_s(t) w(t)\, dt\,,$$

so that

$$\int h_r(t) h_s(t) w(t)\, dt = 0 \text{ for } r \neq s\,,$$

$$c_r = \int h_r(t) y(t) w(t)\, dt \Big/ \int h_r^2(t) w(t)\, dt\,,$$

and the residual sum of squares after m terms equals

$$R_m = \int y^2(t) w(t)\, dt - \sum_{r=1}^{m} c_r^2 \int h_r^2(t) w(t)\, dt\,.$$

Gram's orthogonal A series, 1879, 1883.

Gram applies the method above to get an expansion of a continuous frequency function by setting $x_j = f(x) x^j$ so that

$$g(x) = f(x) \sum_{j=0}^{\infty} b_j x^j = f(x) \sum_{j=0}^{\infty} c_j P_j(x)\,, \tag{2.4}$$

where $\{P_j(x)\}$ are orthogonal polynomials determined from (2.3) by means of

$$a_{rs} = \int x^{r+s} f^2(x) w(x)\, dx\,, \quad (r,s) = 0, 1, \dots\,.$$

In particular, Gram studies this series for $w(x) = 1$ and $w(x) = 1/f(x)$. In the following we shall only discuss the latter case for which

$$a_{rs} = \int x^{r+s} f(x)\, dx = \nu_{r+s}\,,$$

so that $P_0(x) = 1$,

$$P_j(x) = \begin{vmatrix} \nu_0 & \nu_1 & \dots & \nu_j \\ \nu_1 & \nu_2 & \dots & \nu_{j+1} \\ \vdots & \vdots & & \vdots \\ \nu_{j-1} & \nu_j & \dots & \nu_{2j-1} \\ 1 & x & \dots & x^j \end{vmatrix} \quad j = 1, 2, \dots, \tag{2.5}$$

$$\int P_r(x) P_s(x) f(x)\, dx = 0 \ \text{ for } \ r \neq s\,, \tag{2.6}$$

and

$$c_j = \int P_j(x) g(x)\, dx \Big/ \int P_j^2(x) f(x)\, dx\,, \quad j = 0, 1, \dots\,. \tag{2.7}$$

This is Gram's fundamental result for a series for which $f_j(x) = f(x)P_j(x)$ and $w(x) = 1/f(x)$. Since $P_j(x)$ is a polynomial of degree j it follows that c_j is a linear combination of the moments of $g(x)$ of order 1 to j. Hence, the method of least squares with $1/f(x)$ as weight leads to the method of moments.

As special cases Gram (1879) chooses $f(x)$ as the uniform distribution over a finite interval (pp. 45–48), the gamma distribution (pp. 60–66), and the normal distribution (pp. 67–72), the corresponding polynomials being related to the Legendre, Laguerre, and Hermite polynomials. His main results are reproduced in the German version (1883) of his paper. From the moments of the three distributions Gram calculates $P_1(x)$, $P_2(x)$, $P_3(x)$ by (2.5) and by induction he finds the general formulas. In each case he checks the orthogonality using integration by parts.

The series based on the uniform distribution has been discussed by Hald (1998, pp. 544–545). Here we relate Gram's results with the normal and the gamma distributions as leading terms.

Gram sets $f(x) = \vartheta(x)$ and proves that

$$\vartheta(x) P_j(x) = D^j \vartheta(x) = \vartheta(x)(-1)^j H_j^*(x)\,,$$

which leads to the A series

$$g(x) = \vartheta(x) \sum_{j=0}^{\infty} (-1)^j c_j H_j^*(x) \tag{2.8}$$

with

$$\begin{aligned} c_j &= (-1)^j \frac{1}{2^j j!} E[H_j^*(x)] \\ &= (-1)^j \frac{1}{j!} \sum_{k=0}^{[j/2]} (-1)^k \frac{j^{(2k)}}{2^{2k} k!} \mu_{j-2k}\,, \end{aligned} \tag{2.9}$$

according to (2.7) and (1.1).

For the gamma distribution

$$f(x) = x^{\alpha-1}e^{-\beta x}\beta^{\alpha}/\Gamma(\alpha)\,, \quad \alpha > 0\,,\ \beta > 0\,,\ x > 0\,, \tag{2.10}$$

Gram proves that

$$f(x)P_j(x) = D^j[f(x)x^j]\,,$$

and carrying out the differentiation by means of Leibniz's formula he obtains

$$P_j(x) = \sum_{k=0}^{j}(-1)^k\binom{j}{k}(j+\alpha-1)^{(j-k)}(\beta x)^k\,, \quad j = 0, 1, \ldots\,. \tag{2.11}$$

After having proved the orthogonality of these polynomials he gets

$$g(x) = f(x)\sum_{j=0}^{\infty} c_j P_j(x)\,,$$

where

$$\begin{aligned} c_j &= \frac{1}{j!(j+\alpha-1)^{(j)}}E[P_j(x)] \\ &= \sum_{k=0}^{j}(-1)^k\frac{\beta^k}{k!(j-k)!(\alpha+k-1)^{(k)}}\mu_k\,, \quad j = 0, 1, \ldots\,. \end{aligned} \tag{2.12}$$

Gram remarks that it is often useful to replace x in the series (2.4) by $\alpha(x+\beta)$, for example in the normal A series (2.8). It is then possible to choose α and β such that two of the coefficients in the series disappear. It is easy to see that $c_1 = c_2 = 0$ if α and β are chosen such that the first two moments of f and g are equal.

In accordance with his general principle Gram maintains that by fitting a partial sum of the series to an empirical distribution the method of least squares should be used to determine α and β. This leads, however, to complications because the model no longer is linear in the parameters (unless f is a constant) so the solution has to be obtained by iteration. As a first approximation α and β are estimated by means of the first two sample moments, and these estimates may be improved by taking the third moments into regard. Gram recommends to use the first or second approximation as if they were the true values of α and β and then proceed accordingly to estimate the c's.

As an example (1879, pp. 105–107) Gram fits a gamma distribution to the distribution of the marriage age for hitherto unmarried men during the observation period 1855–1869. Taking the origin at 17.5 years of age he estimates the parameters in the gamma distribution (2.10) by means of the first two sample moments about the origin, m_1 and m_2 say, by solving the equations

$$\alpha = \beta m_1 \quad \text{and} \quad \alpha + 1 = \beta m_2/m_1\,.$$

He observes that the fit is not quite satisfactory but does not go on to find the following terms of the series.

This example seems to be the first formulation and application of the gamma distribution.

It is clear that an orthogonal B series, analogous to the A series, may be constructed by Gram's method, using sums instead of integrals. However, Gram discusses only the case for $f(x)$ constant for which he develops an arbitrary function $g(x)$, defined on a finite number of equidistant points, in terms of orthogonal polynomials. He points out that a considerable simplification is obtained by using factorials instead of powers of x.

As an example he first considers the orthogonal A series for an arbitrary continuus function defined on a finite interval which he normalizes to $[0,1]$. For $f(x)=1$ he gets $g(x)=\Sigma c_j P_j(x)$, where

$$P_j(x)=\sum_{k=0}^{j}(-1)^k\binom{j}{k}\binom{j+k}{k}x^k\,, \quad 0\leqq x\leqq 1\,.$$

For the discrete case he assumes that $x=0,1,\ldots,n-1$ and obtains the expansion of $g(x)$ by replacing x^k in $P_j(x)$ by $x^{(k)}/(n-1)^{(k)}$, see Hald (1998, pp. 544–547) for the proof. He uses this result as a polynomial regression, not as a representation of a frequency function.

Thiele's representation of the A series by means of the cumulants, 1889, 1899, 1903.

Thiele (1889, pp. 26–28) improves Gram's proof by introducing $\phi(x)$ instead of $\vartheta(x)$, by using the orthogonality directly instead of the method of least squares for finding c_j, and by introducing the cumulants instead of the moments. He writes the series as

$$\begin{aligned} g(x) &= \sum_{j=0}^{\infty}(-1)^j\frac{1}{j!}c_jD^j\phi(x) \\ &= \phi(x)\sum_{j=0}^{\infty}\frac{1}{j!}c_jH_j(x)\,. \end{aligned}$$

Multiplying by $H_r(x)$, integrating, and using the orthogonality of the H's he finds $c_r=E[H_r(x)]$, which gives c_r in terms of the moments by replacing x^{r-2j} by μ_{r-2j} in (1.14). By means of his formula for the moments in terms of the cumulants he finds $c_1,\ldots,c_8$ as functions of the cumulants, and using the recursion formula for the Hermite polynomials he derives a recursion formula for the c's. Replacing x by the standardized variable $u=(x-\kappa_1)\kappa_2^{-\frac{1}{2}}$ the final form of the series becomes

$$\begin{aligned} g(x)=\kappa_2^{-\frac{1}{2}}\phi(u)[1&+\gamma_1H_3(u)/3!+\gamma_2H_4(u)/4!+ \\ &\gamma_3H_5(u)/5!+(\gamma_4+10\gamma_1^2)H_6(u)/6!+\ldots], \end{aligned} \tag{2.13}$$

where

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

see (1889, p. 28; 1903, p. 35).

Ten years later Thiele (1899; 1903, p. 24) defined the relation between the cumulants and the moments by equating the cumulant generating function to the logarithm of the moment generating function, that is,

$$\sum_{j=1}^{\infty} \kappa_j t^j/j! = \ln \int e^{xt} g(x)\,dx = \ln\left(1 + \sum_{j=1}^{\infty} \mu_j t^j/j!\right)\,, \tag{2.14}$$

from which he (1899), without giving the proof, finds

$$\frac{\mu_j}{j!} = \sum_{r=1}^{j} \sum \frac{1}{a!}\left(\frac{\kappa_\alpha}{\alpha!}\right)^a \frac{1}{b!}\left(\frac{\kappa_\beta}{\beta!}\right)^b \cdots \frac{1}{d!}\left(\frac{\kappa_\delta}{\delta!}\right)^d\,, \quad j \geqq 1\,, \tag{2.15}$$

where $j = a\alpha + b\beta + \ldots + d\delta$ and $r = a + b + \ldots + d$, and

$$\frac{\kappa_j}{j!} = \sum_{r=1}^{j} (-1)^{r-1}(r-1)! \sum \frac{1}{a!}\left(\frac{\mu_\alpha}{\alpha!}\right)^a \frac{1}{b!}\left(\frac{\mu_\beta}{\beta!}\right)^b \cdots \frac{1}{d!}\left(\frac{\mu_\delta}{\delta!}\right)^d\,, \quad j \geqq 1\,. \tag{2.16}$$

He (1899) also derives the symbolic form of the A series as

$$g(x) = \exp\left(-\kappa_3 D^3/3! + \kappa_4 D^4/4! - \ldots\right) f(x)\,, \quad D = d/dx\,, \tag{2.17}$$

where $f(x)$ denotes a density with mean κ_1 and variance κ_2, see Hald (2000a).

To evaluate the right side of (2.14) Thiele (1903, p. 34), using integration by parts, finds

$$\begin{aligned}\int e^{xt} D_x^j \phi\left(\frac{x-\mu}{\sigma}\right)\frac{dx}{\sigma} &= (-t)^j \int e^{xt} \phi\left(\frac{x-\mu}{\sigma}\right)\frac{dx}{\sigma}\\ &= (-t)^j \exp\left(\mu t + \sigma^2 t^2/2\right).\end{aligned}$$

Inserting the series for $g(x)$ and carrying out the integrations he obtains

$$\sum_{j=1}^{\infty} \kappa_j t^j/j! = \mu t + \sigma^2 t^2/2 + \ln \sum_{j=0}^{\infty} c_j t^j/j!\,.$$

Setting $\mu = \kappa_1$ and $\sigma^2 = \kappa_2$ the generating function for the coefficients in the series becomes

$$1 + \sum_{j=3}^{\infty} c_j t^j/j! = \exp\left(\sum_{j=3}^{\infty} \kappa_j t^j/j!\right)\,. \tag{2.18}$$

Independently, Thiele has thus proved Chebyshev's formula (1.11) for $n = 1$. Expanding the right side in a power series and equating coefficients the c's are found. Thiele derives the coefficients up to c_8 and points out that (2.18) is of the same form as (2.14) so that the relation between the c's and the κ's is analogous to the one between the μ's and the κ's, but he does not present the explicit solution. However, this is readily found from (2.15), which shows that

$$\frac{c_j}{j!} = \sum_{r=1}^{[j/3]} \sum \frac{1}{a!}\left(\frac{\kappa_\alpha}{\alpha!}\right)^a \frac{1}{b!}\left(\frac{\kappa_\beta}{\beta!}\right)^b \cdots \frac{1}{d!}\left(\frac{\kappa_\delta}{\delta!}\right)^d, \quad j \geqq 3, \tag{2.19}$$

where $j = a\alpha + b\beta + \ldots + d\delta$ and $r = a + b + \ldots + d$. Hence, c_j is a homogeneous function of the subscripts. As an example we have

$$\frac{c_{10}}{10!} = \frac{1}{2!}\left(\frac{\kappa_3}{3!}\right)^2 \frac{\kappa_4}{4!} + \frac{\kappa_3}{3!}\frac{\kappa_7}{7!} + \frac{\kappa_4}{4!}\frac{\kappa_6}{6!} + \frac{1}{2!}\left(\frac{\kappa_5}{5!}\right)^2 + \frac{\kappa_{10}}{10!}\,.$$

Thiele considered the condensation of the information in a sample by means of a few symmetric functions as one of the main problems in statistics, but which kind of symmetric function should one choose? He discarded the moments compared with the cumulants and looked for an interpretation of these. For a continuous distribution he found this in the A series for which he points out that λ_1 characterizes the skewness and λ_2 the peakedness of the distribution. He (1903, pp. 49–50) concludes that the coefficients of the A series are to be preferred to the cumulants.

In applications he recommends to use the partial sum of the A series with only the first five cumulants as parameters because of the large sanpling error of the following cumulants. He (1889, pp. 62–64) derives the variance of the first four sample cumulants and an approximation for the followings and uses this result to find the corresponding confidence limits.

He remarks that the length of the class-interval should be at most one-fourth of the standard deviation to reduce the effect of grouping.

Hence, the theory and application of the A series are fully discussed in the works of Thiele and Gram.

Thiele's orthogonal B series based on the symmetric binomial, 1889.

Thiele (1889, pp. 9–13; 1903, p. 21) also introduced a B series based on the symmetric binomial and its differences. His exposition implies that the functions $\{f_j(x)\}$ are orthogonal with respect to the weight function $1/f_0(x)$ so

$$\int f_r(x) f_s(x) f_0^{-1}(x)\,dx = \begin{matrix} 0 & \text{for } s \neq r \\ \rho_r & \text{for } s = r\,. \end{matrix}$$

The method of least squares with $1/f_0(x)$ as weight gives

$$c_j = \rho_j^{-1} E[f_j(x)/f_0(x)]\,,$$

which is a generalization of Gram's formula (2.7).

To describe an unnormed frequency function $g(x)$, $x = 0, 1, \dots, m$, $\Sigma g(x) = n$ Thiele uses the binomial coefficient

$$\beta_m(x) = \binom{m}{x}, \quad x = 0, 1, \dots, m, \text{ and } \beta_m(x) = 0 \text{ otherwise,}$$

as leading term and writes the series as

$$g(x) = c_0\beta_m(x) + c_1\nabla\beta_{m-1}(x) + \dots + c_m\nabla^m\beta_0(x), \quad m = 1, 2, \dots, \tag{2.20}$$

where $\nabla\beta_m(x) = \beta_m(x) - \beta_m(x-1)$ and $\Sigma g(x) = 2^m c_0$. By not norming $g(x)$ to unity Thiele obtains that the functions involved take on integer values only. Thiele indicates that the functions

$$\nabla^j\beta_{m-j}(x) = f_j(x), \quad \begin{array}{l} x = 0, 1, \dots, m \\ j = 0, 1, \dots, m, \end{array}$$

are orthogonal with respect to $1/f_0(x)$, but leaves the proof to the reader. He tabulates the matrices $\{f_j(x)\}$ for $m = 1, \dots, 16$ to make the calculations of the coefficients and $g(x)$ easy, the orthogonality of the tabular values is obvious. More details and a proof of the orthogonality are given by Hald (2000b).

Thiele's C series, 1897, 1903.

As a third possibility for representing a frequency function by a series Thiele (1897, pp. 14–15; 1903, p. 16) proposes to use polynomial interpolation on the logarithm of the density, which leads to the C series

$$\ln g(x) = \sum_{j=0}^{2m} c_j x^j, \quad c_{2m} < 0, \quad m = 1, 2, \dots, \quad -\infty < x < \infty. \tag{2.21}$$

Thiele's three examples, 1889, 1897, 1903.

Regarding applications of the three series Thiele (1889, p. 9) writes: "The exact representation of an empirical frequency function with m different results will of course require a series with m terms, but for an approximate representation it is important that the series has been chosen and ordered in such a way that the coefficient of each term can be calculated separately, and that the first terms of the series immediately give the essential characteristics of the function, whereas the later terms more and more lose importance and at last only regard trifles with importance only for a completely detailed representation of the given empirical distribution."

To demonstrate the applications he fits the three series to the same data, viz. 500 observations from a game of patience, and obtains nearly the same goodness of fit, see Table 1. He judges the goodness of fit by looking at the differences between the observed and calculated frequencies. He does not comment on the fact that the A series leads to negative frequencies for $x = 5$ and 6.

Table 1. Thiele's fitting of the A, B, and C series to an observed distribution

No. of points	Observed	Calculated A series	B series	C series
4	0	0.0		
5	0	−0.1		
6	0	−0.3		
7	3	1.6	1	0.3
8	7	12.3	11	7.1
9	35	39.6	40	39.2
10	101	78.2	82	85.9
11	89	104.1	103	105, 4
12	94	97.7	92	93.4
13	70	69.4	70	70.5
14	46	42.8	48	48.5
15	30	26.7	26	29.8
16	15	16.0	13	14.5
17	4	8.0	8	4.6
18	5	3.0	4	0.7
19	1	0.8	1	0.0
20	0	0.2		
21	0	0.0		
Total	500	500.0	499	499.9

Sources:

Observed: Thiele (1889, p. 12; 1897, p. 12; 1903, p. 13)

A series: Thiele (1903, pp. 50–51). $u = (x - 11.86)/2.0408$.

$g_5(x) = \frac{\phi(u)}{2.0408}\{1 + 0.09233H_3(u) + 0.009356H_4(u) - 0.006344H_5(u)\}$.

B series: Thiele (1889, p. 12). $f_0(x) = \beta_{12}(x)$.

$g_5(x) = 0.1221f_0(x) + 0.278f_1(x) + 0.600f_2(x) + 0.216f_3(x) + 0.278f_4(x) - 0.318f_5(x)$,

$x = 0, 1, \ldots, 12$. No. of points $= x + 7$.

C series: Thiele (1897, p. 12; 1903, pp. 13–14).

$$\log g_4(x) = 2.0228 + 0.0030(x - 11) - 0.06885(x - 11)^2$$
$$+0.01515(x - 11)^3 - 0.001678(x - 11)^4 .$$

Thiele introduced the B series to describe discontinuous distributions and the A and C series for the continuous case. Nevertheless he used all three series for analysing the same sample from a discrete population. He did not make a direct comparison of the three series as we have done in Table 1, which we have included to illustrate the following problem: If an empirical distribution is described equally well by several different models, which model should be chosen as the "best"?

In his discussion of this problem for continuous distributions Thiele (1903, p. 22) writes: "[...] that we certainly possess good instruments by means of which we can even in more than one form find general series adapted for the representation of laws of errors.[...] If anything, we have too many forms and too few means of estimating their value correctly.[...] We ask in vain for a fixed rule, by which we can select the most important and trustworthy forms with limited numbers of constants, to be used in predictions."

Thiele states explicitly that among the many possible forms he prefers the four- or five-parameter A series, presumably because of its flexibility, its mathematical and computational simplicity, and the simple interpretation of the four parameters for describing the frequency curve. With their actuarial background Thiele and Gram looked at the problem as one of graduation, and they abstained from speculations about the genesis of the model.

3 The German school

Fechner's Kollektivmasslehre, and the Fechner distribution, 1897.

The background for the German school is the posthumously published *Kollektivmasslehre* (1897) by G. T. Fechner (1801–1887), physicist and psychologist, co-author of the Weber-Fechner law and founder of the discipline psychophysics (experimental psychology), see Stigler (1986, pp. 242–254) on Fechner's multifactor experiments on the stimulus-sensation relation and their statistical analysis by the method of quantal response, and Heidelberger (1987) on Fechner's indeterminism and the *Kollektivmasslehre*.

After Fechner's death the incomplete manuscript to his book was edited and completed by G. F. Lipps (1865–1931), philosopher, psychologist and mathematician. Many of Fechner's propositions are based on empirical investigations, for example by means of random numbers from Saxon lotteries, and Lipps provides the corresponding mathematical proofs and also supplementary empirical data and analyses. Lipps's contributions are so essential that he ought to have figured as co-author of the book.

Fechner (1897, p. 3) defines a collective as "an object consisting of an indefinite number of randomly varying specimens that belong to the same species or genus." A collective is described by means of a frequency function so that "Kollektivmasslehre" in modern terminology becomes the theory of frequency functions. Fechner observes that most frequency functions encountered outside the physical sciences are asymmetric and he aims at supplementing the classical error theory taking this fact into account. He discusses only continuous distributions. His book contains a large number of empirical distributions from lotteries, astronomy, anthropology, botany, meteorology and dimensions of paintings, and it became a challenge for him and his followers to find corresponding theoretical distributions. Comparing the empirical and calculated frequencies he measured the goodness of fit by the sum of the absolute deviations. He (1897, pp. 4–5) states his program as follows: "Does there exist a general law or at least a law applicable for most collectives for the relation between the numbers and the sizes of the specimens? Actually it is possible to obtain such a law and it is the main task in the following to establish it."

Fechner solves this problem in two steps. First (1897, pp. 69–70, 295–299) he generalizes the normal distribution to the "two-sided normal", which is a composition of two normal distributions with different standard deviations and common mode, see Hald (1998, pp. 378–380). He recommends this distribution for describing moderately skew data. Second (1897, pp. 339–351) for more extreme skewness he uses the logarithm of the variable as two-sided normal. For variables taking on only positive values he considers the logarithmic form as fundamental, the arithmetic form being a useful approximation if the relative variation is small. He fits a normal and a two-sided normal distribution to his moderately skew empirical distributions to demonstrate the improvement in the goodness of fit, and for distributions of greater skewness he compares the fits obtained by using the arithmetic and the logarithmic forms of the two-sided normal.

Lipps (1897) gives a summary of the *Kollektivmasslehre* and indicates that Fechner's solution is insufficient. He says that there are two essentially different methods of solution: (1) the direct method, to seek a (more flexible) formula for the distribution, which has recently been done by Bruns (1897), (2) the indirect method, to transform the random variable such that the corresponding distribution has a specified form.

