FITTING GENERAL GRAM-CHARLIER SERIES

By Paul A. Samuelson

Massachusetts Institute of Technology

1. Introduction. Since the last part of the nineteenth century at least, it has been common to represent a probability distribution by means of a linear sum of terms consisting of a parent function and its successive derivatives. Usually the parent function is the Type A or normal curve, as discussed by Gram [1], Bruns [2], Charlier [3], and numerous others. In addition there have been generalizations in various directions: for example, the Type B expansion in terms of the Poisson parent function and its successive finite differences.

Unlike these two types, which have a definite probability interpretation, another generalization involves the use of other parent functions and their derivatives (or differences) to give an approximate representation of a given frequency curve. With this process is associated the names of Charlier, Carver [4], Roa [5], and many others. Two general methods by which the equating of moments of the fitted curve and the given distribution yield the appropriate coefficients have been given by Charlier and Carver respectively. An account of the latter's technique is more accessible to the average English speaking statistician.

It is the purpose of the present discussion to indicate how the Charlier method may be simplified, and can be used to replace the Carver method. In doing so, I am following up the oral suggestion made some years ago by Professor E. B. Wilson of Harvard, that repeated integration by parts will yield the requisite coefficients very simply. At the same time certain methods implicit in the work of Dr. A. C. Aitken [6] show how the use of a moment generating function can often lighten the algebraic analysis. There will also be a brief indication of analogous results for general finite difference parent families; and attention will be called to a troublesome historical blunder which has permeated the statistical literature.

2. Alternative methods. Avoiding the overburdened expression generating function, I shall consider parent functions, called $f(x)$, with the restrictive properties:

a) Moments of all order of $f(x)$ exist.

b) Derivatives of any required order exist with appropriate continuity.

c) There exist high order contact at the extremities of the distribution as defined below.

Mathematically,

a) $\int_{-\infty}^{\infty} x^k f(x)dx$ is finite for all positive integral values of k

and

c) $\lim_{x \to \pm\infty} x^j f^k(x) = 0$ for all positive integral values of j and k.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org

These conditions suffice for many statistically interesting cases, but where desirable they can be lightened. Thus, derivatives may only be defined "almost everywhere," and there may be finite instead of infinite limits to the distribution, etc.

Given an arbitrary frequency curve $F(x)$, we shall suppose it to be formally expanded in the series

$$F(x) \sim a_0 f(x) + a_1 f'(x) + a_2 f''(x) + \cdots + a_n f^n(x) + \cdots . \tag{1}$$

For convenience in what follows, we shall assume that all distributions are given in terms of relative frequency so that the area under both f and F is equal to unity, so that a_0 may be taken as unity. The suppressed absolute frequency can clearly be restored at any time by multiplication of both sides with the appropriate constant. Also for algebraic convenience, many writers consider the slightly modified form of the expansion

$$F(x) \sim A_0 f(x) - \frac{A_1}{1!} f'(x) + \frac{A_2}{2!} f''(x) + \cdots \frac{(-1)^n A_n}{n!} f^n(x) + \cdots .$$

It is assumed without discussion that the first n coefficients in such a series are to be determined by equating the first n moments of each side.

I shall prove the two following identities:

$$(-1)^n a_n = L_n(F) - \sum_{0}^{n-1} L_{n-j}(f)(-1)^j a_j , \tag{2}$$

where

$$L_j(f^i) = \frac{\int_{-\infty}^{\infty} x^j f^i(x) dx}{j!} .$$

Alternatively

$$A_n = \sum_{i=0}^{n} \binom{n}{i} \frac{d^i}{d\alpha^i} \left(\frac{1}{\int_{-\infty}^{\infty} f e^{\alpha x} \, dx} \right)_{\alpha=0} \int_{-\infty}^{\infty} x^{n-i} F(x) \, dx. \tag{3}$$

The first of these which I owe to Prof. Wilson is implicit in Charlier's work. The second which may fairly be attributed to Aitken may reduce the actual work in many special cases met in practice.

Both of these methods are closely related to the Charlier device of finding polynomials $S_n(x)$ with the bi-orthogonal property

$$\int_{-\infty}^{\infty} S_n(x) f^i(x) \, dx = 0, \qquad i \neq n.$$

The subscript indicates the degree of the polynomial. By means of n of the above relationships, the polynomials can be determined except for a factor of proportionality. By formal integration of both sides of our expansion we have the Charlier identity

$$a_n = \int_{-\infty}^{\infty} S_n(x)F(x)\,dx/\text{factor of proportionality}.$$

From a theoretical standpoint, this method leaves little to be desired; but in practice the algebraic work increases rapidly with the number of terms to be included in the series.

In the Carver method, the new parent function in question, as well as the function to be approximated, are both expanded in terms of the normal curve, thus almost doubling the numerical calculations. After some differentiation, the members of the Type A family are eliminated yielding in the process the required coefficients in terms of the new parent family. We shall see below how this method may be related to the three above.