From about 1897 a lively discussion of the new systems of frequency functions took place among natural scientists in the German speaking countries, see for example Ludwig (1898), Duncker (1899), and Ranke and Greiner (1904). We shall in turn discuss the mathematical contributions in the form of series expansions due to Bruns, Lipps and Hausdorff.

Bruns's derivations of the A series, 1897, 1898, 1906a.

H. Bruns (1848–1919) was Professor of astronomy at the University of Leipzig. His main work in statistics is *Wahrscheinlichkeitsrechnung und Kollektivmasslehre* (1906a), which he characterizes as the first textbook on Kollektivmasslehre in general. For priority reasons, and perhaps also as an excuse for the lack of references to recent literature, he refers in the preface to his previous papers on series expansions of distribution functions and states that the manuscript of the book was ready for printing in the beginning of 1900, but publication was delayed because he at the time contemplated to give a more extensive exposition of applications of the theory.

Bruns states that the application of probability theory presupposes that objects exist that at least approximately realize the concepts of random events and theoretical frequency distributions. That this is so is for the first time demonstrated by Fechner in his *Kollektivmasslehre*. However, Fechner's mathematics is rather primitive and it cannot be expected that an arbitrary frequency function can be approximated by the two-sided normal distribution containing only three parameters. Instead Bruns (1898) proposes to use an expansion of the form $g(x) = \Sigma c_j f^{(j)}(x)$ and, for simplicity, to choose $f(x)$ as normal. He considers this series as the general solution of Fechner's problem.

Bruns (1897, 1898, 1906a) gives three derivations of the A series. Except for some changes of terminology and notation we shall first relate the proof given in 1898 and 1906 and later comment on the first proof.

Let $G(x)$ be the distribution function corresponding to the continuous density $g(x)$ so that

$$2G(x) - 1 = \int_{-x}^{x} g(t)\,dt\,, \quad -\infty < x < \infty\,.$$

Bruns expresses this function as a linear combination of the Gaussian error function

$$\theta(x) = \pi^{-\frac{1}{2}} \int_{-x}^{x} \exp(-t^2)\,dt$$

and its derivatives. Both functions increase from -1 to 1 when x increases from $-\infty$ to ∞. Introducing a scale parameter in $\theta(x)$ and letting this parameter tend to zero Bruns gets the degenerate error function

$$\operatorname{sgn} x = \begin{array}{rl} 1 & \text{for } x > 0 \\ 0 & \text{for } x = 0 \\ -1 & \text{for } x < 0\,. \end{array}$$

It follows that

$$\begin{aligned} E[\operatorname{sgn}(y-x)] &= \int_{-\infty}^{y} g(x)\,dx - \int_{y}^{\infty} g(x)\,dx \\ &= 2G(y) - 1\,, \quad -\infty < y < \infty\,. \end{aligned} \tag{3.1}$$

The problem is thus to find a series expansion for $\operatorname{sgn}(y-x)$.

Noting that

$$\theta'(x) = 2\pi^{-\frac{1}{2}} \exp(-x^2)\,,$$

and using the characteristic function for the normal distribution Bruns gets

$$\theta'(x) = \frac{2}{\pi} \int_{-\infty}^{\infty} \exp(2ixt - t^2)\,dt\,.$$

Integrating with respect to x he obtains

$$\theta(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} \exp(2itx - t^2)\frac{dt}{it}\,, \tag{3.2}$$

and

$$\operatorname{sgn} x = \frac{1}{\pi} \int_{-\infty}^{\infty} \exp(2ixt)\frac{dt}{it}\,.$$

Hence,

$$\operatorname{sgn}(y-x) = \frac{1}{\pi}\int_{-\infty}^{\infty} \exp(2iyt - t^2)\exp(-2ixt + t^2)\frac{dt}{it}\,. \tag{3.3}$$

To evaluate the second exponential factor in the integrand Bruns uses the relation

$$\begin{aligned}\theta'(x+t) &= \theta'(x)\exp(-2xt - t^2)\\ &= \theta'(x)\sum_{j=0}^{\infty} R_j(x)(2t)^j\,,\end{aligned} \tag{3.4}$$

where the power series is found by multiplying the series for $\exp(-2xt)$ and $\exp(-t^2)$, which shows that $R_j(x)$ is a polynomial of degree j. It follows that

$$\exp(-2ixt + t^2) = \sum_{j=0}^{\infty} R_j(x)\,(2it)^j\,, \tag{3.5}$$

which inserted in (3.3) gives

$$\operatorname{sgn}(y-x) = \frac{1}{\pi}\sum_{j=0}^{\infty} R_j(x)\int_{-\infty}^{\infty}\exp(2iyt - t^2)\,(2it)^j\frac{dt}{it}\,.$$

Differentiating (3.2) j times it will be seen that the integral in the series above equals $\pi\theta^{(j)}(y)$ so that

$$\operatorname{sgn}(y-x) = \sum_{j=0}^{\infty} R_j(x)\theta^{(j)}(y)\,. \tag{3.6}$$

Hence,

$$2G(y) - 1 = \sum_{j=0}^{\infty} E[R_j(x)]\theta^{(j)}(y)\,, \tag{3.7}$$

which he (1906a, p. 115) calls "the fundamental formula for the interpolatory representation of an arbitrary distribution function [Kollektivreihe]."
Differentiation gives the A series

$$g(y) = \frac{1}{2}\sum_{j=0}^{\infty} E[R_j(x)]\theta^{(j+1)}(y)\,. \tag{3.8}$$

Differentiating (3.4) with respect to t and setting $t=0$ Bruns finds

$$\theta^{(j+1)}(x) = 2^j j!\theta'(x)R_j(x)\,,$$

so

$$g(y) = \frac{1}{2}\theta'(y)\sum_{j=0}^{\infty} 2^j j! E[R_j(x)]R_j(y)\,. \tag{3.9}$$

He (1898, p. 351; 1906a, p. 43) derives the first five polynomials and notes that the rule of formation from then on is obvious. He finds the differential equation and the recursion formula for the R's, but does not mention the orthogonality. To facilitate the applications of the series he (1898) tabulates $\theta^{(j)}(x)/2^{j-1}$ for $j = 1, \dots, 6$ to four decimal places for $x = 0.00(0.01)4.00$. This table is reproduced together with a table of $\theta(x)$ at the end of his book (1906a).

Since $R_j(x)$ is a polynomial of degree j the coefficient $E[R_j(x)]$ is a linear combination of the first j moments of x. Bruns introduces the linear transformations

$$u = h(x - c) \text{ and } v = h(y - c)\,, \quad h > 0\,, \quad -\infty < c < \infty\,,$$

and remarks that by setting

$$c = E(x) \text{ and } h^{-2} = 2E[(x - c)^2]\,,$$

the series takes on its "Normalform" in which $E[R_1(u)] = E[R_2(u)] = 0$.

Bruns stresses that the series for $G(x)$ may be used also for a discontinuous distribution if only the cumulative probabilities are referred to the midpoints of the consecutive values of the argument.

Using characteristic functions Bruns (1906a, pp. 134–137) proves that the successive coefficients in the standardized A series for a sum of random variables are of the order of $n^{-(j-2)/2}$, $j \geqq 2$, if the variances of the components are finite and of the same order of magnitude. He remarks that the main terms of this series is due to Laplace and that further terms have been derived by Bessel (1838).

Turning to the fitting of a partial sum of the series to an observed distribution, Bruns estimates the coefficients by means of the corresponding empirical values, corrected for grouping in the continuous case. The number of terms included in the series depends on the goodness of fit. In his book he presents detailed schemes for carrying out the calculations and one worked example. He refers to a paper by his student F. Werner (1900), who has calculated the first six terms of the series for 18 observed distributions, among them some of Fechner's, and compared the observed and calculated frequencies numerically and graphically, an enormous amount of work. In 16 of the 18 cases Werner considers the fit as satisfactory, only for Fechner's two distributions of dimensions of paintings more than six terms are required. He mentions that the method is unsatisfactory for non-homogeneous data.

Bruns's series is the same as that previously found by Gram and Thiele since

$$R_j(x) = (-1)^j H_j^*(x)/2^j j!\,.$$

It is odd that he does not refer to Gram's paper (1883) published in a German mathematical journal.

Bruns's proof is artificial and cumbersome compared with Thiele's which uses only the orthogonality of the series. One naturally asks the question: Why does

Bruns not mention and use the orthogonality? It seems that the explanation is to be found in the first part of his 1897 paper where he discusses a more general series. Briefly told, he introduces a distribution function defined by replacing t^2 in (3.2) by a power series in t^2, assuming that

$$\exp[-\sum_{j=0}^{\infty} a_j t^{2j}]$$

has the same properties as $\exp(-t^2)$ for $|t| \to \infty$. As a result t^2 in (3.3) is replaced by the power series, and the evaluation of the second factor in the integrand leads to a series of the same form as (3.5) but with a more complicated definition of the polynomials $\{R_j(x)\}$ which in the general case are non-orthogonal.

In the second part of the paper he specializes to the normal distribution and proves the usual properties of $R_j(x)$, among them the orthogonality, and concludes that "if a convergent series of the form

$$g(x) = \sum_{j=0}^{\infty} c_j \theta^{(j+1)}(x)$$

exists, then the coefficients can be found in the same way as by the trigonometric series." Bruns's fine proof of the result for the general series is thus superfluous for the special case, but he nevertheless reproduces it in his two later expositions.

The originality of Bruns's general proof rests upon the relations (3.1) and (3.3). His method of proof is influenced by the classical proofs of the central limit theorem, see Hald (1998, p. 319) on Poisson's proof.

Lipps's derivations of the A and B series, 1897, 1901, 1902.

Lipps considered Bruns's proof of the A series as too complicated and presented two simpler proofs (1897, 1901). He (1901) showed that the A and B series may be derived from a common formula. He determined the coefficients by the method of moments without using the orthogonality.

Lipps's first derivation of the A series, 1897.

In his first proof (1897) of the A series he considers the given frequency function $g(x)$ as defined by the equidistant arguments $x_1, \dots, x_m$ with frequencies $g_1, \dots, g_m$, $\Sigma g_i = 1$, that is, a discontinuous or a grouped continuous distribution. He remarks that it is simpler to use a degenerate normal density with infinitely large precision instead of Bruns's sgn x and he therefore introduces the approximation

$$\zeta(x) = \pi^{-\frac{1}{2}} h \sum_{j=1}^{m} g_j \exp[-h^2(x - x_j)^2] \,,$$

and the corresponding probability

$$\int_{x_\alpha}^{x_\beta} \zeta(x)\, dx = \frac{1}{2} \sum_{j=1}^{m} g_j \{\theta[h(x_\beta - x_j)] - \theta[h(x_\alpha - x_j)]\} \,, \; x_\alpha \leqq x_\beta \,.$$

Developing this function around an arbitrary value x_0, say, and introducing the standardized variables $u = h(x - x_0)$ and $u_j = h(x_j - x_0)$ he gets

$$\sum_{k=0}^{\infty}(-1)^k (k!)^{-1}[\theta^{(k)}(u_\beta) - \theta^{(k)}(u_\alpha)] \sum_{j=1}^{m} u_j^k g_j \,.$$

Differentiating with respect to x_α and x_β and letting the arguments tend to x Lipps obtains

$$\zeta(x) = \frac{1}{2}\sum_{k=1}^{\infty}(-1)^k (k!)^{-1}\theta^{(k)}(u) \sum_{j=1}^{m} u_j^k g_j \,,$$

which shows that $\zeta(x)$ is a linear combination of the derivatives of $\theta(x)$ with the moments of the given distribution as coefficients. Using the properties of the coefficients in Bruns's series (3.9), among them the orthogonality, Lipps proves that the two series are identical.

For $h \to \infty$ the moments of $g(x)$ and $\zeta(x)$ are identical. Lipps remarks that the approximation may be satisfactory also for finite values of h, and he evaluates the differences between the exact and approximate moments up to the fifth order to judge the goodness of the approximation.

Lipps's derivation of the B series, 1901.

Lipps's general theory of the A and B series is to be found in the 215-pages long paper *Die Theorie der Collektivgegenstände* (1901) in the section entitled "A method for representing arbitrary given functions" (p. 166). The paper was published as a book the following year.

Lipps represents the given discontinuous frequency function $g(x)$ by the series

$$g(x) = \sum_{j=1}^{n} c_j f(x + \beta_j)\,,\ \beta_1 < \beta_2 < \ldots < \beta_n\,, \quad n = 1, 2, \ldots\,, \tag{3.10}$$

where the β's are suitably chosen integers. He determines the coefficients by the method of moments.

Noting that

$$\nu_{rj} = \sum_{x} x^r f(x + \beta_j) = \sum_{k=0}^{r}(-1)^k \binom{r}{k} \beta_j^k \nu_{r-k}\,,$$

multiplying (3.10) by x^r, and summing over x, he gets

$$\mu_r = \sum_{j=1}^{n} c_j \nu_{rj} = \sum_{k=0}^{r}(-1)^k \binom{r}{k} \nu_{r-k} \sum_{j=1}^{n} \beta_j^k c_j \,.$$

From this system of equations Σc_j, $\Sigma \beta_j c_j, \ldots$ may be found successively as functions of the μ's and ν's, and by solving the resulting linear equations the c's are found.

For a continuous $g(x)$ the β's are real numbers and the sums are replaced by integrals but otherwise the procedure is the same.

Let $\triangle f(x) = f(x+1) - f(x)$ and $\nabla f(x) = f(x) - f(x-1)$ and set

$$f(x+\beta) = \sum_{j=0}^{\beta} \binom{\beta}{j} \triangle^j f(x) .$$

Generalizing (3.10) Lipps writes the general B series in the form

$$g(x) = c_0 f(x) + \sum_{j=1}^{\infty} c_j^* \triangle^j f(x) + \sum_{j=1}^{\infty} c_j \nabla^j f(x) , \quad x = 0 , \pm 1 , \pm 2, \dots .$$

In the following he assumes that $g(x) = 0$ for $x < 0$ and limits the discussion to the series

$$g(x) = \sum_{j=0}^{\infty} c_j \nabla^j f(x) , \quad x = 0, 1, \dots . \tag{3.11}$$

Lipps chooses $\lambda^x/x!$ as $f(x)$ and finds that all the c's contain the factor $\exp(-\lambda)$. We shall therefore use the Poisson frequency function as $f(x)$, which only requires a trivial change of Lipps's formulas and makes comparisons with later developments easier.

For

$$f(x) = e^{-\lambda}\lambda^x/x! , \quad \lambda > 0 ,$$

Lipps finds

$$\nabla^j f(x) = f(x) P_j(x) , \quad j = 0, 1, \dots ,$$

where

$$P_j(x) = \sum_{k=0}^{j} (-1)^k \binom{j}{k} \lambda^{-k} x^{(k)} , \tag{3.12}$$

$x^{(k)} = x(x-1)\cdots(x-k+1)$, $k \geqq 1$, and $x^{(0)} = 1$.

Instead of the ordinary moments he introduces the binomial moments, which we shall denote by α and β, respectively. Multiplying (3.11) by $\binom{x}{r}$ and summing over x, he gets

$$\alpha_r = \sum_{j=0}^{r} \beta_{rj} c_j , \quad r = 0, 1, \dots , \tag{3.13}$$

where

$$\beta_r = \sum_x \binom{x}{r} f(x) = \lambda^r / r! ,$$

and

$$\begin{aligned} \beta_{rj} &= \sum_x \binom{x}{r} \nabla^j f(x) \\ &= (-1)^j \lambda^{r-j}/(r-j)! , \quad j = 0, 1, \dots , r . \end{aligned} \tag{3.14}$$

To prove this formula Lipps uses that

$$\binom{x}{r}\nabla^j f(x) = \binom{x}{r} f(x)P_j(x)$$
$$= \frac{\lambda^r}{r!} f(x-r)P_j(x)\,, \tag{3.15}$$

and

$$P_j(x) = P_j(x-1) - \frac{j}{\lambda} P_{j-1}(x-1)\,, \tag{3.16}$$

which by iteration gives

$$P_j(x) = P_j(x-r) - \binom{r}{1}\frac{j}{\lambda}P_{j-1}(x-r) + \binom{r}{2}\frac{j^{(2)}}{\lambda^2}P_{j-2}(x-r) - \dots\,.$$

This is a finite series that breaks off when the binomial coefficient or the factorial coefficient become zero. It follows that

$$f(x-r)P_j(x) = \sum_{v=0}(-1)^v \binom{r}{v}\frac{j^{(v)}}{\lambda^v}\nabla^{j-v} f(x-r)\,.$$

Using that

$$\sum_x f(x-r) = \sum_x f(x) = 1\,,$$

and

$$\sum_x \nabla^j f(x-r) = \sum_x \nabla^j f(x) = 0\,, \quad j = 1, 2, \dots\,,$$

(3.14) follows. Hence, the matrix of coefficients in (3.13) is lower triangular, and solving for c_r Lipps (p. 505) gets

$$c_r = \sum_{j=0}^{r}(-1)^j[(r-j)!]^{-1}\lambda^{r-j}\alpha_j\,. \tag{3.17}$$

Lipps shows how α_r may be calculated by repeated summations of $g(x)$. He also expresses μ_r in terms of $\alpha_0, \dots, \alpha_r$, but does not go on to express c_r in terms of $\mu_0, \dots, \mu_r$, which is easily done but the resulting formulas are clumsy compared with (3.17).

Lipps presents two empirical distributions with positive frequencies for $x \leqq 3$ and $x \leqq 7$, respectively. Fitting finite series with 5 and 8 terms, respectively, he naturally obtains a good fit. He uses conveniently chosen values of λ and remarks that $\lambda = \alpha_1/\alpha_0$ gives $c_1 = 0$.

It seems that Lipps is the first to develop a B series that is useful for approximating skew discontinuous distributions, it is simpler and more natural for this purpose than Thiele's B series.

We shall now comment on Lipps's proof from the Thiele-Gram point of view. Like Gram's series (2.4), Lipps's series may be written as

$$g(x) = f(x) \sum_{j=0}^{\infty} c_j P_j(x) \,.$$

It is therefore natural to ask whether the P's are orthogonal with respect to the weight function $f(x)$. To prove that this is so we first observe that $\beta_{rj} = 0$ for $j > r$ which follows by repeated summation by parts. From (3.12) and (3.14) we then get for $r \leqq s$

$$\begin{aligned}
\sum_x P_s(x) P_r(x) f(x) &= \sum_x P_s(x) \nabla^r f(x) \\
&= \sum_{k=0}^{s} (-1)^k s^{(k)} \lambda^{-k} \beta_{kr} \\
&= (-1)^r \lambda^{-r} \sum_{k=r}^{s} (-1)^k s^{(k)} / (k-r)! \\
&= \lambda^{-r} s^{(r)} (1-1)^{s-r} \,,
\end{aligned}$$

which proves the orthogonality and shows that

$$\sum P_r^2(x) f(x) = r! \lambda^{-r} \,.$$

It seems that Ch. Jordan (1926) is the first to note and prove the orthogonality. Using the orthogonality it follows that

$$c_r = \lambda^r [r!]^{-1} \sum_x P_r(x) g(x) \,,$$

which by means of (3.12) immediately gives (3.17).

Lipps's second derivation of the A series, 1901.

Lipps (1901, p. 171) derives the A series analogously to the B series. Let $g(x)$ and $f(x)$ be continuous frequency functions and set

$$g(x) = \sum_{i=0}^{n} c_i^* f(x + \beta_i) \,,$$

where the β's are suitably chosen real numbers. Using Taylor's series,

$$f(x+\beta) = \sum_{j=0}^{\infty} \beta^j f^{(j)}(x) / j! \,,$$

Lipps obtains the A series in the form

$$g(x) = \sum_{j=0}^{\infty} c_j f^{(j)}(x) .$$

He mentions that a suitable choice of $f(x)$ is the normal distribution and notes that the corresponding series has been derived by Bruns (1897, 1898) in another way, and he refers to Werner (1900) and his own 1897 paper.

Setting

$$f(x) = h\vartheta(t) , \quad t = h(x-b) ,$$

Lipps obtains

$$g(x) = \sum_{j=0}^{\infty} (h^j c_j) h\vartheta^{(j)}(t) .$$

The method of moments gives

$$\begin{aligned} h^r \mu_r &= \int [h(x-b)]^r g(x)\, dx \\ &= \sum_{j=0}^{r} (h^j c_j)\nu_{rj} , \quad r = 0, 1, \dots , \end{aligned} \tag{3.18}$$

where

$$\begin{aligned} \nu_{rj} &= \int t^r \vartheta^{(j)}(t)\, dt \\ &= (-1)^j r^{(j)} \nu_{r-j} , \quad j = 0, 1, \dots , r , \end{aligned}$$

and

$$\nu_r = \int t^r \vartheta(t)\, dt ,$$

which takes on the values

$$\nu_{2r} = (2r)!/(2^{2r} r!) \;\text{ and }\; \nu_{2r+1} = 0 , \quad r = 0, 1, \dots .$$

Hence,

$$\begin{aligned} h^{2r}\mu_{2r} &= \sum_{j=0}^{2r} (-1)^j (2r)^{(j)} \nu_{2r-j} (h^j c_j) \\ &= (2r)! \sum_{j=0}^{r} (h^{2j} c_{2j})/[2^{2r-2j}(r-j)!] \end{aligned}$$

and

$$h^{2r+1}\mu_{2r+1} = -(2r+1)! \sum_{j=0}^{r} (h_{2j+1} c_{2j+1})/[2^{2r-2j}(r-j)!] .$$

The matrix of coefficients in this system of linear equations is lower triangular. Lipps (p. 510) solves the equations with respect to $h^r c_r$ with the result that

$$c_{2r} = \sum_{j=0}^{r} (-1)^j \mu_{2r-2j} / [h^{2j}(2r-2j)!2^{2j}j!] , \quad r = 0, 1, \ldots , \tag{3.19}$$

and

$$c_{2r+1} = \sum_{j=0}^{r} (-1)^{j+1} \mu_{2r+1-2j} / [h^{2j}(2r+1-2j)!2^{2j}j!] , \quad r = 0, 1, \ldots . \tag{3.20}$$

His solution is simple and his formula directly applicable. He does not introduce the Hermite polynomials, which give the formally more elegant solution (2.9). Lipps does not give an example of the application of the A series, presumably because a wealth of examples had been provided by Werner (1900).

Lipps's criticism of other systems of distributions, 1901.

Lipps (1901, pp. 152–166) critizises the ideas and systems of distributions proposed by Gauss, Hagen, Fechner, and Pearson. He distinguishes between frequency functions based on hypotheses of a probabilistic nature and "empirical" graduation formulas.

The Gaussian distribution is based on the hypothesis of the arithmetic mean, which was generalized by Fechner to a hypothesis about the mode. Hagen derived the normal distribution from the hypothesis of elementary errors and the symmetric binomial. This was generalized by Pearson, who first derived the gamma distribution by means of the skew binomial and afterwards his four-parameter system of distributions from the hypergeometric. Lipps remarks that Pearson's criterion to find out whether a distribution has a finite or an infinite range is illusionary, because the empirical distribution has finite support so supplementary a priori knowledge is required to reach a decision.

He (p. 163) points out the shortcoming of the probabilistic hypotheses, because "every distribution can be produced by an unlimited number of different systems of elementary causes. The often occurring inclination to draw conclusions about the nature of the elementary causes from the form of the empirical distribution or from the estimates of its characteristics has no justification and leads to untenable suppositions without scientific value."

As an example he demonstrates that a certain discrete distribution cannot be generated by independent but by dependent causes. He remarks that elementary causes generally are dependent.

Fechner and Bruns believed that a single formula would be sufficient for describing any regular frequency function, but Lipps showed that a multitude of series expansions exist for this purpose, and he determined the coefficients in the two most important cases. However, he (p. 176) remarks that the coefficients depend not only on the moments of $g(x)$ but also on the choice of the leading term $f_0(x)$ of the series and therefore "an arbitrary and unnatural element is introduced in the characterization of the distribution."

Instead of the series he prefers the symmetric functions, in particular the means, which he defines as $\varepsilon_r = (\mu_r)^{1/r}$, that is,

$$\varepsilon_r^r = \sum_{i=1}^{m} p_i(x_i - b)^r\,, \quad r = 0, 1, \dots\,,$$

and the corresponding empirical means, $\hat{\varepsilon}_r$ says. By algebraic methods he first derives inequalities between the ε's for positive values of $x - b$, and next he discusses the general case, summarizing his results on pp. 499–503.

He (pp. 538–564) analyses five examples from psychology, anthropology, botany and meteorology by calculating the first six means and finding confidence limits for the theoretical values by means of the asymptotic variance

$$\mathrm{var}(\hat{\varepsilon}_r^r) = (\varepsilon_{2r}^{2r} - \varepsilon_r^{2r})/n\,,$$

where n is the number of observations. He (p. 551) says that "These values are the basis for characterizing the properties of the distribution and for comparisons with other distributions of a similar kind."

Summarizing the history so far it will be seen that the contributions of Thiele, Gram, Bruns, and Lipps give a complete solution of the approximation problem by determining the coefficients in the A and B series and by giving the asymptotic distribution of the coefficients in the partial sums.

Hausdorff's derivation of the normal A series, 1901.

F. Hausdorff (1868–1942) vas Lecturer at the Business School in Leipzig when he wrote his paper (1901), in 1910 he became Professor of mathematics at the University of Bonn. His many contributions to probability theory, some of them unpublished, and their importance for later developments have been discussed by Girlich (1996).

There are no new results in Hausdorff's paper, but his method of proof is simpler than previous ones. Independently of Thiele (1899), he introduces the cumulants, which he calls "canonical parameters", by the equation (2.14). He refers to Laplace, Bessel, and Bruns for previous derivations of the A series and uses the classical method of characteristic functions and the inversion formula combined with the definition of the cumulants.

Let $\psi(t) = E[\exp(itx)]$ be the characteristic function of $g(x)$so that

$$\psi(t) = \exp(\sum_{j=1}^{\infty}(it)^j\kappa_j/j!)\,.$$

Using the inversion formula Hausdorff gets

$$g(x) = \frac{1}{2\pi}\int \exp[-ixt + \sum_{j=1}^{\infty}(it)^j\kappa_j/j!]\,dt\,, \tag{3.21}$$

which for $\kappa_1 = 0$ and $\kappa_2 = 1/2$ becomes

$$g(x) = \frac{1}{2\pi} \int \exp[-ixt - t^2/4] \exp[\sum_{j=3}^{\infty} (it)^j \kappa_j / j!] . \tag{3.22}$$

To carry out the integration Hausdorff remarks that

$$\begin{aligned} \int (it)^j \exp[-ixt - t^2/4]\, dt &= (-1)^j D_x^j \int \exp[-ixt - t^2/4]\, dt \\ &= (-1)^j 2\pi \vartheta^{(j)}(x) \\ &= 2\pi \vartheta(x) H_j^*(x) , \quad j = 0, 1, \dots . \end{aligned} \tag{3.23}$$

Espanding the second factor of the integrand in (3.22) in a power series of (it) and using (3.23) he finds the first six terms of the A series in the form (2.13).

Using the orthogonality he gets the series in the form (2.8) and (2.9).

Finally, he assumes that x is the sum of n random variables with finite cumulants so that the rth cumulant of the standardized variable $(x - n\bar{\kappa}_1)(n\bar{\kappa}_2)^{-\frac{1}{2}}$ is of the order of $n^{1-(r/2)}$. Inserting this result in the A series he gets (1.4). Hence, his method of proof is a streamlined version of the classical proofs of the extended central limit theorem.

Bruns's A series for discontinuous distribution functions and his derivation of the B series, 1906b.

Bruns (1906b) characterizes his paper as a supplement and extension of his book (1906a). He discusses four topics:
(1) A slight simplification of his derivation of the A series.
(2) A series expansion for discontinuous distribution functions by means of the A series.
(3) A derivation of a general form of the B series and in particular the series with the binomial and Poisson distributions as leading terms.
(4) A numerical example of the Poisson B series.

Let $g(x_k)$, $k = 1, \dots, m$, be a frequency function, $\Sigma g(x_k) = 1$, $x_1 < x_2 < \dots < x_m$, with the distribution function

$$G(x_i) = g(x_1) + g(x_2) + \dots + g(x_i) .$$

Bruns's series expansion is based on the function $E[\operatorname{sgn}(y - x)]$, see (3.1), which in the present case equals

$$\sum_{k=1}^{m} \operatorname{sgn}(y - x_k) g(x_k) . \tag{3.24}$$

For $y = x_i$ he gets

$$\begin{aligned} \sum_{k=1}^{n} \operatorname{sgn}(x_i - x_k) g(x_k) &= [g(x_1) + \dots + g(x_{i-1})] - [g(x_{i+1}) + \dots + g(x_m)] \\ &= G(x_{i-1}) + G(x_i) - 1 . \end{aligned}$$

The resulting series is thus an expansion of the function

$$G^*(x_i) = \frac{1}{2}[G(x_{i-1}) + G(x_i)]\,, \tag{3.25}$$

corresponding to the midpoints of the vertical parts of the stepfunction $G(x_i)$.

Inserting the A series expansion (3.6) for $\operatorname{sgn}(y-x)$ in (3.24) Bruns finds

$$\begin{aligned} 2G^*(y) - 1 &= E[\operatorname{sgn}(y-x)] \\ &= \sum_{j=0}^{\infty}\sum_{k=1}^{m} R_j(x_k)g(x_k)\theta^{(j)}(y) \\ &= \sum_{j=0}^{\infty} E[R_j(x)]\theta^{(j)}(y)\,, \quad y = x_1, \dots, x_m\,, \end{aligned} \tag{3.26}$$

where

$$E[R_j(x)] = \sum_{k=1}^{m} R_j(x_k)g(x_k)\,, \tag{3.27}$$

which has the same form as the continuous version (3.7).

The corresponding Edgeworth series for a lattice distribution with span h has been discussed by Feller (1966, pp. 512–515), who uses the auxiliary variable $x_i + \varepsilon_i$, where ε_i is uniformly distributed on $(-h/2, h/2)$.

To find the frequency function Bruns notes that

$$\nabla G^*(x_i) = \frac{1}{2}[g(x_i) + g(x_{i-1})]\,,$$

and

$$\nabla^2 G^*(x_i) = \frac{1}{2}[g(x_i) - g(x_{i-2})]\,,$$

so that

$$\nabla^2 G^*(x_1) = \frac{1}{2}g(x_1)\,, \quad \nabla^2 G^*(x_3) = \frac{1}{2}[g(x_3) - g(x_1)]\,, \dots\,,$$

which leads to

$$g(x_{2i+1}) = \sum_{k=0}^{i} \nabla^2[2G^*(x_{2k+1}) - 1]\,, \quad i = 0, 1, \dots, [(m-1)/2]\,.$$

An analogous formula holds for $g(x_{2i})$.

The same result may be obtained by solving (3.25) for $G(x_i)$, which gives

$$\begin{aligned} G(x_i) &= 2\sum_{k=0}^{i-1}(-1)^k G^*(x_{i-k}) \\ &= 2\sum_{k=0}^{[i/2]} \nabla G^*(x_{i-2k}) \end{aligned}$$

and

$$g(x_i) = 2\sum_{k=0}^{[i/2]} \nabla^2 G^*(x_{i-2k}) \,.$$

Bruns refers to the derivations of the B series by Lipps (1902) and Charlier (1905b, 1905c) and remarks that he will derive the series from fundamental principles. His method of proof is analogous to that of the A series with the modification that characteristic functions are replaced by generating functions and $\operatorname{sgn}(y-x)$ is replaced by the function $e(x,y)$ which equals zero for $y \neq x$ and unity for $y = x$. He limits the investigation to expansions of the same form as (3.3), that is,

$$e(x,y) = \sum_{j=0}^{\infty} a_j(x) b_j(y)\,, \quad (x,y) = 0, 1, \dots\,,$$

which means that the discontinuous distribution can be written as

$$\begin{aligned} g(y) &= \sum_{k=0}^{\infty} e(x_k, y) g(x_k) \\ &= \sum_{j=0}^{\infty} c_j b_j(y)\,, \quad c_j = \sum_{k=0}^{\infty} a_j(x_k) g(x_k)\,. \end{aligned}$$

Bruns proves that $a_j(x)$ is a polynomial in x of degree j so that c_j is a linear combination of the first j moments of $g(y)$. Requiring that $b_0(y)$ be a frequency function he gets $b_j(y) = \nabla^j b_0(y)$ so the B series becomes

$$g(y) = \sum_{j=0}^{\infty} E[a_j(x)] \nabla^j b_0(y)\,,$$

which is analogous to (3.8).

Using the Poisson distribution for $b_0(y)$ he obtains the same results as Lipps and Charlier. He applies this series to Bortkewitsch's example of the number of deaths by horsekicks of soldiers in an army corps.

He introduces the standardized variable and proves that the Poisson B series tends to the normal A series for the Poisson parameter tending to infinity.

As another example he sets $b_0(y)$ equal to the binomial (n,p) and proves that

$$(-1)^j a_j(x) = \sum_{k=0}^{\infty} (-1)^k \binom{x}{j-k} \binom{n+k-1}{k} p^k\,. \tag{3.28}$$

However, he does not evaluate $\nabla^j b_0(y)$.

R. von Mises's discussion of the A and B series, 1912, 1931.

In his widely read textbook E. Czuber (1908, vol. 1, pp. 356–372) reproduces Bruns's (1906a) proof of the A series and his example; he adds two examples and reprints Bruns's table of the derivatives of $\vartheta(x)$. He refers to Hausdorff (1901) and Lipps (1902) but does not indicate their methods and results.

R. von Mises (1883–1953) was Associate Professor of applied mathematics at Strasbourg from 1909, Professor at Berlin from 1920 to 1933, and at Harvard from 1939. He (1912) points out that the A series is a special case of the Sturm-Liouville orthogonal expansion of solutions to a second order differential equation, which is satisfied by $\vartheta(x)$ and its derivatives. Using the orthogonality and referring to the Hermite polynomials he gives a simple derivation of the coefficients in Bruns's series. His method is thus the same as Thiele's with the exception that he does not introduce the cumulants.

In his textbook von Mises (1931) gives a complete account of the A and B series, which he calls "Die Brunssche Reihe" (pp. 250–265) and "Die Charliersche Entwicklung" (pp. 265–269), respectively. He does not refer to Hausdorff and Lipps.

The A series for a continuous density may be written as

$$g(x) = \vartheta(x) + \sum_{j=0}^{\infty} c_{j+1}\vartheta^{(j+1)}(x) .$$

By integration we get the A series for a continuous distribution function

$$G(x) = F(x) + \sum_{j=0}^{\infty} c_{j+1}\vartheta^{(j)}(x) , \quad F(x) = \int_{-\infty}^{x} \vartheta(x)\,dx , \tag{3.29}$$

where

$$c_{j+1} = \frac{(-1)^{j+1}}{2^{j+1}(j+1)!}\int H_{j+1}^*(x)g(x)\,dx \tag{3.30}$$

$$= \frac{(-1)^{j}}{2^{j}j!}\int H_{j}^*(x)[G(x)-F(x)]\,dx , \quad j = 0, 1, \dots . \tag{3.31}$$

Formulas (3.29) and (3.31) are von Mises's reformulations of Bruns's results.

Considering the discrete case von Mises assumes that $G(x)$ is a stepfunction with steps of size $g(x_k)$ at $x = x_k$, $k = 1, 2, \dots , m$, $\Sigma g(x_k) = 1$. He evaluates (3.31) by splitting up the integral into its $m+1$ components corresponding to the intervals $(-\infty, x_1)$, $(x_1, x_2), \dots , (x_m, \infty)$. Using the fact that

$$\int_{x_k}^{x_{k+1}} H_j^*(x)\,dx = \frac{1}{2(j+1)}\int_{x_k}^{x_{k+1}} D_x H_{j+1}^*(x)\,dx$$
$$= \frac{1}{2(j+1)}[H_{j+1}^*(x_{k+1}) - H_{j+1}^*(x_k)] ,$$

he gets

$$c_{j+1} = \frac{(-1)^{j+1}}{2^{j+1}(j+1)!} \sum_{k=1}^{m} H^*_{j+1}(x_k) g(x_k) , \tag{3.32}$$

that is, the discrete analogue to (3.30). Von Mises believes that this result is new but it had previously been derived by Bruns (1906b) by another method of proof, see (3.27).

Von Mises uses the series (3.29) with the coefficients (3.32) for representing a given stepfunction $G(x)$, whether this is a discontinuous distribution function or a grouped continuous distribution. This is, however, unsatisfactory because the series takes on the value $G^*(x)$ at any point of increase. It is peculiar that he does not use Bruns's correction for grouping in the continuous case.

4 The British School

F. Y. Edgeworth (1845–1926) was Lecturer in logic at the University of London from 1880, Professor from 1888, and from 1891 Professor of political economy at Oxford University. His contributions to mathematical statistics have been discussed by Bowley (1928), who also gives an annotated bibliography of Edgeworth's 74 statistical papers, see also Stigler (1978, 1986) on modern aspects of Edgeworth's work and his importance for the emergence of the British School of statistics. Here we shall mainly discuss his (1905) derivation of "the generalised law of error," today known as the Edgeworth series.

Edgeworth knew the proofs of the central limit theorem by Laplace and Poisson from Todhunter (1865) and Czuber (1891), to whom he repeatedly refers, but he was ignorant of the works of Bienaymé, Chebyshev, Thiele, Bruns, Lipps, and Hausdorff.

As discussed by Stigler (1986, pp. 338–341) Edgeworth and Pearson competed on developing and first presenting a generalized system of frequency functions. After a preliminary paper on "Poisson's asymmetrical frequency function, " i.e., the first two terms of (1.4), Edgeworth read a paper on the generalized law of error to the Royal Society, which however rejected it for publication. Only an abstract was published in which Edgeworth (1894) presented Poisson's result and added that it could be obtained independently from "a general form for the asymmetrical probability curve." Edgeworth (1895) points out that his 1894 paper is "preserved in the archives of the [Royal] Society," which implies that it is available for other interested statisticians. Pearson thus won the first round of the competition; his four-parameter system of continuous distributions was published in 1895 and successfully applied to many sets of data. However, Edgeworth came back; first, he (1900, pp. 75–77) presented the third term of his series, discussed its importance and applied it to one of Pearson's examples; next, at his instigation Bowley (1902) discussed the three terms of the series, supplied some auxiliary tables and gave many examples of applications to British wage statistics; and finally Edgeworth derived the complete series and gave some applications in three papers (1905, 1906, 1907).

Edgeworth's generalized law of error, 1905.

Edgeworth's 1905 paper is the decisive contribution to the development of the extended central limit theorem begun by the French School. He derives the general term of the A series and rearranges the terms according to magnitude. He gives two proof, one by the method of moments and one by Laplace's method of characteristic functions. Except for the reordering of terms Edgeworth's results are closely related to Thiele's (1899). However, Edgeworth's paper is important for the historical development because Thiele's paper was overlooked. Edgeworth's paper is somewhat difficult to read; we have numbered and reordered his arguments slightly.

(1) Order of magnitude.

Edgeworth considers the sum $s_n = x_1 + \ldots + x_n$ of n independently distributed "elements" (he avoids the term error except for the name of his law) with zero expectations and moments μ_r and μ_{ri}, respectively. Assuming that the n distributions have finite support and that $\mu_2 = \Sigma\mu_{2i}$ is finite it follows that $\mu_{2i} = O(n^{-1})$, $\mu_{ri} = O(n^{-r/2})$ and $\mu_r = O(n^{1-r/2})$, $r = 2, 3, \ldots$.

Edgeworth introduces the cumulants by developing the logarithm of the moment generating function in a power series which shows that

$$\kappa_{2i} = \mu_{2i}\,, \quad \kappa_{3i} = \mu_{3i}\,, \quad \kappa_{4i} = \mu_{4i} - \frac{4!}{2 \times (2!)^2}\,\mu_{2i}^2\,, \ldots\,.$$

He stops at the fourth order but remarks that κ_{ri} "is a homogeneous function" of $\mu_{2i}, \ldots, \mu_{ri}$ wherefore $\kappa_{ri} = O(n^{-r/2})$. The general formula for κ_{ri} in terms of the μ's is given by Thiele, see (2.16), from which Edgeworth's result follows by replacing μ_α by $n^{-\alpha/2}$ etc.

From the independence of the x's it follows that

$$\ln M_{s_n}(t) = \sum \ln M_{x_i}(t)\,,$$

which implies that $\kappa_r = \Sigma\kappa_{ri}$ so that $\kappa_r = O(n^{1-r/2})$. Edgeworth denotes the cumulants by k_r, $r = 0, 1, \ldots$, corresponding to our κ_{r+2}.

(2) The moments in terms of the cumulants and the grouping of terms in order of magnitude.

Developing the right side of the relation

$$1 + \mu_2 t^2/2! + \mu_3 t^3/3! + \ldots = \exp(\kappa_2 t^2/2! + \kappa_3 t^3/3! + \ldots) \tag{4.1}$$

into a power series Edgeworth finds

$$1 + \sum_{r=1}^{\infty} \frac{1}{r!}\left(\frac{\kappa_2}{2!}t^2 + \frac{\kappa_3}{3!}t^3 + \ldots\right)^r$$
$$= 1 + \sum_{r=1}^{\infty} \sum \frac{1}{r_0!r_1!\ldots}\left(\frac{\kappa_2}{2!}\right)^{r_0}\left(\frac{\kappa_3}{3!}\right)^{r_1}\cdots t^{2r_0+3r_1+\ldots}\,, \quad r_0 + r_1 + \ldots = r\,,$$

where the inner summation is over all partitions of r into non-negative integers $r_0, r_1, \dots$. Introducing the notation

$$Q(r^*) = \frac{1}{r_1!}\left(\frac{\kappa_3}{3!}\right)^{r_1} \frac{1}{r_2!}\left(\frac{\kappa_4}{4!}\right)^{r_2} \cdots , \qquad \begin{array}{l} r^* = (r_1, r_2, \dots) \\ r_1 + r_2 + \dots = r - r_0 , \end{array} \tag{4.2}$$

we have

$$1 + \sum_{r=1}^{\infty} \sum \frac{1}{r_0!}\left(\frac{\kappa_2}{2!}\right)^{r_0} t^{2r_0} Q(r^*) t^{3r_1 + 4r_2 + \dots} . \tag{4.3}$$

Identifying the coefficients of t^j in (4.1) and (4.3) Edgeworth gets

$$\frac{\mu_j}{j!} = \sum_{r=1}^{[j/2]} \sum \frac{1}{r_0!}\left(\frac{\kappa_2}{2!}\right)^{r_0} Q(r^*) , \quad j \geqq 2 , \qquad \begin{array}{l} r_0 + r_1 + \dots = r \\ 2r_0 + 3r_1 + \dots = j . \end{array} \tag{4.4}$$

He also writes this result in the same form as Thiele, see (2.15).

However, for Edgeworth's purpose (4.4) is the more practical form because the first term is found by setting r_0 as large as possible whereafter terms of lower order are obtained by gradually diminishing r_0 and increasing $r_1, r_2, \dots$ as demonstrated by Edgeworth in the following two cases:

$$\begin{aligned} \mu_{2j} = {} & \frac{(2j)!}{j!}\left(\frac{\kappa_2}{2!}\right)^{j} + \frac{(2j)!}{(j-2)!}\left(\frac{\kappa_2}{2!}\right)^{j-2} \frac{\kappa_4}{4!} + \frac{(2j)!}{(j-3)!}\left(\frac{\kappa_2}{2!}\right)^{j-3} \frac{1}{2!}\left(\frac{\kappa_3}{3!}\right)^{2} \\ & + \frac{(2j)!}{(j-3)!}\left(\frac{\kappa_2}{2!}\right)^{j-3} \frac{\kappa_6}{6!} + \dots , \quad j = 1, 2, \dots , \end{aligned} \tag{4.5}$$

and

$$\mu_{2j+1} = \frac{(2j+1)!}{(j-1)!}\left(\frac{\kappa_2}{2!}\right)^{j-1} \frac{\kappa_3}{3!} + \frac{(2j+1)!}{(j-2)!}\left(\frac{\kappa_2}{2!}\right)^{j-2} \frac{\kappa_5}{5!} + \dots , \quad j = 1, 2, \dots . \tag{4.6}$$

For μ_{2j} the first term is of order unity, the next two constitute a group of order n^{-1}, the last term is the first in a group of order n^{-2}, and so on. For μ_{2j+1} the first term is of order $n^{-\frac{1}{2}}$, the next is the first in a group of order $n^{-3/2}$, and so on.

(3) Each group of terms in the moments corresponds to a linear combination of the normal density and its derivatives.

Edgeworth notes that the first term of μ_{2j} in (4.5) equals the $(2j)$th moment of the normal distribution with zero mean and variance κ_2. Denoting the normal density by $f(x)$ he proves that the series

$$g(x) = f(x) - \frac{\kappa_3}{3!} D^3 f(x) + \left[\frac{\kappa_4}{4!} D^4 + \frac{1}{2!}\left(\frac{\kappa_3}{3!}\right)^2 D^6\right] f(x) + \dots \tag{4.7}$$

has the same moments as s_n and that the terms are ordered according to magnitude in the same way as in (4.5) and (4.6).

To evaluate $\int x^j g(x)\,dx$ he uses the formula

$$\int x^j(-D)^{j-2s} f(x)\,dx = \frac{j!}{s!}\left(\frac{\kappa_2}{2!}\right)^s, \quad 2s \leqq j, \tag{4.8}$$

which is proved by integration by parts and from which the result follows.

The first term of (4.7) is of order unity, the next of order $n^{-\frac{1}{2}}$, the following two terms constitute the third term of the Edgeworth series and are of order n^{-1}, and so on. The series is thus an asymptotic expansion of the density in question.

(4) A symbolic representation of the series.

The proof given above is unsatisfactory because it stops at terms of order n^{-1}. To give a general proof Edgeworth writes the series in the form

$$g(x) = \exp(\kappa_3(-D)^3/3! + \kappa_4(-D)^4/4! + \dots)f(x). \tag{4.9}$$

Developing the operator in powers of D by the same method as in (4.1) he obtains

$$1 + \sum_{s=1}^{\infty}\sum Q(s^*)(-D)^{3s_1+4s_2+\dots}, \quad \begin{array}{l} s_1 + s_2 + \dots = s \\ s^* = (s_1, s_2, \dots). \end{array} \tag{4.10}$$

The problem is to prove that the moments of $g(x)$ are the same as those given by (4.4). The moment of order $2j$ equals

$$\int x^{2j}(1 + \sum_{s=1}^{\infty}\sum Q(s^*)(-D)^{3s_1+4s_2+\dots})f(x)\,dx.$$

Setting $3s_1 + 4s_2 + \dots = 2j - 2r_0 \geqq 0$ we get by means of (4.8) that

$$\mu_{2j} = \sum_{s=1}^{j}\sum \frac{(2j)!}{r_0!}\left(\frac{\kappa_2}{2!}\right)^{r_0} Q(s^*), \quad \begin{array}{l} s_1 + s_2 + \dots = s \\ 2r_0 + 3s_1 + \dots = 2j, \end{array}$$

which is identical to (4.4). A similar proof holds for μ_{2j+1}.

The symbolic form of the series had been derived by Thiele (1899) by another method of proof, see Hald (2000a).

To find the general expression for the series ordered according to magnitude Edgeworth introduces the generating function

$$\exp(\kappa_3(-D)^3 z/3! + \kappa_4(-D)^4 z^2/4! + \dots), \tag{4.11}$$

remarking that the successive groups of terms in the series are obtained as the coefficients of the successive powers of z in the expansion of this function. The expansion follows immediately from (4.10), which shows that the coefficient of z^m, $m = s_1 + 2s_2 + \dots + ms_m$, equals

$$\sum_{s=1}^{m}(-D)^{m+2s} f(x)\sum Q(s^*), \quad m = 1, 2, \dots, \tag{4.12}$$

since $3s_1+4s_2+\ldots=m+2s$. The order of $Q(s^*)$, $s=1,\ldots,m$, is independent of s and equals $n^{-m/2}$, which follows from the facts that

$$Q(s^*)\propto \kappa_3^{s_1}\kappa_4^{s_2}\ldots,$$

and $\kappa_r=O(n^{1-r/2})$, $r=2,3,\ldots$, so that the exponent of n in the order of $Q(s^*)$ becomes

$$(1-\frac{3}{2})s_1+(1-\frac{4}{2})s_2+\ldots=s_1+s_2+\ldots-\frac{1}{2}(3s_1+4s_2+\ldots)=-\frac{1}{2}m.$$

Hence, (4.12) is the compact expression for the $(m+1)$st group of terms, all of order $n^{-m/2}$, of the series. It is easy to check that the first three groups are given by (4.7); the next group is

$$-\left[\frac{\kappa_5}{5!}D^5+\frac{\kappa_3}{3!}\frac{\kappa_4}{4!}D^7+\frac{1}{3!}\left(\frac{\kappa_3}{3!}\right)^3D^9\right]f(x),$$

as given by Edgeworth (1905, p. 61).

(5) Edgeworth's completion of Laplace's proof.

Edgeworth remarks that the same series may be found by an extension of the method of Laplace and Poisson. We shall sketch Edgeworth's proof using modern notation and setting $u=s_n/\sqrt{\kappa_2}$.

By the same procedure as used by Bienaymé and Hausdorff, see (3.21)–(3.23), Edgeworth obtains

$$p(s_n)=\frac{1}{2\pi}\int\exp(-iut-t^2/2)\exp(\sum_{j=3}^{\infty}(it)^j\kappa_j/j!)\,dt.$$

He remarks that an expansion of the second factor of the integrand may be found by replacing $(-D)$ by (it) in the expansion (4.10) of the operator. Hence,

$$p(s_n)=\frac{1}{2\pi}\int\exp(-iut-t^2/2)(1+\sum_{s=1}^{\infty}\sum Q(s^*)(it)^{3s_1+4s_2+\ldots})\,dt.$$

Setting

$$3s_1+4s_2+\ldots=j\,,\quad j\geqq 3\,,$$

and using (3.23) with $\phi(x)$ instead of $\vartheta(x)$, it follows that

$$\sqrt{\kappa_2}\,p(s_n)=\phi(u)+\sum_{j=3}^{\infty}(-1)^j\phi^{(j)}(u)\sum_{s=1}^{[j/3]}\sum Q(s^*)\,,\tag{4.13}$$

which is the completion of the extended central limit theorem in the form of an A series. Edgeworth also writes the last factor in the form (2.19).

To bring (4.13) in a form analogous to (1.4) we introduce

$$\kappa_{r+2}/\kappa_2^{(r+2)/2} = \bar{\kappa}_{r+2}/[\bar{\kappa}_2^{(r+2)/2} n^{r/2}] = \gamma_r/n^{r/2}\,, \quad r = 1, 2, \dots\,,$$

and

$$Q_\gamma(s^*) = \frac{1}{s_1!}\left(\frac{\gamma_1}{3!}\right)^{s_1} \frac{1}{s_2!}\left(\frac{\gamma_2}{4!}\right)^{s_2} \cdots\,,$$

so that

$$Q(s^*) = Q_\gamma(s^*)/n^{j/2-s}\,, \quad s = 1, \dots, [j/3]\,.$$

Hence, (4.13) becomes

$$\sqrt{n\bar{\kappa}_2}\, p(s_n) = \phi(u) + \sum_{j=3}^{\infty} (-1)^j \phi^{(j)}(u) n^{1-j/2} \sum_{s=1}^{[j/3]} n^{s-1} \sum Q_\gamma(s^*)\,. \tag{4.14}$$

For $j \geqq 6$ the last sum consists of at least two terms of various orders, which explains why the series is not ordered according to increasing powers of $n^{-1/2}$.

Transcribing (4.12) in the same manner we get Edgeworth's version as

$$\sqrt{n\bar{\kappa}_2}\, p(s_n) = \phi(u) + \sum_{m=1}^{\infty} (-1)^m n^{-m/2} \sum_{s=1}^{m} \phi^{(m+2s)}(u) \sum Q_\gamma(s^*)\,. \tag{4.15}$$

Edgeworth does not summarize his results by presenting the two general expressions (4.14) and (4.15) but the formulas are clearly implied by his presentation of the first terms of the series.

It is clear that Edgeworth's result holds whether the support of the distributions is finite or infinite if only the moments of the components of s_n are of the order indicated in (1). Likewise, the law holds for a linear combination of components instead of a sum. Generalizations to non-linear functions and to correlated components are discussed by Edgeworth (1906), who concludes that the law holds for these cases if the standardized cumulants are small and decreasing, although not necessarily as $n^{-1/2}$, $n^{-1}, \dots$. However, he does not reach a general result for these cases.

Applications and discussions of the Edgeworth series compared with other systems.

In the 1906 paper the first three terms of the series are fitted to three sets of data, not very abnormal, the analysis being carried out by Bowley. The goodness of fit is comparable to that obtained by fitting a four-parameter Pearson curve. Edgeworth also presents some diagrams showing the effects of varying γ_1 and γ_2.

After having discovered the works of Thiele, Bruns and Charlier, Edgeworth (1907) points out that his series represents the distribution of aggregated random variables, whereas the other authors aim at approximating a given frequency function by a suitably chosen series. Moreover, the "Bruns-Charlier" series differs from his series in the third and following terms.

From 1895 and for the rest of his life Edgeworth carried on a polemical discussion on the advantages and disadvantages of his system of distributions (the generalized law and the method of translation) compared with Pearson's system, for a summary see Edgeworth (1917, pp. 411–437). He repeatedly underlined that the generalized law of error is based on the physical hypothesis that observed quantities in many fields depend nearly linearly on the effects of many independent causes and therefore has the status of a law of nature, whereas other systems of distributions are merely empirical. However, he added that the generalized law describes only moderately abnormal distributions, so that the more flexible empirical systems are important practical supplements.

In the applications of the series as a frequency function Edgeworth uses at most three terms. He notes that this may lead to negative frequencies in the tails of the distribution. He (1924) summarizes the results of a lifelong work with fitting this formula to 19 empirical distributions by listing the values of the skewness $m_3 m_2^{-3/2}$, which varies from 0.063 to 0.29, and the kurtosis $m_4 m_2^{-2} - 3$, varying from -0.051 to 0.327, for slightly and moderately abnormal cases. If the two coefficients are calculated from a sample of N observations their standard errors are of order $N^{-1/2}$ and N^{-1}, respectively, whereas the coefficients themselves are of order $n^{-1/2}$ and n^{-1}, where n is the unknown number of components.

K. Pearson (1857–1936), from 1884 Professor of applied mathematics and mechanics at University College, London, based his model (1895) on the hypergeometric distribution, that is, Pearson's s_n is the sum of n negatively correlated binary variables. Edgeworth points out that this model is too special as the basis for a general theory of distributions for sums of interdependent variables. It seems that Pearson reached the same conclusion because he (1905b) writes that "all discussion of asymmetrical frequency must turn in one form or another on the proper form to be given to $F(x)$ in the equation

$$\frac{1}{y}\frac{dy}{dx} = \frac{-x}{\sigma_0^2 F(x)} \text{."}$$

Setting $F(x) = a_0 + a_1 x + a_2 x^2 + \ldots$ and disregarding terms of higher order Pearson obtains his system of frequency curves, $y = y(x)$, without reference to urn models and underlying causes. In the same paper Pearson criticizes Edgeworth's method of translation.

Yule (1906) used the opportunity at the discussion of Edgeworth's (1906) paper to defend Pearson's original idea. He argues that Edgeworth has not reached the bottom of the problem because the distributions of the components are unspecified, and he asks the question: How did these distributions arise? He answers that "he would regard every distribution as being built up from a series of such elements, each capable of taking only one of two values. That was the general process, of which the special processes adopted by Quetelet, Pearson, and others were particular instances." Moreover, Edgeworth's "process of analysis into elements was purely mental; you could classify the elements out of which any variable was built up in an indefinite number of ways." Edgeworth answered that "The fact that sporadic variation was widely prevalent might legitimately be used as the foundation of a theory, though the ultimate causes underlying that fact were imperfectly apprehended."

Pearson read widely and in some cases superficially to judge from the following two examples. In a Note at the end of his 1895 paper Pearson remarks that he has procured and read Thiele's (1889) Danish textbook. He notes correctly that Thiele has introduced the cumulants and measures of skewness and kurtosis similar to his own but he overlooks Thiele's discussion of the A series. He remarks that Thiele does not use the cumulants for estimating the parameters in frequency functions although Thiele calculates the first six cumulants for his example from which the coefficients of the A series may be calculated, see Table 1. Pearson's mistake is presumably due to the fact that Thiele did not calculate the corresponding frequencies. On the other hand Pearson correctly relates Thiele's discussion and application of the B series and shows that a somewhat better fit than the one recorded in our Table 1 may be obtained by his four-parameter Type I distribution. Pearson (1905a) criticizes Thiele and Lipps. He suggests that their purpose is "to reproduce the complete frequency" by including as many terms of the series as there are classes minus one. He warns against this procedure, partly because the problem is graduation, not interpolation and partly because of the large sampling errors of the higher moments. However, this is exactly the arguments of Thiele and Lipps; in their applications they do not use moments of higher order than six. Pearson does not in this context mention the applications of the four-parameter Edgeworth series due to Edgeworth and Bowley.

Among British statisticians only Bowley used the Edgeworth series as a frequency distribution; in his textbook (1926) he gave a simplified version of Edgeworth's proofs and used the first two terms of the series as a skew distribution.

A compact summary of the results of Thiele and Edgeworth, without reference to these authors, is given by Cornish and Fisher (1937) together with two new applications of the series. They derive unbiased estimates of the cumulants and give an expansion of the percentiles of $g(x)$ in terms of the standardized normal percentiles by inverting the relation between the distribution function of x and the corresponding Edgeworth series. A supplementary paper by Fisher and Cornish (1960) gives further formulas and tables.

Barton and Dennis (1952) determine the region in the (λ_1, λ_2)-plane within which the four-parameter series is unimodal and positive, see also the improved version by Draper and Tierney (1972).

5 The Swedish School

It was rather late in his career that C. V. L. Charlier (1862–1934), Professor of astronomy at the University of Lund, became interested in mathematical statistics. His main ideas and results on the A and B series are contained in five papers published between 1905 and 1908. It seems that he was ignorant of most of the literature on these topics when he wrote the first two papers. In the third paper (1905c) "On the representation of arbitrary functions" he refers to Gram and Thiele and remarks that he in November 1904 visited Bruns in Leipzig and that he considers his own paper as a generalization of Bruns's work. It is, however, better characterized as a modification of Lipps's paper (1901), which he surprisingly did not know.

We shall sketch the contents of the five papers and also make some remarks on the special cases and examples discussed in the following papers. In a survey paper Charlier (1914) lists 19 papers as his "Contributions to the mathematical theory of statistics" so far; many of these papers contain improved versions of the original proofs and are mostly didactic.

(1) **Charlier's probabilistic derivation of the normal A series, 1905a.**

Referring to Hagen's hypothesis of elementary errors Charlier derives (1.4) by means of the characteristic function for a sum and the inversion formula. His proof is essentially the same as Poisson's (1824), which he presumably knew from Todhunter (1865). His proof contains two errors that neutralize each other; he points out the error in his 1914 paper. There is nothing new in this paper, both the method of proof and the results were well known.

(2) **Charlier's probabilistic derivation of the B series, 1905b, 1908.**

Let x_j , $j = 1, 2, \dots$, be independent random variables taking on the values 0 and 1 with probabilities q_j and p_j, respectively, $p_j + q_j = 1$. Laplace (1812, II, §38) shows that the characteristic function for $s = x_1 + \dots + x_n$ equals

$$\psi(t) = \prod_{j=1}^{n} (q_j + p_j e^{it}) \,, \tag{5.1}$$

and by means of the inversion theorem he proves that s is asymptotically normal with mean Σp_j and variance $\Sigma p_j q_j$. The same result is obtained by Poisson (1837, §109).

It is well known that Poisson (1837, §81) derived the distribution

$$f_\lambda(x) = e^{-\lambda}\lambda^x/x! \,, \quad x = 0, 1, \dots \,,$$

from the binomial for $np = \lambda, 0 < \lambda < \infty$, and $n \to \infty$. The characteristic function for the Poisson distribution is

$$e^{-\lambda} \sum_{x=0}^{\infty} \lambda^x e^{ixt}/x! = \exp(-\lambda + \lambda e^{it}) \,.$$

Charlier considers the limiting distribution for the model with varying probabilities, often called Poisson trials.

Setting $E(s) = \Sigma p_j = \lambda, 0 < \lambda < \infty$, so that $p_j = O(n^{-1})$ and using (5.1) Charlier gets

$$\begin{aligned} \ln \psi(t) &= \sum [\ln q_j + \ln(1 + (p_j/q_j)e^{it})] \\ &= \sum [-p_j + (p_j/q_j)e^{it} + \dots] \\ &= -\lambda + \lambda e^{it} + \dots \,. \end{aligned}$$

Hence,Charlier finds that s is asymptotically Poisson distributed with parameter λ. By means of the inversion formula he writes the limit distribution in the form

$$\begin{aligned} p_\lambda(s) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp(-\lambda + \lambda e^{it} - ist)\, dt \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp[-\lambda + \lambda\cos t + i(\lambda\sin t - st)]\, dt \\ &= \frac{1}{\pi}e^{-\lambda}\int_{0}^{\pi} e^{\lambda\cos t}\cos(\lambda\sin t - st)\, dt\,, \quad s = 0, 1, \dots\,. \end{aligned} \tag{5.2}$$

In his enthusiasm for this result Charlier writes: "This is the function which in the present case plays a similar rôle as the Gaussian distribution in the usual theory of errors."

Next, he extends the definition of $p_\lambda(s)$ from non-negative integers to "arbitrary real or imaginary values of the argument." To study the properties of this function he notes that

$$p_0(s) = \frac{1}{\pi}\int_0^{\pi} \cos(st)\, dt = \frac{\sin\pi s}{\pi s}\,,$$

and developing $p_\lambda(s)$ in Maclaurin's series he obtains

$$\begin{aligned} p_\lambda(s) &= \sum_{j=0}^{\infty}(-1)^j\frac{1}{j!}\lambda^j\nabla^j p_0(s) \\ &= e^{-\lambda}\sum_{j=0}^{\infty}\frac{1}{j!}\lambda^j p_0(s-j)\,. \end{aligned}$$

Using that

$$p_0(s-j) = (-1)^j\frac{\sin\pi s}{\pi(s-j)}\,,$$

he gets the expansion

$$p_\lambda(s) = e^{-\lambda}\frac{\sin\pi s}{\pi}\sum_{j=0}^{\infty}(-1)^j\frac{1}{j!(s-j)}\lambda^j\,, \quad -\infty < s < \infty\,. \tag{5.3}$$

It follows that

$$p_\lambda(s) = \begin{matrix} 0 & \text{for } s = -1, -2, \dots \\ e^{-\lambda}\lambda^s/s! & \text{for } s = 0, 1, \dots\,. \end{matrix} \tag{5.4}$$

The function $p_\lambda(s)$ may thus be called a continuous version of the Poisson distribution.

Charlier does not discuss the properties of $p_\lambda(s)$ further. However, the function may be considered as an interpolation formula based on the values given by (5.4), which shows that $p_\lambda(s)$ for $s<0$ alternates between positive and negative values in successive intervals of unit length. Hence, $p_\lambda(s)$ is no frequency function, the connection with the original model is lost and the function has no probabilistic interpretation.

Charlier was not the first to invent a continuous version of a discrete distribution. Thiele (1903, p. 21) proposes to write the binomial coefficient in the form

$$\binom{n}{s} = \frac{\Gamma(n+1)}{\Gamma(s+1)\Gamma(n-s+1)}$$
$$= \frac{\sin \pi s}{\pi s}\frac{\Gamma(n+1)}{(n-s)\dots(1-s)}\,, \qquad -\infty < s < \infty\,,\ s \neq 1, 2, \dots, n\,.$$

He remarks, however, that the corresponding continuous version of the binomial is inadmissable as a frequency function because it alternates between negative and positive values periodically for $s<0$ and $s>n$.

Finally, Charlier remarks that the complete expansion of $\ln\psi(t)$ is

$$\ln\psi(t) = \sum_{j=1}^{n}[\ln q_j + (p_j/q_j)e^{it}] + \sum_{m=2}^{\infty}\frac{1}{m}e^{itm}\sum_{j=1}^{n}(p_j/q_j)^m\,,$$

which for $n\to\infty$ shows that

$$\ln\psi(t) = -\lambda + \lambda e^{it} + \text{ a linear combination of } e^{itm}\,, \quad m = 0, 1, \dots\,,$$

the first two coefficients being of order n^{-1} and the following of order n^{1-m}. Since

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\exp(-\lambda + \lambda e^{it} - ist)e^{itm}dt = p_\lambda(s-m)\,,$$

the complete expression for $p(s)$ has the form

$$p(s) = \sum_{m=0}^{\infty}\alpha_m p_\lambda(s-m)$$
$$= \sum_{m=0}^{\infty}\beta_m \nabla^m p_\lambda(s)\,, \tag{5.5}$$

which is Charlier's first derivation of the B series. He does not discuss how the coefficients depend on λ.

In 1908 Charlier realized that his enthusiasm for the function $p_\lambda(s)$ as a means for describing distributions with positive frequencies for $s<0$ was unfounded. He explains that this is due to the fact that the elementary errors are assumed to be non-negative and he therefore generalizes his previous model by considering a trinomial error distribution instead of the binomial. Let x_j take on the values

$-1, 0, 1$ with probabilities p_{1j}, q_j, p_{2j}, respectively, $p_{1j} + q_j + p_{2j} = 1$, assuming that p_{1j} and p_{2j} are of the order n^{-1} and $p_{2j} \geqq p_{1j}$, say. Charlier states that the same method of proof as above leads to the "auxiliary function"

$$p_{\lambda,\mu}(s) = \frac{1}{\pi} e^{-\lambda} \int_0^\pi e^{\lambda \cos t} \cos(\mu \sin t - st)\, dt\,, \quad -\infty < s < \infty\,. \tag{5.6}$$

He remarks that $p_{\lambda,\mu}(s) > 0$ for $s = -1, -2, \ldots$, and indicates that he on a later occasion will return to a more detailed study of this function. However, he never did so, it seems that the mathematical and numerical problems were too difficult. We shall return to this matter under (6).

To see how the parameters depend on the error distribution we shall derive (5.6) under the assumption that the error distribution is the same for all the components; the proof in the general case is the same.

Setting $np_2 - np_1 = \mu$ and $np_2 + np_1 = \lambda$ we get $E(s) = \mu$ and

$$\mathrm{var}(s) = np_2 + np_1 - n(p_2 - p_1)^2 = \lambda - \mu^2/n\,.$$

It is easy to prove that $\kappa_{2r-1}(s) \to \mu$ and $\kappa_{2r}(s) \to \lambda$, $r = 1, 2, \ldots$, for $n \to \infty$.

From the characteristic function

$$\psi(t) = (p_1 e^{-it} + q + p_2 e^{it})^n$$

we get

$$\begin{aligned}
\ln \psi(t) &= n \ln q + n \ln \left[1 + (p_1/q)e^{-it} + (p_2/q)e^{it}\right] \\
&= -\lambda + np_1 e^{-it} + np_2 e^{it} + \ldots \\
&= -\lambda + \frac{1}{2}\lambda(e^{it} + e^{-it}) + \frac{1}{2}\mu(e^{it} - e^{-it}) + \ldots \\
&= -\lambda + \lambda \cos t + i\mu \sin t + \ldots\,,
\end{aligned}$$

from which (5.6) immediately follows by means of the inversion formula. Hence, μ is the mean and λ the variance of s, $s = 0, \pm 1, \pm 2, \ldots$, for $n \to \infty$.

For $\mu = \lambda$ we have $p_{\lambda,\mu}(s) = p_\lambda(s)$, which for $s = 0, 1, \ldots$ equals the Poisson distribution. Like Jørgensen (1916), we shall therefore call $p_{\lambda,\mu}(s)$ the Poisson-Charlier distribution.

(3) **Charlier's representation of an arbitrary function by the A series, 1905c, 1906.**

Charlier (1905c) writes the A series as $g(x) = \Sigma c_j f^{(j)}(x)$ and determines the coefficients by the method of moments using (1.12) with

$$\nu_{rj} = \frac{1}{r!} \int_{-\infty}^{\infty} x^r f^{(j)}(x)\, dx\,, \qquad \begin{array}{l} j = 0, 1, \ldots \\ r = 0, 1, \ldots\,. \end{array}$$

Integration by parts gives the recursion formula $\nu_{rj} = -\nu_{r-1,j-1}$, which leads to

$$\nu_{rj} = (-1)^j \nu_{r-j,0} = (-1)^j \nu_{r-j}\,,$$

where

$$\nu_r = \frac{1}{r!}\int_{-\infty}^{\infty} x^r f(x)\,dx\,, \quad r = 0, 1, \dots .$$

Moreover, $\nu_{rj} = 0$ for $j > r$.

Hence, (1.12) leads to

$$\mu_r = \sum_{j=0}^{r} (-1)^j \nu_{r-j} c_j\,, \quad r = 0, 1, \dots ,$$

which shows that the c's are determined successively as linear combinations of the μ's with coefficients that are independent of $g(x)$. Charlier therefore writes

$$c_r = \int_{-\infty}^{\infty} S_r(x) g(x)\,dx\,, \tag{5.7}$$

where

$$S_r(x) = s_{r0} + s_{r1}x/1! + s_{r2}x^2/2! + \dots + s_{rr}x^r/r!\,.$$

Inserting the series for $g(x)$ into (5.7) Charlier obtains

$$c_r = \sum_{j=0}^{\infty} c_j \int_{-\infty}^{\infty} S_r(x) f^{(j)}(x)\,dx\,,$$

so that

$$\int_{-\infty}^{\infty} S_r(x) f^{(j)}(x)\,dx = \begin{matrix} 0 \text{ for } j < r \\ 1 \text{ for } j = r\,, \end{matrix} \tag{5.8}$$

which leads to the equations

$$(-1)^j \sum_{k=0}^{r} \nu_{k-j} s_{rk} = \begin{matrix} 0 \text{ for } j = 0, 1, \dots , r-1 \\ 1 \text{ for } j = r\,, \end{matrix}$$

for the determination of $\{s_{rk}\}$. These equations may be written in matrix form as

$$\begin{pmatrix} \nu_0 & 0 & \cdots & 0 \\ \nu_1 & \nu_0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \nu_r & \nu_{r-1} & \cdots & \nu_0 \end{pmatrix} \times \begin{pmatrix} s_{rr} \\ s_{r,r-1} \\ \vdots \\ s_{r0} \end{pmatrix} = \begin{pmatrix} (-1)^r \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{5.9}$$

Charlier uses determinantal expressions instead of the matrix form above. He works out the explicit expressions for $S_r(x)$ for $r = 1, \dots , 4$ and gives the recursion formula

$$s_{rk} = -s_{r-1,k-1}\,, \quad k = 1, \dots , r\,. \tag{5.10}$$

The problem of determining the coefficients in Charlier's A series is thus formally solved by means of (5.9) and (5.7). If more than four terms are needed (5.10) may be used for calculating further values of $S_r(x)$.

As a special case Charlier sets $f(x)$ equal to the normal density and derives the first six coefficients. He observes that $S_r(x) \propto H_r(x)$ and that his results previously have been found by Gram, Thiele, and Bruns. He overlooks that Lipps (1901) had formulated the same general model and given the complete solution for the normal case, see (3.19)–(3.20).

As a second special case Charlier sets

$$f(x) = \frac{2}{\pi\alpha}\left(e^{x/\alpha} + e^{-x/\alpha}\right),$$

and determines the first four coefficients of the expansion.

In a following paper he (1906) presents the formulas for the normal A series and gives detailed schemes for the calculation of the parameters. He illustrates the theory by five examples.

(4) **Charlier's representation of an arbitrary function by the B series, 1905c, 1906.**

Charlier's procedure is analogous to that for the A series. He writes the B series as $g(x) = \Sigma c_j \nabla^j f(x)$ and determines the coefficients by the method of moments using (1.12) with

$$\nu_{rj} = \sum_{x=-\infty}^{\infty} x^r \nabla^j f(x), \quad \begin{array}{l} j = 0, 1, \dots \\ r = 0, 1, \dots . \end{array}$$

To express ν_{rj} in terms of $\nu_r = \Sigma x^r f(x)$ Charlier uses that

$$\nabla(u_x v_x) = u_x \nabla v_x + v_{x-1} \nabla u_x,$$

which sums to zero under the assumption that $u_x v_x \to 0$ for $|x| \to \infty$. Hence,

$$\sum u_x \nabla v_x = -\sum v_{x-1} \nabla u_x,$$

so that

$$\sum x^r \nabla^j f(x) = -\sum \nabla^{j-1} f(x-1) \nabla x^r.$$

Repeated applications of this formula give

$$\begin{aligned} \nu_{rj} &= (-1)^j \sum f(x-j) \nabla^j x^r \\ &= (-1)^j \sum f(x) \nabla^j (x+j)^r, \quad j = 0, 1, \dots, r, \end{aligned}$$

and $\nu_{rj} = 0$ for $j > r$, which shows that ν_{rj} is a linear combination of $\nu_0, \nu_1, \dots, \nu_r$, where $\nu_r = \Sigma x^r f(x)$.

It thus follows from (1.12) that the c's are linear combinations of the μ's with coefficients that are independent of $g(x)$. Charlier writes

$$c_r = \sum_{x=-\infty}^{\infty} T_r(x) g(x) \,, \tag{5.11}$$

where

$$T_r(x) = t_{r0} + t_{r1}x + \ldots + t_{rr}x^r \,.$$

Inserting the series for $g(x)$ in (5.11) he finds that

$$\sum T_r(x) \nabla^j f(x) = \begin{matrix} 0 \ \text{ for } j < r \,, \\ 1 \ \text{ for } j = r \,, \end{matrix} \tag{5.12}$$

which leads to the equations

$$\sum_{k=0}^{r} \nu_{r-k,j} t_{r,r-k} = \begin{matrix} 0 \ \text{ for } j = 0, 1, \ldots, r-1 \,, \\ 1 \ \text{ for } j = r \,. \end{matrix} \tag{5.13}$$

The matrix of coefficients is lower triangular so that the t's are linear combinations of the ν_{rj}, $j = 0, 1, \ldots, r$, which may be expressed as linear combinations of ν_j, $j = 0, 1, \ldots, r$. Charlier derives the four polynomials $T_1(x), \ldots, T_4(x)$ and using (5.11) he finds the first four coefficients in terms of the moments of $g(x)$ and $f(x)$.

He remarks that if $f(x)$ contains m disposable parameters then we may use them to make m coefficients disappear. He uses this idea in his first example which is a discussion of Pearson's (1895) approximation by means of the binomial to an asymmetrical frequency function. The B series is

$$g(x) = \sum c_j \nabla^j f\left(\frac{x}{a}\right), \quad a > 0 \,,$$

where

$$f(x) = \frac{n!}{(b+x)!(n-b-x)!} p^{b+x} q^{n-b-x}, \quad 0 < b + x < n \,,$$

and $f(x) = 0$ otherwise. Suppose that a, b, p, n have been chosen such that $c_1 = \ldots = c_4 = 0$. We then have

$$\begin{aligned} \mu_r &= c_0 \sum x^r f\left(\frac{x}{a}\right) + c_5 \sum x^r \nabla^5 f\left(\frac{x}{a}\right) + \ldots \\ &= a^r (c_0 \nu_r + c_5 \nu_{r5} + \ldots) \,, \quad r = 0, 1, \ldots \,, \end{aligned}$$

where

$$\nu_r = \sum x^r f(x) = \sum (x+b)^r f(x+b) \,,$$

which are the well-known moments of the binomial. Charlier thus finds that the four parameters a, b, p, n must satisfy the four equations $\mu_r = a^r \nu_r$, $r = 1, \ldots, 4$, in agreement with Pearson. He remarks that the following terms of the B series give a "correction" to Pearson's result but he does not determine these terms.

He (1905c, 1906) then turns to his main result: The Poisson B series. After having found ν_{rj} as function of λ he determines the first four coefficients from the general formula with the result that $c_1 = \lambda - \mu_1$,

$$\begin{aligned} 2!c_2 &= \lambda^2 - (2\lambda + 1)\mu_1 + \mu_2\,, \\ 3!c_3 &= \lambda^3 - (3\lambda^2 + 3\lambda + 2)\mu_1 + 3(\lambda + 1)\mu_2 - \mu_3\,, \\ 4!c_4 &= \lambda^4 - (4\lambda^3 + 6\lambda^2 + 8\lambda + 6)\mu_1 + (6\lambda^2 + 12\lambda + 11)\mu_2 - (4\lambda + 6)\mu_3 + \mu_4\,, \end{aligned} \tag{5.14}$$

where

$$\mu_r = \sum x^r g(x)\,, \quad r = 1, 2, \ldots\,.$$

He derives the simpler formulas obtained by setting $\lambda = \mu_1$ and introducing the central moments.

Finally, he introduces two more parameters, a and b say, by replacing $g(x)$ by $g(ax + b)$. He derives the first four terms of the series for the following special cases: (1) Setting $a = 1$, he chooses b and λ such that $c_1 = c_2 = 0$. (2) He chooses a, b, λ such that $c_1 = c_2 = c_3 = 0$. (3) He chooses a and λ such that $c_1 = c_2 = 0$. He fits a B series to three sets of data.

To facilitate the use of the B series he (1906) reproduces Bortkewitsch's (1898) table of the Poisson frequency function.

Regarding the Poisson B series Charlier was preceded by Lipps (1901) who gave the elegant formula (3.17) for the coefficients. Charlier's achievement lies in the derivation of the polynomials $T_r(x)$ for the general B series and the orthogonality relation (5.12).

(5) Charlier's applications of the series to the Bernoulli, Poisson, and Lexis models, 1909, 1911.

Charlier (1909) derives the first six cumulants of the binomial distribution and the usual A series approximation may thus be found, see (2.13). However, Charlier wanted to improve this result and developed a slightly different series.

Setting

$$b(x, y) = \binom{n}{x} p^x q^y\,, \quad x + y = n\,, \quad x = 0, 1, \ldots, n\,,$$

and introducing the auxiliary function

$$\psi(t) = pe^{it} + qe^{-it}\,,$$

he observes that

$$\psi^n(t) = \sum_{x=0}^{n} b(x, y) e^{it(x-y)}\,,$$

so that

$$b(x,y) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp\{n\kappa(t) - it(x-y)\}\,dt\,, \tag{5.15}$$

where $\kappa(t) = \ln\psi(t)$, $\kappa_1 = p - q$, $\kappa_2 = 4pq$.

Setting $v = x - np$ and $\sigma^2 = npq$, he finds

$$n\kappa(t) - it(x-y) = -2\sigma^2 t^2 - 2vit + n\sum_{j=3}^{\infty} \kappa_j (it)^j/j!\,.$$

The main term of the expansion becomes

$$\begin{aligned} f(v) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp(-\frac{1}{2}\sigma^2 t^2 - ivt)\,dt \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp(-\frac{1}{2}\sigma^2 t^2)\cos(vt)\,dt\,, \end{aligned} \tag{5.16}$$

which for $\sigma^2 \to \infty$ tends to normality. The improved series is

$$b(x,y) = f(v) + \sum_{j=3}^{\infty} c_j f^{(j)}(v)\,, \quad v = x - np\,, \tag{5.17}$$

where the coefficients are the same as for the normal A series. He calls this the "strict" form and recommends it when σ^2 is small. It seems that nobody has used it.

To obtain the B series he sets $np = \lambda$ and finds that

$$n\kappa(t) - it(x-y) = -\lambda(1 - e^{2it}) - 2itx - n\sum_{j=2}^{\infty} \frac{1}{j} p^j (1 - e^{2it})^j\,. \tag{5.18}$$

Using (5.15) the main term becomes $p_\lambda(x)$ and the series becomes the ordinary Poisson B series (5.5), but in the present case he is able to find the coefficients because of the simple form of (5.18). He does not use (5.14), presumably because c_5 and c_6 are rather complicated.

Charlier finds the first six coefficients in the usual way by identifying the coefficients of t in two power series. We shall show how the general formula may be derived from Lipps's formula (3.17). Noting that

$$\begin{aligned} \alpha_j &= n^{(j)} p^j/j! \\ &= (p^j/j!)\sum_{k=0}^{j} (-1)^{j-k} n^k D_{kj}\,, \quad D_{kj} = D^k 0^{(-j)}/k!\,, \quad D_{0j} = 0\,, \end{aligned}$$

we get

$$\begin{aligned} r!c_r &= \sum_{j=0}^{r} \binom{r}{j} \lambda^{r-j} \sum_{k=0}^{j} (-1)^{j-k} \lambda^{j-k} p^k D_{j-k,j} \\ &= \sum_{k=0}^{r} (-1)^{r-k} \lambda^{r-k} p^k \sum_{j=0}^{r-k} (-1)^j \binom{r}{r-j} D_{r-k-j,r-j} \\ &= \sum_{k=[r/2]}^{r-1} (-1)^{r-k} \lambda^{r-k} p^k \sum_{j=0}^{r-k-1} (-1)^j \binom{r}{r-j} D_{r-k-j,r-j}, \end{aligned} \tag{5.19}$$

because the last sum equals zero for $k < [r/2]$. Since the D's are tabulated the c's are easily found. For example, for $r = 6$ we get

$$D_{36} - 6D_{25} + 15D_{14} = 225 - 6 \times 50 + 15 \times 6 = 15\,,$$

$$D_{26} - 6D_{15} = 274 - 6 \times 24 = 130\,, \quad D_{16} = 120\,,$$

so

$$6!c_6 = -15\lambda^3 p^3 + 130\lambda^2 p^4 - 120\lambda p^5\,,$$

in agreement with Charlier.

It follows from (5.19) that c_r for $r > 3$ consists of several terms of different orders of magnitude since p is of the order n^{-1}. Charlier extends Edgeworth's discussion of the order of magnitude to the B series.

Referring to Edgeworth, Charlier notes that the series may be written in a symbolic form analogous to that for the A series, namely

$$g(x) = \exp\left[-n\left(\frac{1}{2}p^2\nabla^2 + \frac{1}{3}p^3\nabla^3 + \dots\right)\right] p_\lambda(x)\,.$$

In the following paper (1911) Charlier employs the results for the Bernoulli model to the models of Poisson and Lexis, which so far had been discussed only in terms of the mean and the variance. He derives the third and fourth moments for use in the improved A series and the B series.

(6) **Jørgensen's analysis and implementation of the Poisson-Charlier distribution, 1916.**

The Danish actuary N. R. Jørgensen (1879–1967) took up the challenge in Charlier's 1905b and 1908 papers by providing a complete theory for the two new distributions and the corresponding series. His contributions are contained in a theoretical paper (1915) and his thesis "Investigations of frequency surfaces and correlation" (1916, 208 pp.) The title of the thesis is incomplete because the first 51 pages contain an exposition of the univariate theory and the last 72 pages contain a comprehensive set of tables for the calculation of the A and B series and the corresponding distribution functions. We shall sketch his main results for the univariate case.

Noting that

$$p_{00}(x) = \frac{\sin \pi x}{\pi x},$$

and using the derivatives of $e^{\lambda} p_{\lambda,\mu}(x)$ with respect to λ and μ he derives the Maclaurin series

$$p_{\lambda,\mu}(x) = e^{-\lambda} \frac{\sin \pi x}{\pi} \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \sum_{k=0}^{j} \binom{j}{k} \left(\frac{\lambda+\mu}{2}\right)^{j-k} \left(\frac{\lambda-\mu}{2}\right)^{k} \frac{1}{x-j+2k}, \quad -\infty < x < \infty, \tag{5.20}$$

$0 \leqq \mu \leqq \lambda$, which for integral values of x becomes

$$\begin{aligned} p_{\lambda,\mu}(\pm r) &= e^{-\lambda} \sum_{j=0}^{\infty} \frac{1}{j!(r+j)!} \left(\frac{\lambda \pm \mu}{2}\right)^{r+j} \left(\frac{\lambda \mp \mu}{2}\right)^{j} \\ &= e^{-\lambda} \left(\frac{\lambda \pm \mu}{2}\right)^{r} \sum_{j=0}^{\infty} \frac{1}{j!(r+j)!} \left(\frac{\lambda^2-\mu^2}{4}\right)^{j}, \quad r = 0, 1, \dots . \end{aligned} \tag{5.21}$$

This is a new discrete distribution which Jørgensen calls the Poisson-Charlier distribution. It is positive for all values of r, tends to zero for $r \to \infty$, its mean is μ and its variance λ, for $\mu = \lambda$ it is the Poisson distribution, for $\mu = 0$ it is symmetric and otherwise skew, and for large values of μ and λ it tends to the normal distribution.

Introducing the Bessel function of the first kind, J_r, Jørgensen writes (5.21) as

$$p_{\lambda,\mu}(r) = e^{-\lambda} (\lambda+\mu)^r F_r(t), \quad t = (\lambda^2-\mu^2)^{1/2}, \quad r = 0, \pm 1, \dots, \tag{5.22}$$

where

$$F_r(t) = J_r(it)/(it)^r .$$

He tabulates $\log F_r(t)$ to seven decimal places for $r = 0, 1, \dots, 11$ and $t = 0.0(0.1)6.0$, so that the values of $p_{\lambda,\mu}(r)$ are easily calculated.

Similarly he writes (5.20) as

$$p_{\lambda,\mu}(x) = e^{-\lambda} \frac{\sin \pi x}{\pi} \sum_{j=0}^{\infty} (-1)^j \left(\frac{(\lambda+\mu)^j}{x-j} + \frac{(\lambda-\mu)^j}{x+j}\right) F_j(t), \quad -\infty < x < \infty . \tag{5.23}$$

Finally he expresses the relation between (5.23) and (5.22) as

$$p_{\lambda,\mu}(x) = \sum_{r=-\infty}^{\infty} \frac{\sin \pi (x-r)}{\pi (x-r)} p_{\lambda,\mu}(r),$$

which shows that $p_{\lambda,\mu}(x)$ is a weighted average or the expectation of $p_{00}(x-r)$ with $p_{\lambda,\mu}(r)$ as weight. Hence, for large absolute values of the standardized variable $(x-\mu)/\sqrt{\lambda}\,, p_{\lambda,\mu}(x)$ takes on (small) positive and negative values periodically. He remarks that the continuous version may be considered as an interpolation formula based on the discrete version.

He tabulates the functions

$$\frac{\sin \pi x}{\pi x} \quad \text{and} \quad \int_0^x \frac{\sin \pi x}{\pi x}\, dx$$

to seven decimal places for $x = 0.00(0.01)10.00$.

Jørgensen's main tool in the following analysis is Thiele's cumulant generating function. He finds

$$\begin{aligned} M_p(t) &= \int_{-\infty}^{\infty} e^{xt} p_{\lambda,\mu}(x)\, dx \\ &= \exp\left[\frac{1}{2}\lambda(e^t + e^{-t} - 2) + \frac{1}{2}\mu(e^t - e^{-t})\right], \end{aligned}$$

so

$$\kappa_p(t) = \lambda(t^2/2! + t^4/4! + \dots) + \mu(t/1! + t^3/3! + \dots)\,,$$

which shows that $M_p(0) = 1$ and that the cumulants of even and uneven order equal λ and μ, respectively.

The B series is

$$g(x) = \sum_{j=0}^{\infty} (-1)^j \frac{1}{j!} c_j \nabla^j p_{\lambda,\mu}(x)\,.$$

It is easy to prove that

$$\int_{-\infty}^{\infty} e^{xt} \nabla^j p_{\lambda,\mu}(x)\, dx = (1-e^t)^j M_p(t)\,,$$

so that the moment generating function for $g(x)$ becomes

$$M_g(t) = M_p(t) \sum_{j=0}^{\infty} (-1)^j \frac{1}{j!} (1-e^t)^j c_j\,.$$

Hence,

$$\kappa_g(t) - \kappa_p(t) = \ln \sum_{j=0}^{\infty} (-1)^j \frac{1}{j!} (1-e^t)^j c_j\,,$$

which leads to

$$\begin{aligned} &(\kappa_1 - \mu)t + (\kappa_2 - \lambda)t^2/2! + (\kappa_3 - \mu)t^3/3! + (\kappa_4 - \lambda)t^4/4! + \dots \\ &= c_1 t + (c_1 + c_2 - c_1^2)t^2/2! + (c_1 + 3c_2 + c_3 - 3c_1^2 - 3c_1c_2 + 2c_1^3)t^3/3! + \dots\,, \end{aligned}$$

from which the c's are found in terms of μ, λ and the cumulants of $g(x)$.

Setting $\mu = \kappa_1$ and $\lambda = \kappa_2$ Jørgensen gets $c_1 = c_2 = 0$ and $c_3 = \kappa_3 - \kappa_1$. Continuing we get $c_4 = \kappa_4 - \kappa_2 - 6(\kappa_3 - \kappa_1)$. Jørgensen points out that c_4 becomes $\kappa_4 - \kappa_2$ if central differences are used in the expansion instead of backward differences.

Referring to the hypothesis of elementary errors Jørgensen points out that for the series with $\mu = \kappa_1, c_2$ is of order n^{-1}, c_3 and c_4 of order n^{-2}, c_5 and c_6 of order n^{-3}, in accordance with Charlier's result for the binomial, see (5.19).

Jørgensen maintains that applications of the B series by means of his tables are just as easy as applications of the A series. He demonstrates this by analysing three sets of data, previously analysed by Pearson, Charlier and others, comparing the graduated values by his method by those obtained by a type A series and a Pearson distribution. In the applications the origin of x should be chosen such that $0 \leqq \mu \leqq \lambda$ for the formulas to be directly applicable.

He finds the existing tables for calculating the A series unsatisfactory and he therefore tabulates the normal density $\phi(x)$, its integral and its first six derivatives to seven decimal places for $x = 0.00(0.01)4.00$. Moreover, he tabulates the Hermite polynomials of order 2 to 6 for the same arguments.

Jørgensen ends by saying that there are two unsolved mathematical problems in his thesis: The convergence of the series and the justification of the operations leading to the cumulants. However, in statistical applications only a finite number of terms is used and the usefulness of the series should be judged from the goodness of fit.

Another Danish actuary, J. F. Steffensen (1873–1961), from 1919 Professor of Actuarial Mathematics at the University of Copenhagen, proved (1916) that $M_p(t)$ diverges so not even the first moment of $p_{\lambda,\mu}(x)$ exists. Jørgensen's formal operations leading to the cumulants are thus invalid. This seemed to be a serious blow to the applications of Jørgensen's results for how could one estimate the coefficients by the method of moments when the theoretical moments do not exist? We have not found any reply from Jørgensen, but the answer is simple. The "defect" refers only to the continuous version of the distribution, for the discrete version the moments exist. Jørgensen's successful fittings of $p_{\lambda,\mu}(x)$ to data depend on the fact that for a grouped continuous distribution the areas are replaced by ordinates so he really fits the discrete version to the data even if he afterwards interprets the result as a continuous distribution.

We conclude that Charlier's derivation of $p_{\lambda,\mu}(x)$ from the hypothesis of elementary errors led to a new discrete distribution but that his extension to the continuous version was a failure. The reason for this is not that the moments do not exist but the fact that the continuous version is not a frequency function and therefore should not be used as a first approximation to an arbitrary frequency function.

(7) **Charlier's C series, 1928.**

Charlier (1928) remarks that the A series has the defects that its partial sums sometimes give negative frequencies and that the successive terms of the series do not decrease regularly. However, by developing the logarithm of the density in a series instead of the density itself these defects are remedied. He defines the C series as

$$\ln g(x) = \sum_{j=0}^{\infty} c_j H_j(u)\,, \quad u = (x-\mu)/\sigma\,, \tag{5.24}$$

and compares this series with the A series in standard form

$$g(x) = \sigma^{-1}\phi(u)[1 + \sum_{j=3}^{\infty} a_j H_j(u)]\,,$$

from which he gets

$$\ln g(x) = -\ln\sigma - H_0 \ln\sqrt{2\pi} - \frac{1}{2}H_0 - \frac{1}{2}H_2 + \ln(1 + \sum_{j=3}^{\infty} a_j H_j)\,, \quad H_j = H_j(u)\,. \tag{5.25}$$

Expanding the last term in powers of H_j and expressing these powers as linear combinations of the H's he gets (5.25) written as a linear combination of H's, which compared with (5.24) gives the c's as functions of the a's. In this way he derives the first nine c's, and since the order of magnitude of the a's is known, he can find the order of the c's, which are $c_0 = c_2 = O(1)$, $c_1 = O(n^{-3/2})$ and $c_j = O(n^{1-j/2})$, $j = 3, 4, \ldots, 9$. He conjectures that the last relation holds for all $j \geqq 3$.

Using the orthogonality of the Hermite polynomials he finds

$$c_j = \frac{1}{j!}\int_{-\infty}^{\infty} H_j(u) \ln g(x)\, du\,.$$

He gives two examples of fitting a C series to large data sets.

Charlier's conjecture about the order of magnitude of the c's was proved by Aitken and Oppenheim (1930). They remark that estimates of the c's will be much influenced by the large negative values of the empirical $\ln g(x)$, which may occur at extreme values of x, because the corresponding relative frequencies are small and unreliable.

(8) **Cramér's completion of the theory for the normal A series and the Edgeworth series, 1926a, 1928.**

The analysis of the normal A series and the Edgeworth series for a continuous distribution function culminated with the works of H. Cramér (1893–1985), Professor of Actuarial Mathematics and Mathematical Statistics at the University of Stockholm, the results are summarized in his textbook (1946, pp. 221–231). After a preliminary paper (1926a), he published the important 1928 paper, which contains four main results: (1) the determination of the coefficients of the two series, (2) conditions for the convergence of the series, (3) conditions for the remainder term of the Edgeworth series to be of the same order as the first term neglected, and (4) fitting of the two four-parameter partial sums of the series to five large data sets. The proofs under (2) and (3) are mathematically difficult and Cramér's paper induced other probabilists to look for simpler proofs.

We shall sketch Cramér's elegant derivation of the normal A series and the Edgeworth series for the case of equal components. Let the cumulants of $x_j - E(x_j)$ be $\bar{\kappa}_1, \bar{\kappa}_2, \ldots$, and let $\psi_1(t)$ be the characteristic function for $x_j - \bar{\kappa}_1$ so that the characteristic function for the standardized sum

$$u = \frac{s_n - n\bar{\kappa}_1}{\sqrt{n\bar{\kappa}_2}} = \sum_{j=1}^{n} \frac{x_j - \bar{\kappa}_1}{\sqrt{n\bar{\kappa}_2}}$$

equals

$$\psi(t) = \psi_1^n(t/\sqrt{n\bar{\kappa}_2}) .$$

Cramér remarks that the A series is obtained by expanding $\psi(t)$ in powers of t whereas the Edgeworth series is obtained by expanding the same function in powers of $n^{-1/2}$.

Setting $\lambda_j = \bar{\kappa}_j(\bar{\kappa}_2)^{-j/2}$, so that $\lambda_1 = 0$, and $\lambda_2 = 1$, he obtains the two expansions

$$\ln \psi(t) = -t^2/2 + n \sum_{j=3}^{\infty} (\lambda_j/j!)(it/\sqrt{n}\,)^j \tag{5.26}$$

and

$$\ln \psi(t) = -t^2/2 + (it)^2 \sum_{j=1}^{\infty} n^{-j/2} \lambda_{j+2}(it)^{j+2}/(j+2)! . \tag{5.27}$$

Noting that the logarithm of the characteristic function for the normal A series equals

$$-t^2/2 + \ln(1 + \sum_{j=3}^{\infty} c_j(-it)^j/j!) ,$$

and comparing with (5.26) he gets the generating function for the c's in terms of the cumulants, that is, he rediscovers Thiele's formula (2.18). He does not as Edgeworth go on to find the explicit formula for the c's, see (2.19) and (4.13).

To find the coefficient of $n^{-j/2}$ in the expansion of $\psi(t)$ Cramér uses (5.27) to get

$$\psi(t) = \exp(-t^2/2) \sum_{k=0}^{\infty} \frac{(it)^{2k}}{k!} \left[\sum_{j=1}^{\infty} \frac{\lambda_{j+2}}{(j+2)!} \left(\frac{it}{\sqrt{n}} \right)^j \right]^k ,$$

which shows that the coefficient of $n^{-j/2} \exp(-t^2/2)$ is of the form

$$\sum_{k=1}^{j} b_{j,j+2k}(it)^{j+2k} .$$

Using the formula

$$(-it)^j \exp(-t^2/2) = \int_{-\infty}^{\infty} \exp(itu)\phi^{(j)}(u)\, du\,,$$

he obtains the inversion

$$\sqrt{n\bar{\kappa}_2}\, p(s_n) = \phi(u) + \sum_{j=1}^{\infty}(-1)^j n^{-j/2} \sum_{k=1}^{j} b_{j,j+2k}\phi^{(j+2k)}(u)\,. \tag{5.28}$$

Writing t^r for $(-1)^r\phi^{(r)}(u)$ Cramér observes that the coefficient of $n^{-j/2}$ is a polynomial in t, $P_j(t)$ say, with the generating function

$$1 + \sum_{j=1}^{\infty} P_j(t)z^j = \exp\left[\sum_{j=1}^{\infty} \lambda_{j+2}t^{j+2}z^j/(j+2)!\right]\,. \tag{5.29}$$

It will be seen that Cramér's (5.28) is the same as Edgeworth's (4.12) and (4.15), and that (5.29) is the same as (4.11) for $-D = t$.

Although Cramér with respect to the derivation of the two series did not produce new results or a new method of proof his paper had a great influence because of its straightforward mathematics and clear formulations compared with Edgeworth's somewhat obscure exposition.

In contradistinction to Lipps (1901), Cramér (1928, p. 156) (rashly) writes: "In some cases, the agreement between the observed values and the theoretical curves [the Edgeworth series] is even so striking that it strongly suggests the conjecture that the fundamental hypothesis may contain something which resembles the actual truth."

(9) **Wicksell's derivation of the B series, 1935.**

S. D. Wicksell (1890–1939), Professor of Statistics at the University of Lund, simplified the theory for the general B series by introducing factorial moments and cumulants. Let $g(x)$ and $f(x)$, $x = 0, 1, \ldots$, be frequency functions and set $g(x) = f(x) = 0$ for $x = -1, -2, \ldots$. Wicksell (1935) introduces the probability generating functions $G(t) = \Sigma t^x g(x)$ and $F(t) = \Sigma t^x f(x)$ and the generating function for $\nabla^j f(x)$ which equals $(1-t)^j F(t)$. By means of Maclaurin's formula he gets

$$G(1-t)/F(1-t) = \sum_{j=0}^{\infty} c_j t^j/j!\,, \quad c_j = D^j[G(1-t)/F(1-t)]_{t=0}\,, \tag{5.30}$$

so that

$$G(t) = \sum_{j=0}^{\infty} c_j(1-t)^j F(t)/j!\,,$$

which implies that

$$g(x) = \sum_{j=0}^{\infty} c_j \nabla^j f(x)/j!\,.$$

Having thus found the B series for $g(x)$ in terms of $f(x)$ he derives various conditions for the convergence of the series.

To find a more manageable expression for c_j than (5.30) he uses the Maclaurin expansion

$$1/F(1-t) = \sum_{j=0}^{\infty} a_j t^j/j!\,, \quad a_j = D^j[1/F(1-t)]_{t=0}\,,$$

which inserted into (5.30) gives

$$G(1-t)\sum_{j=0}^{\infty} a_j t^j/j! = \sum_{j=0}^{\infty} c_j t^j/j!\,, \tag{5.31}$$

so that c_j equals the jth derivative of the left side for $t = 0$.

However, $G(1-t)$ is the generating function for the descending factorial moments,

$$G(1-t) = \sum_x (1-t)^x g(x) = \sum_{k=0}^{\infty} (-1)^k \mu_{(k)} t^k/k!\,,$$

so

$$D^k G(1-t)\,|_{t=0} = (-1)^k \mu_{(k)}\,.$$

Using Leibniz's formula for differentiating (5.31) Wicksell obtains

$$c_j = \sum_{k=0}^{j} (-1)^k \binom{j}{k} \mu_{(k)} a_{j-k}\,, \quad j = 1, 2, \dots\,, \tag{5.32}$$

which shows that c_j is a linear combination of the factorial moments of $g(x)$ with coefficients depending on the factorial moments of $f(x)$. Wicksell does not find the a's, but differentiating the relation

$$F(1-t)\sum_{j=0}^{\infty} a_j t^j/j! = 1\,,$$

we get

$$\sum_{k=0}^{j} (-1)^k \binom{j}{k} \nu_{(k)} a_{j-k} = 0\,, \quad j = 1, 2, \dots\,, \tag{5.33}$$

which gives a recursion formula for the determination of a_j in terms of the factorial moments $\nu_{(1)}, \dots, \nu_{(j)}$ of $f(x)$.

Although Wicksell's formula for c_j in terms of factorial moments is considerably simpler than Charlier's (5.11) in terms of power moments, it is still rather complicated for $j \geqq 4$. Wicksell therefore looks for a recursion formula, which he finds by introducing factorial cumulants defined by the equation

$$\ln G(1-t) = \sum_{j=1}^{\infty} \kappa_{(j)} t^j / j!\,,$$

so that the relation between the factorial cumulants and the factorial moments becomes

$$\exp\left[\sum_{j=1}^{\infty} \kappa_{(j)} t^j / j!\right] = \sum_{j=0}^{\infty} (-1)^j \mu_{(j)} t^j / j!\,, \tag{5.34}$$

analogous to the relation between ordinary cumulants and moments. Denoting the factorial cumulants of $f(x)$ by $\lambda_{(j)}$, $j = 1, 2, \ldots$, it follows that

$$G(1-t)/F(1-t) = \exp\left[\sum_{j=1}^{\infty} (\kappa_{(j)} - \lambda_{(j)}) t^j / j!\right] = \sum_{j=0}^{\infty} c_j t^j / j!\,.$$

Since Thiele's recursion formula for the cumulants,

$$\mu_j = \sum_{k=0}^{j-1} \binom{j-1}{k} \mu_{j-1-k} \kappa_{k+1}\,, \quad j = 1, 2, \ldots\,,$$

obviously holds also for the factorial cumulants, Wicksell gets the recursion formula

$$c_j = \sum_{k=0}^{j-1} \binom{j-1}{k} (\kappa_{(k+1)} - \lambda_{(k+1)}) c_{j-1-k}\,, \quad j = 1, 2, \ldots\,, \tag{5.35}$$

which corresponds to the recursion formula for the coefficients in the A series derived by Thiele (1889), see Hald (2000a).

This result may be used to write the B series in a form analogous to the A series by setting

$$\gamma_j = \kappa_{(j+2)} - \lambda_{(j+2)}\,, \quad j = 0, 1, \ldots$$

see (2.13).

Wicksell illustrates his method by several examples. Considering the expansion of $g(x)$ in terms of the Poisson distribution he finds $1/F(1-t) = \exp(\lambda t)$ and thus $a_j = \lambda^j$ so that (5.32) gives

$$c_j = \sum_{k=0}^{j} (-1)^k \binom{j}{k} \mu_{(k)} \lambda^{j-k}\,, \quad j = 1, 2, \ldots\,, \tag{5.36}$$

which is a considerable simplification of Charlier's result (5.14). Wicksell believes that his formula is new, it is however due to Lipps, see (3.17), and it had also been derived by Ch. Jordan (1927, p. 39). As a special case he finds the B series for the binomial (n, p) using that $\mu_{(k)} = n^{(k)} p^k$ and $\lambda = np$, a result also found by Jordan (1927, pp. 39, 99). Wicksell also derives the recursion formula

$$c_j = -np \sum_{k=1}^{j-1} (j-1)^{(k)} p^k c_{j-1-k}\,, \; c_1 = 0\,, \quad j = 2, 3, \ldots\,.$$

Finally, Wicksell gives a comprehensive discussion of expansions with $f(x)$ equal to the binomial (n, p) using that

$$a_j = (n + j - 1)^{(j)} p^j\,, \quad j = 1, 2, \ldots\,.$$

He discusses the Bernoulli, Poisson and Lexis models and moreover the hypergeometric and Pascal distributions. In each case he gives the explicit formula for c_j.

(10) **Andersson's derivation of the A and B series, 1944, and his applications of the Gram series, 1941, 1942.**

The final refinement of the derivation of Charlier's two series is due to the Swedish actuary W. Andersson (1903–1984). Using a modified version of Wicksell's method, Andersson (1944) shows, like Charlier, how both series can be derived by the same method, the simplification is obtained by using moment generating functions and cumulants instead of ordinary moments.

For a continuous distribution we have

$$M_g(t) = \sum_{j=0}^{\infty} \mu_j t^j / j! = \exp(\sum_{j=1}^{\infty} \kappa_j t^j / j!)\,,$$

$$M_f(t) = \sum_{j=0}^{\infty} \nu_j t^j / j! = \exp(\sum_{j=1}^{\infty} \lambda_j t^j / j!)\,,$$

so that the Maclaurin expansion of the ratio gives

$$M_g(t) = M_f(t) \sum_{j=0}^{\infty} c_j t^j / j!\,, \quad c_j = D^j [M_g(t)/M_f(t)]_{t=0}\,. \tag{5.37}$$

Replacing t by it and using the inversion formula Andersson finds the A series

$$\begin{aligned} g(x) &= \sum_{j=0}^{\infty} (c_j/j!) \frac{1}{2\pi} \int e^{-ixt} M_f(it)(it)^j \, dt \\ &= \sum_{j=0}^{\infty} (c_j/j!)(-1)^j D_x^j \frac{1}{2\pi} \int e^{-ixt} M_f(it) \, dt \\ &= \sum_{j=0}^{\infty} (-1)^j (c_j/j!) f^{(j)}(x)\,. \end{aligned}$$

The new expression for c_j compared with Charlier's formula (5.7) implies that

$$S_j(x) = D^j[e^{xt}/M_f(t)]_{t=0}\,. \tag{5.38}$$

Setting

$$1/M_f(t) = \sum_{j=0}^{\infty} a_j t^j/j!\,, \quad a_j = D^j[1/M_f(t)]_{t=0}\,,$$

and using Leibniz's formula Andersson finds

$$S_j(x) = \sum_{k=0}^{j} \binom{j}{k} x^k a_{j-k} = \sum_{k=0}^{j} s_{jk} x^k/k!\,,$$

so that

$$s_{jk} = j^{(k)} a_{j-k}\,, \quad k = 0, 1, \ldots, j\,,$$

where a_j is found recursively from the formula

$$\sum_{k=0}^{j} \binom{j}{k} \nu_k a_{j-k} = 0\,,$$

which corresponds to (5.33).

Introducing the cumulants into (5.37) we have

$$\exp\left[\sum_{j=1}^{\infty} (\kappa_j - \lambda_j) t^j/j!\right] = \sum_{j=0}^{\infty} c_j t^j/j!\,,$$

and setting $t = -D$ Andersson obtains the symbolic expression for the A series

$$g(x) = \exp\left[\sum_{j=1}^{\infty} (-1)^j \frac{1}{j!} (\kappa_j - \lambda_j) D^j\right] f(x)\,,$$

which previously had been derived by Thiele (1899) by another method.

From the recursion formula for the cumulants Andersson gets

$$c_j = \sum_{k=0}^{j-1} \binom{j-1}{k} (\kappa_{k+1} - \lambda_{k+1}) c_{j-k-1}\,, \quad j = 1, 2, \ldots\,. \tag{5.39}$$

For a discontinuous distribution Andersson sets $M(t) = E[(1+t)^x]$ so that

$$M_g(t) = \sum_{j=0}^{\infty} \mu_{(j)} t^j/j! = \exp\left[\sum_{j=1}^{\infty} \kappa_{(j)} t^j/j!\right]\,,$$

$$M_f(t) = \sum_{j=0}^{\infty} \nu_{(j)} t^j/j! = \exp\left[\sum_{j=1}^{\infty} \lambda_{(j)} t^j/j!\right]\,,$$

and

$$M_g(t) = M_f(t) \sum_{j=0}^{\infty} c_j t^j / j!\,, \quad c_j = D^j [M_g(t)/M_f(t)]_{t=0}\,. \tag{5.40}$$

Noting that

$$\sum_x t^x \nabla^j f(x) = (-1)^j (t-1)^j M_f(t-1)\,,$$

it follows that

$$M_g(t-1) = \sum_{j=0}^{\infty} c_j (t-1)^j M_f(t-1)/j!\,,$$

which implies that

$$g(x) = \sum_{j=0}^{\infty} (-1)^j c_j \nabla^j f(x)/j!\,,$$

which is Charlier's general B series.

By the same reasoning as for the A series Andersson proves that

$$T_j(x) = D^j [(1+t)^x / M_f(t)]_{t=0}\,,$$

which leads to the relation

$$t_{jk} = j^{(k)} a_{(j-k)}\,, \quad k = 0, 1, \ldots, j\,,$$

where $a_{(j)}$ is found recursively from the formula

$$\sum_{k=0}^{j} \binom{j}{k} \nu_{(k)} a_{(j-k)} = 0\,.$$

Introducing the factorial cumulants into (5.40) and setting $t = -\nabla$ Andersson obtains the symbolic expression for the B series

$$g(x) = \exp\left[\sum_{j=1}^{\infty} (-1)^j \frac{1}{j!} (\kappa_{(j)} - \lambda_{(j)}) \nabla^j \right] f(x)\,,$$

which had been indicated by Charlier (1909) for the special case of the Poisson B series for the binomial.

Finally, Andersson derives the recursion formulas for the c's in the same form as (5.39) but with factorial cumulants instead of ordinary cumulants.

He generalizes the formulas for the B series to forward and central differences instead of backward differences.

Andersson (1941) gives a clear account of Gram's (1879) orthogonal series expansion, $g(x) = f(x) \Sigma c_j P_j(x)$, which he proposes to call the Gram series. He points out that the two main examples are the series with the normal and the Poisson distributions as leading terms and supplements these by setting $f(x) = b(x)$, the binomial with parameters (n, p).

We remark that this series had been treated by Bruns (1906b) and Wicksell (1935) but only in the form $g(x) = \Sigma c_j \nabla^j b(x)$, which is unsuitable for applications because tables of $b(x)$ were lacking.

Andersson succeeds in finding the orthogonal polynomials with $b(x)$ as weight function. He introduces $P_j(x)$ by the equation

$$\begin{aligned} b(x)P_j(x) &= (-1)^j q^j \Delta^j [b(x)x^{(j)}] \\ &= q^j \sum_{k=0}^{j} (-1)^k \binom{j}{k} b(x+k)(x+k)^{(j)} . \end{aligned} \tag{5.41}$$

Repeated applications of the relation

$$qb(x+k)(x+k) = (n-x-k+1)pb(x+k-1)$$

give

$$q^k b(x+k)(x+k)^{(k)} = (n-x)^{(k)} p^k b(x) ,$$

so that

$$\begin{aligned} P_j(x) &= \sum_{k=0}^{j} (-1)^k \binom{j}{k} p^k q^{j-k} (n-x)^{(k)} x^{(j-k)} \\ &= \sum_{k=0}^{j} (-1)^j \binom{j}{k} (n-j+k)^{(k)} p^k x^{(j-k)} . \end{aligned}$$

Using summation by parts he proves the orthogonality and gets

$$\sum_{x=0}^{n} P_j^2(x) b(x) = j! n^{(j)} p^j q^j .$$

The c's are then easily found from (2.7).

As a further application of the Gram series Andersson (1942) derives the A series with the Pearson curves as leading terms; this is a generalization of Gram's two A series and of Romanovsky's (1924, 1928) series, which will be discussed in the next section.

Pearson's frequency functions are defined as solutions to the differential equation

$$f'(x)/f(x) = (a+x)/(b_0 + b_1 x + b_2 x^2) ,$$

where the four parameters have to satisfy certain conditions for the solution to be a frequency function with finite moments. Andersson sets

$$f(x)P_j(x) \propto D^j [f(x)(b_0 + b_1 x + b_2 x^2)^j] \tag{5.42}$$

and proves the orthogonality using integration by parts. Choosing the coefficient of x^j as unity he finds the norming constant and a recurrence formula for the coefficients of $P_j(x)$, from which he obtains the first four rather complicated polynomials.

It will be seen that (5.41) and (5.42) are generalizations of Gram's results for the gamma distribution.

He derives a recurrence formula for $P_j(x)$, expressing P_{j+2} in terms of P_{j+1} and P_j, and a second order differential equation for $P_j(x)$, and points out that the Jacobi, Hermite, Laguerre and Legendre polynomials are special cases of $P_j(x)$. These results are also proved by Jackson (1941, pp. 161–165).

Andersson's three papers are the end of the story. They contain the simplest possible derivation of the A and B series and generalize Gram's results.

6 The contributions of Romanovsky, Jordan, and Steffensen

(1) **Romanovsky's generalization of Pearson's frequency functions by means of the A series, 1924, 1928.**

V. I. Romanovsky (1879–1954), Professor of statistics at the University of Tashkent, uses (1924) some of Pearson's frequency functions as leading terms of the A series, which he writes in the same form as Gram, whose work he did not know.

His main example is the beta distribution

$$f(x) = (a+x)^{\alpha}(b-x)^{\beta}/[(a+b)^{\alpha+\beta+1}B(\alpha+1,\beta+1)], \; -a \leqq x \leqq b\,, \; \alpha > -1\,, \; \beta > -1.$$

Setting

$$f(x)P_j(x) = D^j[f(x)(a+x)^j(b-x)^j]\,, \quad j = 0, 1, \dots\,,$$

denoting the ascending factorial by

$$(\beta+h)_k = (\beta+h)(\beta+h+1)\dots(\beta+k)\,, \quad k \geqq h\,, \quad \text{and} \quad (\beta+h)_{h-1} = 1\,,$$

and using Leibniz's formula for the differentiation he gets

$$P_j(x) = \sum_{k=0}^{j}(-1)^k\binom{j}{k}(\alpha+k+1)_j(\beta+j-k+1)_j(a+x)^k(b-x)^{j-k}\,.$$

He proves the orthogonality of the P's by integration by parts and finds

$$\int_{-a}^{b} P_j^2(x)f(x)\,dx = B(\alpha,\beta)j!(\alpha+\beta+j+1)_{2j}(a+b)^{\alpha+\beta+2j+1}(\alpha)_j(\beta)_j/(\alpha+\beta)_{2j+1}\,.$$

The coefficient c_j is then found from (2.7) and becomes a linear combination of the first j moments of $g(x)$.

Romanovsky proves that if the method of moments is used for determining the four parameters of $f(x)$ then $c_1 = \ldots = c_4 = 0$, so the A series becomes

$$g(x) = f(x)[1 + c_5 P_5(x) + c_6 P_6(x) + \ldots].$$

Pearson's Type II distribution

$$f(x) = (a^2 - x^2)^\alpha / [(2a)^{2\alpha+1} B(\alpha+1, \alpha+1)], \quad -a \leqq x \leqq a, \quad \alpha > -1,$$

is a special case of the Type I (beta) distribution and requires no further comments.

The Type III (gamma) distribution is

$$f(x) = (a + x)^\alpha e^{-\beta x} \beta^{\alpha+1} / [\Gamma(\alpha+1) e^{a\beta}], \quad x \geqq -a, \quad \alpha > -1, \quad \beta > 0,$$

and leads to

$$P_j(x) = \sum_{k=0}^{j} (-1)^k \binom{j}{k} (\alpha + k + 1)_j \beta^k (a + x)^k,$$

which for $a = 0$ equals Gram's result (2.11).

Romanovsky remarks that the A series based on the normal distribution, Type VII, is well known.

He notes that the application of series involving moments of order five or more "is not always desirable" because of the large sampling error of these moments. He points out that the gamma distribution with one correction term depends on four moments only.

In a supplementary paper (1928) he derives analogous results for the remaining Pearson frequency functions and notes the least squares property of the expansions, see also Romanovsky (1927). As noted in the previous section Andersson's (1942) proof covers all Romanovsky's results.

Pearson (1924) comments that an alternative generalization of his system may be obtained by adding terms of the third and higher powers of x in the denominator of his differential equation. [It is, however to be expected that this system will be more complicated than the A series.] He maintains that for practical applications one should not use moments of higher order than four.

(2) **Ch. Jordan's orthogonal expansions of frequency functions, 1926, 1927.**

Ch. Jordan (1871–1959), Professor at the University of Technical and Economical Sciences of Budapest, gives a clear and comprehensive exposition of orthogonal expansions of frequency functions in his textbook (1927) on mathematical statistics, preceded by a paper (1926) with applications to the binomial distribution. Nearly the same material may he found in his books on the calculus of finite differences (1947) and on the history of probability theory (1972), whose Hungarian editions were published in 1939 and 1956, respectively. Like his contemporaries Jordan did not know the works of Gram and Lipps, so his results on the B series were considered as new.

The 1927 book begins with a discussion of the mathematical properties of the four systems of orthogonal polynomials due to Chebyshev, Legendre, Hermite, and Lipps. Referring to Chebyshev and Charlier and using the orthogonality Jordan derives the expressions for the coefficients of the normal A series and the Poisson B series previously found by Gram and Lipps, respectively.

Let $f(x,\lambda)$ denote the Poisson frequency function with parameter λ, $f(x,\lambda) = 0$ for $x < 0$, let D denote differentiation with respect to λ and Δ differencing with respect to x. Hence,

$$Df(x,\lambda) = -\Delta f(x-1,\lambda)\,.$$

Jordan (1926; 1927, p. 36) defines the polynomial $G_j(x,\lambda)$ by the relation

$$D^j f(x,\lambda) = f(x,\lambda)G_j(x,\lambda) = (-1)^j \Delta^j f(x-j,\lambda)\,.$$

If we, like Lipps and Charlier, introduce backward differences the right side becomes

$$(-1)^j \nabla^j f(x,\lambda) = (-1)^j f(x,\lambda) P_j(x,\lambda)\,,$$

see (3.12), so that

$$G_j(x,\lambda) = (-1)^j P_j(x,\lambda)\,.$$

Jordan is the first to prove the orthogonality of the G's with respect to the weight function f. We shall give his proof in terms of the P's and ∇. Using summation by parts,

$$\sum_{x=1}^{\infty} u_x \nabla v_x = [u_x v_x]_0^{\infty} - \sum_{x=1}^{\infty} v_{x-1} \nabla u_x\,,$$

he gets

$$\begin{aligned}\sum_{x=0}^{\infty} \binom{x}{r} P_s(x) f(x) &= \sum_{x=0}^{\infty} \binom{x}{r} \nabla^s f(x)\\ &= -\sum_{x=1}^{\infty} \binom{x}{r-1} \nabla^{s-1} f(x-1)\,, \quad (r,s) = 1,2,\dots\,,\end{aligned}$$

which by iteration gives

$$(-1)^r \sum_{x=r}^{\infty} \nabla^{s-r} f(x-r) = \begin{matrix} (-1)^r & \text{for } s=r \\ 0 & \text{for } s>r\,. \end{matrix}$$

Since $P_r(x)$ is a linear combination of $\binom{x}{k}$, $k = 0, 1, \ldots, r$, it follows that

$$\sum_{x=0}^{\infty} P_r(x)P_s(x)f(x) = 0 \quad \text{for} \quad s \neq r\,.$$

Combining the relation

$$(-1)^r \sum_{x=0}^{\infty} \binom{x}{r} P_r(x)f(x) = 1\,, \qquad r = 0, 1, \ldots\,,$$

with the fact that

$$\binom{x}{r} = (-1)^r(\lambda^r/r!)P_r(x) + \text{ a linear combination of } P_k(x)\,,\ k < r\,,$$

Jordan gets

$$(-1)^{2r}(\lambda^r/r!)\sum_{x=0}^{\infty} P_r^2(x)f(x) = 1\,,$$

so that

$$\sum_{x=0}^{\infty} P_r^2(x)f(x) = r!\lambda^{-r}\,.$$

The coefficient c_r then follows from (2.7), which leads to the same result as Lipps's (3.17). Jordan points out that c_r may be written in symbolic form as $(\lambda - \mu)^r/r!$, where μ^k has to be replaced by the factorial moment $\mu_{(k)}$.

We shall now indicate some of Jordan's other ideas from the 1927 book. He (pp. 235–236) is the first to give a serious discussion of the properties of the three- and four-parameter A series considered as frequency functions. Let $g_3(x)$ be the three-parameter A series. By investigating the roots of the equation $g_3(x) = 0$ he finds the conditions for $g_3(x)$ to be non-negative expressed in terms of μ_2 and μ_3. He also finds the conditions for unimodality. He indicates that similar results hold for $g_4(x)$ and concludes that the applicability of these formulas as frequency functions is severely restricted, as later confirmed by Barton and Dennis (1952) who did not know Jordan's work.

He (p. 93) remarks that the binomial cannot be expressed rigorously by means of the A series, instead one should use the B series (p. 99).

He (pp. 237–239) mentions that the goodness of fit may be measured by the residual sum of squares, using either the relative frequencies or the cumulative relative frequencies, but he does not discuss the distribution of these statistics. Independently the latter measure was studied in more detail by Cramér (1926b, pp. 111–112; 1928, pp. 144–156) and by von Mises (1931, pp. 316–335), who named it the ω^2 test.

Jordan (p. 275) suggests that $\log g(x)$ may be represented by a polynomial, which he writes as a linear combination of Chebyshev polynomials; $g(x)$ is thus a

C series in the terminology of Charlier (1928). Jordan determines the coefficients by the method of least squares.

Besides the method of moments Jordan discusses the method of least squares without weights. This leads him (p. 277) to introduce the modified Hermite polynomials

$$\psi_r(u) = H_r(u)\sqrt{\phi(u)}\,, \quad u = [x - E(x)]/\sigma\,, \quad r = 0, 1, \dots\,,$$

which he uses for an orthogonal expansion with coefficients determined by the method of least squares. The coefficients are thus linear combinations of

$$\int u^r \sqrt{\phi(u)}\, g(x)\, du\,, \quad r = 0, 1, \dots\,.$$

He (p. 279) claims that this series is better than the usual A series because it is based on the method of least squares and the influence of large deviations is modified by the factor $\sqrt{\phi(u)}$. He (p. 280) makes an analogous modification of the B series.

In two notes at the end of his book he proves that the method of moments and unweighted least squares give the same expansion for Chebyshev and Legendre polynomials. Apart from a remark on the A series (p. 238) he does not realize that the method of moments and weighted least squares lead to the same expansion for the normal A series and the Poisson B series.

Uspensky (1931) characterizes Jordan's (1926) B series as "a remarkable series capable of representing a given infinite sequence of numbers under rather general conditions." He proves that the coefficients and the series converge if the convergence radius for the generating function $\Sigma g(x)t^x$ is larger than 2. He writes the distribution function as

$$\sum_{x=0}^{m} g(x) = (c_0/m!)\int_{\lambda}^{\infty} e^{-t}t^m dt + \sum_{j=1}^{\infty}(-1)^j c_j \Delta^{j-1} f(m+1-j, \lambda)\,, \quad m = 0, 1, \dots\,.$$

He applies this formula to the binomial (n, p) and finds the rapidity of the convergence by determining an upper limit for $|c_j|$.

Referring to Jordan (1926), Aitken (1931–32) derives the properties of the Lipps polynomials and points out that they satisfy a recurrence relation

$$P_{j+1}(x) = P_1(x)P_j(x-1) + (j/\lambda)P_{j-1}(x-1)\,, \quad P_1(x) = 1 - (x/\lambda)\,, \quad j = 1, 2, \dots\,,$$

which is analogous to that for the Hermite polynomials. He studies the corresponding C series, that is, the expansion of $\log g(x)$ in terms of $\{P_j(x)\}$, and concludes that it is unsatisfactory because of the large influence of the logarithm of the small probabilities.

(3) **Steffensen's unified derivation of the finite A and B series as probability distributions, 1924, 1930.**

Steffensen looks at the problem of series expansion of frequency functions from a purely statistical point of view. He (1930) writes: "We are therefore of opinion that the labour that has been expended by several authors in examining the conditions under which the A-series is ultimately convergent, interesting as it is from the point of view of mathematical analysis, has no bearing on the question of the statistical applications." The statistical problem is, he says, to fit a frequency function, $g_m(x)$ say, containing m parameters to a sample of n observations, $m < n$, and therefore the series has to be finite. Moreover, $g_m(x)$ should be a probability distribution, that is, it should be non-negative and its sum or integral over the whole domain should be unity. He therefore writes the series as

$$g_m(x) = \sum_{j=0}^{m} a_j f(x - j\omega)\,, \quad \sum_{j=0}^{m} a_j = 1\,, \tag{6.1}$$

where $f(x)$ is a probability distribution, ω an arbitrary real number, and the constants $\{a_j\}$, without being necessarily all positive, are chosen such that $g_m(x)$ is non-negative.

The basic ideas and the solution of the problem for $\omega = 1$ are given in the 1924 paper; in 1930 he derives both the A and B series as special cases of a general formula.

Steffensen introduces the moments (in our notation)

$$\mu_r = \sum_x x^r g_m(x)\,, \quad \nu_r = \sum_x x^r f(x)\,, \quad \gamma_r = \sum_j j^r a_j\,, \quad r = 0, 1, \dots\,,$$

where γ_r is defined in analogy with μ_r and ν_r although $\{a_j\}$ is not a probability distribution. The corresponding cumulants are denoted by κ_r^g, κ_r^f, and κ_r^a. For the continuous case the sums are replaced by integrals.

Steffensen derives a relation between the three moments by inserting

$$x^r = \sum_{s=0}^{r} \binom{r}{s} (x - j\omega)^s (j\omega)^{r-s}\,,$$

into

$$\mu_r = \sum_x x^r \sum_j a_j f(x - j\omega)\,,$$

which leads to

$$\begin{aligned} \mu_r &= \sum_j a_j \sum_s \binom{r}{s} (j\omega)^{r-s} \nu_s \\ &= \sum_{s=0}^{r} \binom{r}{s} \omega^{r-s} \nu_s \gamma_{r-s}\,. \end{aligned} \tag{6.2}$$

The corresponding relation between the cumulants is obtained by multiplying the two cumulant generating functions

$$\exp\left[\sum_{j=1}^{\infty} \kappa_j^f t^j/j!\right] = \sum_{j=0}^{\infty} \nu_j t^j/j!\,,$$

and

$$\exp\left[\sum_{j=1}^{\infty} \kappa_j^a (\omega t)^j/j!\right] = \sum_{j=0}^{\infty} \gamma_j (\omega t)^j/j!\,,$$

and identifying the coefficients of t^r, which on the left side equals

$$(\kappa_r^f + \omega^r \kappa_r^a)/r!\,,$$

and on the right side

$$\sum_{j=0}^{r} \frac{\nu_j}{j!} \frac{\gamma_{r-j}}{(r-j)!} \omega^{r-j} = \frac{\mu_r}{r!}\,,$$

according to (6.2). Comparing with the expansion

$$\exp\left[\sum_{j=1}^{\infty} \kappa_j^g t^j/j!\right] = \sum_{j=0}^{\infty} \mu_j t^j/j!\,,$$

it follows that

$$\kappa_r^g = \kappa_r^f + \omega^r \kappa_r^a\,. \tag{6.3}$$

Steffensen develops $g_m(x)$ as a linear combination of difference quotients of $f(x)$ setting

$$\begin{aligned} \nabla_\omega f(x) &= [f(x) - f(x-\omega)]/\omega \\ &= \frac{1 - E^{-\omega}}{\omega} f(x)\,. \end{aligned}$$

Hence,

$$\begin{aligned} f(x - j\omega) &= E^{-j\omega} f(x) \\ &= (1 - \omega\nabla_\omega)^j f(x) \\ &= \sum_{s=0}^{j} (-1)^s \binom{j}{s} \omega^s \nabla_\omega^s f(x)\,, \end{aligned}$$

which inserted into (6.1) gives

$$\begin{aligned} g_m(x) &= \sum_{s=0}^{m} \frac{(-1)^s}{s!} \omega^s \nabla_\omega^s f(x) \sum_{j=0}^{m} j^{(s)} a_j \\ &= \sum_{s=0}^{m} c_s \nabla_\omega^s f(x)\,, \quad \text{where } c_s = \frac{(-1)^s}{s!} \omega^s \gamma_{(s)}\,, \end{aligned} \tag{6.4}$$

$\gamma_{(s)}$ denoting the factorial moment. This is Steffensen's finite generalized B series from which the ordinary B series is obtained for $\omega = 1$ and the A series for $\omega \to 0$ since $\lim \nabla_\omega^s f(x) = f^{(s)}(x)$. The first term of the series is $f(x)$ since

$$c_0 = \gamma_{(0)} = \sum a_j = 1\,.$$

To find c_s, $s = 1, \ldots, m$, suppose that the theoretical or empirical value of κ_s^g is known and that κ_s^f has been calculated from $f(x)$. From (6.3) we get

$$\kappa_s^a = (\kappa_s^g - \kappa_s^f)/\omega^s\,, \tag{6.5}$$

whereafter γ_s and $\gamma_{(s)}$ are found by the usual formulas connecting the moments and the cumulants. However, for the normal A series and the Poisson B series simple expressions for c_s in terms of the moments are known.

From the formula

$$\gamma_{(s)} = \sum_{j=s}^{m} j^{(s)} a_j$$

it follows that

$$a_j = (1/j!) \sum_{s=0}^{m-j} (-1)^s \gamma_{(j+s)}/s!\,,$$

but the a's are of secondary importance compared with the c's.

If $f(x)$ contains $k < m$ parameters, determined by the first k moments, then $\kappa_s^g = \kappa_s^f$, $s = 1, \ldots, k$, so, according to (6.5), the corresponding values of $\kappa_s^a = 0$, which leads to $\gamma_{(s)} = 0$ and thus $c_{(s)} = 0$ for $s = 1., \ldots, k$.

Steffensen (1930) concludes: "There are, however, considerable drawbacks. We cannot, as with Pearson's types, be sure beforehand that negative values will not occur; as a matter of fact they often do occur, and this can only be ascertained at the end of the calculation. We have not even very good reason to expect that by adding another term such negative values may be made to disappear. (...) It may finally be observed that the frequency function (28) [our (6.4)] often presents several maxima and minima. This may be an advantage if the experience also do so; but then, such an experience is often of little value, as the presence of maxima and minima is, perhaps, due to the fact that the material is not homogeneous, or too small. We are therefore inclined to think that the apparent generality of (28) is rather a disadvantage than otherwise, and that Pearson's types are as a rule preferable."

7 Concluding remarks

Series expansions of frequency functions, which blossomed at the beginning of the century, disappeared from common statistical practice in the late 1920s.

The many authors who developed the normal A series believed that the partial sum based on the first four moments could be used as an approximation to unimodal skew distributions having contact with the axis at both ends of the range.

The justification of this belief was amply demonstrated by the many successful fittings of partial sums to empirical distributions by Werner, Bowley, Charlier, Edgeworth, and Cramér. However, the four-parameter normal A series had to compete with other systems of distributions, in particular the four-parameter Pearson system. Edgeworth (1917) remarks that his series covers only moderately nonnormal distributions and therefore has to be supplemented by the method of translation to cover nearly the same field as the Pearson system.

Pearson carried out a feud with all other systems. Not being satisfied with Edgeworth's admission of the limitation of series expansions he, in 1922, planned what he considered as the definitive blow against the normal A series. Noting that the three-parameter gamma distribution and the four-parameter beta distribution had been fitted satisfactorily to many empirical distributions he proposed to investigate whether the two theoretical distributions could be adequately represented by the normal A series. If not, he took this as a sign that the A series could not represent the empirical distributions either. He left the demonstration of this proposal to J. Henderson (1922–1923).

Henderson carried out his investigation in terms of Pearson's tetrachoric functions defined as

$$\begin{aligned}\tau_j(x) &= (-1)^{j-1}(d/dx)^{j-1}\phi(x)/\sqrt{j!}\\ &= H_{j-1}(x)\phi(x)/\sqrt{j!}\,, \quad j = 0, 1, \dots ,\end{aligned}$$

$\tau_0(x)$ being the normal probability integral. As noted by Henderson, an expansion in tetrachoric functions is the same as a normal A series when the norming factor $1/\sqrt{j!}$ is taken into account.

By a suitable change of origin and scale the two densities may be written as

$$x^{\alpha-1}e^{-x}/\Gamma(\alpha) \quad \text{and} \quad x^{\alpha-1}(1-x)^{\beta-1}/B(\alpha,\beta)\,.$$

We shall discuss Henderson's series expansion of the gamma distribution.

Introducing the standardized variable $u = (x-\alpha)/\sqrt{\alpha}$ Henderson writes the expansion as

$$x^{\alpha-1}e^{-x}/\Gamma(\alpha) = \sum_{j=0}^{\infty} c_j\sqrt{(j+1)!}\,\tau_{j+1}(u)\,. \tag{7.1}$$

Multiplying by $\exp(ut)$ and integrating he gets the relation between the moment generating functions

$$\alpha^{-\frac{1}{2}}\exp(-t\alpha^{\frac{1}{2}})(1-t\alpha^{-\frac{1}{2}})^{-\alpha} = \exp(t^2/2)\sum_{j=0}^{\infty} c_j t^j\,. \tag{7.2}$$

Expanding the functions of t into power series it is easy to see that $c_0 = \alpha^{-\frac{1}{2}}$, $c_1 = c_2 = 0$, so that the equation becomes

$$\exp\left(\sum_{j=3}^{\infty}\alpha^{-(j-2)/2}t^j/j!\right) = \alpha^{\frac{1}{2}}\sum_{j=3}^{\infty} c_j t^j\,, \tag{7.3}$$

from which the c's may be determined. It follows immediately that

$$c_3 = 1/3\alpha\,, \quad c_4 = 1/4\alpha^{3/2}\,, \quad c_5 = 1/5\alpha^2\,.$$

To determine the following c's Henderson uses recursion. Logarithmic differentiation of (7.2) leads to the equation

$$\sum_{j=0}^{\infty} c_j t^{j+2} = (\alpha^{\frac{1}{2}} - t)\sum_{j=1}^{\infty} j c_j t^{j-1}\,,$$

which gives the recursion formula

$$c_{j+1} = (jc_j + c_{j-2})/[(j+1)\alpha^{\frac{1}{2}}]\,, \quad j = 3, 4, \ldots\,.$$

By means of this formula Henderson calculates $c_6, \ldots, c_{12}$. It will be seen that c_j is a linear combination of

$$\alpha^{r-(j+1)/2}\,, \quad r = 1, \ldots, [j/3]\,, \quad j \geq 3\,.$$

If Henderson (and Pearson) had read Edgeworth (1905) more carefully they would have found that the general solution of their problem is given by (2.19) for

$$\kappa_j = (j-1)!\alpha^{-(j-2)/2}\,, \quad j = 2, 3, \ldots\,.$$

Note that (7.3) is a special case of (2.18).

Integrating (7.1) Henderson gets the A series for the distribution function, which he uses in his numerical investigations. He calculates the first 31 partial sums for $\alpha = 49$ and $u = (x - 49)/7 = -2.8, -1, 0$, corresponding to the probability integrals 0.0005850, 0.1577387, 0.5189993, respectively. The deviation of the partial sums from the exact value vary about zero in a wavelike fashion and "we have as good an approximation at the 5th or 6th terms as at the 15th, say, and better than at the 30th." Furthermore, Henderson remarks that the practical value of the expansion "depends on the convergency of the series and our experience has shown us that in the most common cases the convergency is so slight or non-existent as to render the expansion idle." He concludes that it is impossible to know where to stop to get a good approximation and that the series is of no practical utility as a representation of the gamma probability integral. He reaches a similar conclusion for the beta distribution. He does not discuss under which conditions the series converges or diverges.

It seems that Henderson is looking at the results from the point of view of numerical analysis, that is, as if the problem is to obtain an approximation formula for tabulating the incomplete gamma and beta functions to four significant figures, say, and in this respect the series fails. However, from a statistical point of view the relative error of the partial sum should be compared with the standard error of the relative frequencies of the empirical distribution in question. For the four-parameter A series the relative error in per cent for the three cases discussed is -10.4, -0.983, and -0.000385, respectively, so Henderson's argument against the A series is not so strong as he believed.

We have not found any references to Henderson's paper so whether it contributed to the decline of the statistical applications of the A series is uncertain. It is, however, certain that the Pearsonian system won the battle because of its simplicity of generation from a single formula, its coverage of distributions of widely different shapes, its easy classification of these distributions by means of the first four moments, and the goodness of fit obtained in many cases.

In due time also the Pearson system got out of fashion, instead distributions were derived from specific assumptions on the random variation in question. This development was foreshadowed, in a much more limited context, by Ranke and Greiner (1904) in their criticism of the Pearson system and its applicability in anthropology. The main problem is, they say, the analysis and comparison of several series of observations of the same phenomenon from different populations. To make that feasible we need a probabilistic model containing a small number of parameters which are simple to calculate. They underline that the parameters should have a biological interpretation, which is not the case for Pearson's frequency functions and they therefore characterize his system as purely descriptive and empirical. They remark that the interpretation of the parameters of the hypergeometric distribution, which is Pearson's starting point, is lost in the differential equation defining his system. Their own solution for anthropological data is to use the lognormal distribution, which is generated by a multiplicative combination of elementary errors. If the coefficient of variation is small, then the normal distribution may be used as an approximation. In case the lognormal distribution does not fit the data, they suspect inhomogeneity and recommend breaking up the sample into rational subgroups for which the lognormal holds. Their conclusion (p. 330) is as follows: "The mean and standard deviation [of $\log x$] give an exhaustive description of the sample, and since the probable error of these quantities is known an exact comparison of the samples is possible, and our problem is thus completely solved, if we have a reliable criterion for distinguishing between essential and inessential deviations between the empirical and theoretical distributions. This has been provided by Pearson [the χ^2 test]." They add that Pearson's system has proved very useful outside anthropology. They remark that the shape of organs is determined on the one hand by hereditary factors and main conditions of living, characterized by the mean, and on the other hand by an infinite number of elementary causes each with an infinitely small effect leading to the variation, characterized by the standard deviation.

In his reply Pearson (1905b) takes up the whole question of graduation of frequency functions by his own system and by other systems as well. By many examples he demonstrates "The need for Generalized Frequency Curves, even in Anthropological Science." He points out that Ranke and Greiner have overlooked the facts that many anthropological distributions are symmetrical without being normal, and that the lognormal distribution covers only a small part of asymmetrical distributions. He asks for a definition of "homogeneity" and notes that "for long series in economics, sociology, zoology, botany and anthropology the Gaussian curve over and over again fails. If in all these cases Ranke and Greiner assert that the material is heterogeneous they are arguing in a circle. The distributions are as continuous and smooth as those which occur in the case of the Gaussian curve, and *they occur for characters in the same group of individuals which present for other characters the normal distribution.*" He admits that his own system is empirical, and thus, by implication, that the parameters do not have a biological interpretation. Thiele (1903, p. 50) referring to Pearson (1895) notes that "Here he [Pearson] makes very interesting efforts to develop the refractory binomial functions into a basis for the treatment of skew laws of error. But there are evidently no natural links between these functions and the biological problems, and the above formulae (31) [the normal A series] will prove to be easier and more powerful instruments."

Examples of fitting partial sums of the Poisson B series to empirical distributions are rather few, see Charlier (1906), Bruns (1906b), Jørgensen (1916), A. Fisher (1922), and Aroian (1938).

Applications of the C series are given by Thiele (1903) and Charlier (1928).

Series expansions proved to be a useful tool for developing approximations to theoretical distributions with known moments. Simple examples are the approximations to the binomial by the Edgeworth series and by the Poisson B series.

The asymptotic properties of the Edgeworth series, its inversion (the Cornish-Fisher series), and its generalization are discussed by Wallace (1958), Feller (1966), Hill and Davis (1968), and subsequently by many others.

The sampling distribution of test statistics under normality, derived by R. A. Fisher in the 1920s, led in many cases to Pearson frequency functions. Several authors studied the robustness of these statistics under sampling from the normal A series, and the Romanovsky-Andersson formulas for series expansions proved useful. For surveys of this topic we refer to Wallace (1958) and Särndal (1972).

A brief history of the Gram-Charlier series is due to Davis (1983), and the historical development of approximations to distributions is discussed by Bowman and Shenton (1982).

References

Aitken, A. C., *On the orthogonal polynomials in frequencies of Type B*, Proc. Roy. Soc. Edinb. **52** (1931–1932), 174–182.

Aitken, A. C. and A. Oppenheim, *On Charlier's new form of the frequency function*, Proc. Roy. Soc. Edinb. **51** (1930), 35–41.

Andersson, W., *The binomial type of Gram's series*, Skand. Aktuarietidskr. **24** (1941), 203–213.

———, *On the Gram series on Pearson's system of frequency functions*, Skand. Aktuarietidskr. **25** (1942), 141–149.

———, *Short notes on Charlier's method for expansion of frequency functions in series*, Skand. Aktuarietidskr. **27** (1944), 16–31.

Aroian, L. A., *The type B Gram-Charlier series*, Ann. Math. Statist. **8** (1937), 183–192.

Barton, D. E. and K. E. Dennis, *The conditions under which Gram-Charlier and Edgeworth curves are positive definite and unimodal*, Biometrika **39** (1952), 425–427.

Bessel, F. W., *Untersuchungen über die Wahrscheinlichkeit der Beobachtungsfehler*, Astron. Nachrichten **15**, 358–359, 369–404; Reprinted in Abhandlungen **2**.

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

Boas, R. P., *The Charlier B-series*, Trans. Amer. Math. Soc. **67** (1949a), 206–216.

______, *Representation of probability distributions by Charlier series*, Ann. Math. Statist. **20** (1949b), 376–392.

Bortkewitsch, L. von, *Das Gesetz der kleinen Zahlen*, Teubner, Leipzig, 1898.

Bowley, A. L., *Methods of representing statistics of wages and other groups not fulfilling the normal law of error; Part II. Applications to wage statistics and other groups*, J. Roy. Statist. Soc **65** (1902), 331–354.

______, *Elements of Statistics, Fifth ed*, 1926, P. S. King & Son, London, First ed. 1901.

______, *F. Y. Edgeworth's Contributions to Mathematical Statistics*, Royal Statistical Society, London, 1928; Reprinted, Kelley, Clifton, New Jersey, USA, 1972.

Bowman, K. O. and L. R. Shenton, *Approximations to distributions*, Enc. Statist. Sci. **1** (1982), 107–118.

Bruns, H., *Ueber die Darstellung von Fehlergesetzen*, Astron. Nachrichten **143** (1897), 329–340.

______, *Zur Collectiv-Masslehre*, Wundt's Philos. Studien **14** (1898), 339–375.

______, *Wahrscheinlichkeitsrechnung und Kollektivmasslehre*, Teubner, Leipzig, 1906a.

______, *Beiträge zur Quotenrechnung*, Ber. Verh. Kgl. Sächs. Ges. Wiss. Leipzig **58** (1906b), 571–613.

Charlier, C. V. L., *Über das Fehlergesetz*, Ark. Mat. Astr. Fys **2,** No. 8, 9 pp. (1905a).

______, *Die zweite Form des Fehlergesetzes*, Ark. Mat. Astr. Fys. **2** No. 15, 8 pp. (1905b).

______, *Über die Darstellung willkürlicher Funktionen*, Ark. Mat. Astr. Fys. **2** No. 20, 35 pp. (1905c).

______, *Researches into the theory of probability*, Lunds Univ. Årsskrift, N. F. Avd. 2 **1** 5, 59 pp.; Kung. Fysiografiska Sälsk. Hand., N. F **16** 5 (1906).

______, *Weiteres über das Fehlergesetz*, Ark. Mat. Astr. Fys. **4** No. 13, 9 pp. (1908).

______, *Die strenge Form des Bernoulli'schen Theorems*, Ark. Mat. Astr. Fys. **5** 15, 22 pp. (1909).

______, *Contributions to the mathematical theory of statistics.I. Theorems of Poisson and Lexis*, Ark. Mat. Astr. Fys. **7** 17, 20 pp. (1911).

______, *Contributions to the mathematical theory of statistics*, Svenska Aktuarieför. Tidskr. **1** (1914), 86–107.

______, *A new form of the frequency function*, Lunds Univ. Årsskrift, N. F. Avd. 2 **24** 8, 28 pp.; Kungl. Fysiografiska Sälsk. Handl., N. F **39,** 8 (1928).

Chebyshev, P. L, *Sur le développement des fonctions à une seule variable*, Bull. phys.-math. Acad. Sci, Sct. Pétersbourg **1** (1859), 193–200; Oeuvres **1**, 501–508.

______, *Sur deux théorèmes relatifs aux probabilités*, Bull. phys.-math. Acad. Sci., Sct. Pétersbourg **55** (1887); Acta Math. **14**, 1890–1891; Oeuvres **2**, 481–491.

Cornish, E. A. and R. A. Fisher, *Moments and cumulants in the specification of distributions*, Rev. Inst. Intern. Statist. **5** (1937), 307–320. Fisher CP 148.

Cramér, H., *On some classes of series used in mathematical statistics*, Den sjette skandinaviska Matematikerkongres, København (Proc. 6th Scand. Math. Congr.) (1926a), 399–425; Reprinted in *CW1.*

______, *Sannolikhetskalkylen och några av dess Användningar*, Gjallarhornets Förlag, Stockholm, 1926b; [Probability Theory and some of its Applications].

______, *On the composition of elementary errors*, Skand. Aktuarietidskr. **11** (1928), 13–74, 141–180; Reprinted in *CW* **1**.

______, *Random Variables and Probability Distributions*, Cambr. Univ. Press, 1937, 2nd ed. 1962, 3rd ed. 1970.

______, *Mathematical Methods of Statistics*, Princeton Univ. Press, 1946.

______, *Collected Works, 2 vols,* (A. Martin-Löf, ed.), Springer, Berlin, 1994.

Czuber, E., *Theorie der Beobachtungsfehler*, Teubner, Leipzig, 1891.

______, *Wahrscheinlichkeitsrechnung, I.*, Teubner, Leipzig; 1908.

Davis, A. W., *Gram-Charlier series*, Enc. Statist. Sci **3** (1983); Wiley, New York. 474–477.

Draper, N. R. and D. E. Tierney, *Regions of positive and unimodal series expansion of the Edgeworth and Gram-Charlier approximation*, Biometrika **59** (1972), 463–465.

Duncker, G., *Die Methode der Variationsstatistik*, Arch. Entwicklungsmechanik der Organismen **8** (1899), 112–183.

Edgeworth, F. Y., *The asymmetrical probability curve*, Proc. Roy. Soc., London **56** (1894), 271–272.

______, *On some recent contributions to the theory of statistics*, J. Roy. Statist. Soc. **58** (1895), 506–515.

______, *On the representation of statistics by mathematical formulæ.* (Supplement.), J. Roy. Statist. Soc. **63** (1900), 72–81.

______, *The law of error*, Trans. Cambr. Phil. Soc. **20** (1905), 35–54, 113–141.

______, *The generalised law of error, or law of great numbers, (with discussion)*, J. Roy. Statist. Soc. **69** (1906), 497–539.

———, *On the representation of statistical frequency by a series*, J. Roy. Statist. Soc. **70** (1907), 102–106.

———, *On the mathematical representation of statistical data.* Section V. *Comparison with other methods*, J. Roy. Statist. Soc. **80** (1917), 411–437.

———, *Untried methods of representing frequency*, J. Roy. Statist. Soc. **87** (1924), 571–594.

Fechner, G. T., *Kollektivmasslehre* (Heraugegeben von G. F. Lipps.), Engelmann, Leipzig, 1897.

Feller, W., *An Introduction to Probability Theory and its Applications, Vol. 2*, Wiley, New York, 1966.

Fisher, A., *The Mathematical Theory of Probabilities and Its Application to Frequency Curves and Statistical Methods,* 2nd. ed. greatly enlarged, 1922, 1st ed. 1915, Macmillan, New York.

Fisher, R. A. and E. A. Cornish, *The percentile points of distributions having known cumulant*, Technometrics **2** (1960), 209–225. Fisher CP 281.

Girlich, H.-J., *Hausdorffs Beiträge zur Wahrscheinlichkeitstheorie*, Felix Hausdorff zum Gedächtnis (E. Brieskorn, Band I, 31–70, ed.), Westdeutscher Verlag, Opladen, 1996.

Gnedenko, B. V. and Kolmogorov, A. N., *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley, Cambridge, Mass., 1954; Translated from the Russian and annotated by K. L. Chung.

Gram, J. P., *Om Rækkeudviklinger, bestemte ved Hjælp af de mindste Kvadraters Methode.* (On series expansions determined by the method of least squares.), Høst, Kjøbenhavn, 1879.

———, *Über die Entwickelung reeller Functionen in Reihen mittelst der Methode der kleinsten Quadrate*, J. reine angew. Math. **94** (1883), 41–73.

Hagen, G., *Grundzüge der Wahrscheinlichkeits-Rechnung*, Dümmler, Berlin, 1837.

Hald, A., *A History of Mathematical Statistics from 1750 to 1930*, Wiley, New York, 1998.

———, *The early history of the cumulants and the Gram-Charlier series*, Intern. Statist. Rev. **68** (2000a), 137–153.

———, *On the early history of skew distributions. Thiele's three frequency functions. Preprint No. 6*, Dept. Theoretical Statistics, University of Copenhagen (2000b).

———, *On the history of the correction for grouping*, Scand. J. Statist. **28** (2001), 417–428.

Hausdorff, F., *Beiträge zur Wahrscheinlichkeitsrechnung*, Ber. Verh. Kgl. Sächs. Res. Wiss. Leipzig **53** (1901), 152–178.

Heidelberger, M., *Fechner's indeterminism: From freedom to laws of chance*, The Probabilistic Revolution, vol. I, pp. 117–156 (By L. Krüger, L. J. Daston, and M. Heidelberger, eds.), The MIT Press, Cambridge, Mass., 1987.

Henderson, J., *On expansions in tetrachoric functions*, Biometrika **14** (1922–1923), 157–185.

Hill, G. W. and A. W. Davis, *Generalized asymptotic expansions of Cornish-Fisher type*, Ann. Math. Statist. **39** (1968), 1264–1273.

Jackson, D., *Fourier Series and Orthogonal Polynomials*, The Carus Mathematical Monographs, No. 6, The Math. Soc., America, 1941.

Jordan, Ch., *Sur la probabilité des épreuves répétées, Le théorème de Bernoulli et son inversion*, Bull. Soc. Math. France **54** (1926), 101–137.

———, *Statistique Mathématique*, Gauthier-Villars, Paris, 1927.

———, *Calculus of Finite Differences*, Chelsea, New York, 1947, Hungarian ed. 1939.

Jordan, Ch. [Kàroly], *Chapters on the Classical Calculus of Probabiltiy*, Akad. Kiadó, Budapest, 1972, Hungarian ed. 1956.

Jørgensen, N. R., *Note sur la fonction de répartition de type B de M. Charlier*, Ark. Mat. Astr. Fys. **10,** No. 15, 18 pp. (1915).

———, *Undersøgelser over Frequensflader og Korrelation*; (Investigations on Frequency Surfaces and Correlation.) Busck, København, 1916.

Kendall, M. and R. L. Plackett eds., *Studies in the History of Statistics and Probability,* **2**, Griffin, London, 1977.

Laplace, P. S., *Mémoire sur les intégrales définies et leur application aux probabilités, et spécialement a la recherche du milieu qu'il faut choisir entre les résultats des observations*, Mém. Acad. Sci. Paris 1810, (1811), 279–347. OC **12** pp. 357–412.

———, *Théorie analytique des probabilités*, Courcier, Paris, 1812. 2nd ed. 1814. 3rd ed. 1820. OC **7**.

Lipps, G. F., *Ueber Fechner's Collectivmasslehre und die Vertheilungsgesetze der Collectivgegenstände*, Wundt's Philos. Studien **13** (1897), 579–612.

———, *Die Theorie der Collectivgegenstände*, Wundt's Philos. Studien **17** (1901), 79–184; 467–575.

———, *Die Theorie der Collectivgegenstände*, Engelmann, Leipzig, 1902, Reprint of Lipps (1901).

Ludwig, F., *Die Variabilität der Lebewesen und das Gauss'sche Fehlergesetz*, Zeit. Math. Phys. **43** (1898), 230–242.

Maistrov, L. E., *Probability Theory. A Historical Sketch*; Translated and edited by Samuel Kotz, Academic Press, New York, 1974.

Meyer, A., *Calcul des Probabilités,* publié sur les manuscrits de l'auteur par F. Folie, Soc. Roy. Sci., Liège, Memoires, (2) **4** (1874), 1-458; German translation by E. Czuber: *Vorlesungen über Wahrscheinlichkeitsrechnung von Dr. A. Meyer* (1879), Teubner, Leipzig.

Mises, R. von, *Über die Grundbegriffe der Kollektivmasslehre*, Jahrber. Deutsch. Math-Verein. **21** (1912), 9–20.

———, *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik*, Deuticke, Leipzig, 1931, Reprinted by Rosenberg, New York, N.Y., 1945.

Molina, E. C., *The theory of probability: Some comments on Laplace's* Théorie analytique, Bull. Amer. Math. Soc. **36** (1930), 369–392.

Pearson, K., *Contribution to the mathematical theory of evolution*, Phil. Trans., A **185** (1894), 71–110; Reprinted in *Karl Pearson's Early Statistical Papers,* Camb. Univ. Press, 1948, 1–40.

———, *Contributions to the mathematical theory of evolution*; II. *Skew variation in homogeneous material*, Phil. Trans., A **186** (1895), 343–414; Reprinted in *Karl Pearson's Early Statistical Papers,* Camb. Univ. Press, 1948, 41–112.

———, *Mathematical contributions to the theory of evolution. XIV. On the general theory of skew correlation and non-linear regression*, Drapers' Company Research Memoirs. Biometric Series, II. (1905a); Reprinted in *Karl Pearson's Early Statistical Papers,* 1948, 477–528.

———, *"Das Fehlergesetz und seine Verallgemeinerungen durch Fechner und Pearson."*, A Rejoinder. Biometrika **4** (1905b), 169–212.

———, *Note on Professor Romanovsky's generalisation of my frequency curves*, Biometrika **16** (1924), 116–117.

Poisson, S. D., *Sur la probabilité des résultats moyens des observations*, Conn. des tems pour 1827 (1824), 273–302.

———, *Recherches sur la Probabilité des Jugements en Matière Criminelle et en Matière Civile, précedées des Règles Générales du Calcul des Probabilités*, Bachelier, Paris, 1837; Translated into German by C. H. Schnuse as *Lehrbuch der Wahrscheinlichkeitsrechnung und deren wichtigsten Anwendungen,* Braunschweig, 1841.

Ranke, K. E. und R. Greiner, *Das Fehlergesetz und seine Verallgemeinerungen durch Fechner und Pearson in ihrer Tragweite für die Anthropologie*, Arch. Anthrop., N. F. **2** (1904), 295–332.

Romanovsky, V., *Generalisation of some types of the frequency curves of Professor Pearson*, Biometrika **16** (1924), 106–116.

———, *Note on orthogonalising series of functions and interpolation*, Biometrika **19** (1927), 93–99.

———, *Sur la généralisation des courbes de Pearson*, Atti del Congresso Intern. dei Matematici **6** (1928), Bologna, 107–110.

Särndal, C.-E., *The hypothesis of elementary errors and the Scandinavian school in statistical theory*, Biometrika **58** (1971), 375–392; Reprinted in Kendall and Plackett, eds., 1977.

Sheppard, W. F., *On the calculation of the most probable values of frequency-constants, for data arranged according to equidistant divisions of a scale*, Proc. London Math. Soc. **29** (1898), 353–380.

Steffensen, J. F., *Review of N. R. Jørgensen: Note sur la fonction de type B de M. Charlier*, Svenska Aktuarf. Tidskr. **3** (1916), 226–228.

———, *On Charlier's generalized frequency-function*, Skand. Aktuarietidskr. **7** (1924), 147–152.

———, *The theoretical foundation of various types of frequency-functions*, In: *Some recent researches in the theory of statistics and actuarial science*, Published for the Institute of Actuaries, Cambridge Univ. Press, 1930, pp. 35–48.

Stigler, S. M., *Francis Ysidro Edgeworth, Statistician*, J. Roy. Statist. Soc. A **141** (1978), 287–322.

———, *The History of Statistics: The Measurement of Uncertainty before 1900*, The Belknap Press of Harvard University Press, Camb. Massachusetts, 1986.

Thiele, T. N., *Om en Tilnærmelsesformel*; (On an approximation formula.) Tidsskrift for Math., Tredie række **3** (1873), 22–31.

———, *Almindelig Iagttagelseslære: Sandsynlighedsregning og mindste Kvadraters Methode.* (The General Theory of Observations: Probability Calculus and the Method of Least Squares.), Reitzel, København, 1889; Translated into English by S. L. Lauritzen in *Thorvald Nicolai Thiele. 19th Century Pioneer in Statistics.* Translation and editing: Steffen L. Lauritzen., Oxford Univ. Press (2002).

———, *Elementær Iagttagelseslære*, Gyldendal, København, 1897.

———, *Om Iagttagelseslærens Halvinvarianter.* (On the halfinvariants in the theory of observations.), Kgl. danske Videnskabernes Selskabs Forhandlinger, 1899, Nr. 3 (1899), 135–141; Translated into English in Hald (2000a)

———, *Theory of Observations*, Layton, London, 1903; Reprinted 1931, Ann. Math. Statist. **2**, 165–307.

Todhunter, I., *A History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace*, Macmillan, London, 1865, Reprinted by Chelsea, New York, 1949, 1965 and by Thoemmes, Bristol, 1993.

Uspensky, J. V., *On Ch. Jordan's series for probability*, Ann. Math., (2) **32** (1931), 306–312.

———, *Introduction to Mathematical Probability*, McGraw-Hill, New York, 1937.

Wallace, D. L., *Asymptotic approximations to distributions*, Ann. Math. Statist. **29** (1958), 635–654.

Werner, F., *Beiträge zur Collectivmasslehre*, Wundt's Philos. Studien **15** (1900), 453–500.

Wicksell, S. D., *Expansions of frequency functions for integer variates in series*, Åttonde Skand. Matematikerkongressens Förh. (1935), 306–325, (Proc. 8th Scand. Math. Congr.).

Yule, G., *Discussion on Edgeworth (1906)*, J. Roy. Statist. Soc. **69** (1906), 534–537.

Zachariae, G., *De mindste Qvadraters Methode*, Nyborg, 1871.