3. Useful relationship. First, two simple identities may be presented:

$$\begin{aligned} L_j(f^i) &= (-1)^i L_{j-i}(f), \quad j \geqq i \\ &= 0 \qquad\qquad\quad\;\;, \quad j < i. \end{aligned}$$

Given the above assumptions of high contact, this follows immediately from repeated integration by parts.

Remembering that the reduced moments defined just above are the coefficients of the powers of α in the series expansion of the moment generating function

$$M(\alpha; f^i) = \int_{-\infty}^{\infty} e^{\alpha x} f^i(x)\,dx = L_0(f^i) + L_1(f^i)\alpha + L_2(f^i)\alpha^2 + \cdots$$

we have the useful Aitken identity

$$M(\alpha;\, f^i) = (-1)^i M(\alpha;\, f)\alpha^i. \tag{4}$$

This, too, is the immediate consequence of repeated integration by parts.

4. Derivation of first method. Formally multiplying each side of (1) by $x^n/n!$ and integrating, we have the formal identity

$$L_n(F) = a_0 L_n(f) - a_1 L_{n-1}(f) + \cdots + (-1)^n a_n L_0(f)\,.$$

This is a "triangular" system of linear equations in the unknown a's. It may be written in matrix terms

$$\begin{bmatrix} L_0(F) \\ L_1(F) \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} = \begin{bmatrix} L_0(f) & 0 & 0 & \cdots \\ L_1(f) & L_0(f) & 0 & \cdots \\ L_2(f) & L_1(f) & L_0(f) & \cdots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \end{bmatrix} \begin{bmatrix} a_0 \\ -a_1 \\ a_2 \\ -a_3 \\ \cdot \\ \cdot \end{bmatrix}.$$

The triangular matrix has the very special property that all of its elements are known as soon as the first column is given. For this reason, as we shall see, it is essentially equivalent to a simple sequence of numbers. This we shall call the *sequence property*. Because of this special form, the above system by simple rearrangement may be written in the modified form

$$\begin{bmatrix} L_0(F) & 0 & \cdots \\ L_1(F) & L_0(F) & \cdots \\ \cdot & \cdot & \\ \cdot & \cdot & \\ \cdot & \cdot & \\ \cdot & \cdot & \end{bmatrix} = \begin{bmatrix} L_0(f) & 0 & \cdots \\ L_1(f) & L_0(f) & \cdots \\ \cdot & \cdot & \\ \cdot & \cdot & \\ \cdot & \cdot & \\ \cdot & \cdot & \end{bmatrix} \begin{bmatrix} a_0 & 0 & 0 \cdots \\ -a_1 & a_0 & 0 \cdots \\ a_2 & -a_1 & a_0 \cdots \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \end{bmatrix}$$

By appropriate definition of symbolism, this may be written in the simple matrix form:

$$L(F) = L(f)\, a(F, f),$$

since multiplication of two triangular, "sequence" matrices is commutative.

It is usually simplest to invert this triangular solution directly as in (2). But if necessary, we may express our answer in the equivalent form

$$a(F, f) = L(F)\, L(f)^{-1}, \tag{5}$$

where the inverse to any special triangular matrix, also possesses the sequence property.

If g is a second parent function with the properties of Section 2, we have the relationship

$$a(F, g) = a(F, f)\, a(f, g)$$

which follows directly from (5). This may be generalized to

$$a(f_1, f_2)\, a(f_2, f_3) \cdots a(f_{n-1}, f_n) = a(f_1, f_n)\,.$$

If F itself is a parent function, we have

$$a(F, f)\, a(f, F) = a(F, F) = I$$

or

$$a(f, F) = a(F, f)^{-1}\,.$$

5. Relation to old methods. In terms of our notation, the Carver method seems to reduce to computing $a(F, f)$ by the relationship

$$a(F, f) = a(F, \phi)\, a(f, \phi)^{-1}$$

where ϕ is the Type A parent function. It involves a doubling of the work of coefficient determination. However, if only a few terms in the expansion are retained, this is of negligible importance.

The Charlier polynomials are clearly summed rows of the matrix product

$$L(f)^{-1}\cdot\begin{bmatrix} 1/1! & 0 & 0 & \cdots \\ 0 & x/1! & 0 & \cdots \\ 0 & 0 & x^2/2! & \cdots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \end{bmatrix}.$$

To know the first n of these polynomials, it is not necessary to derive $n(n+1)/2$ different coefficients. Because of the sequence property, it is only necessary to derive n elements of the first column of $L(f)^{-1}$. These can be expressed in terms of the reduced moments of f, as did Charlier; but the relationships are non-linear and algebraically become tedious for high n. They are better computed from sequence relationships.

The above discussion suggests that the bi-orthogonal relationship between a parent family and suitable polynomials has no deep significance. In particular, there is no essential relationship to least squares as in orthogonal expansions. It does, however, share one important property with orthogonal functions—determination of later coefficients does not affect the earlier ones. But this is a property of all triangular reductions, orthogonal expansions being only special cases of these.

6. Sequence properties. Ordinarily to derive the inverse of an n^2 matrix, n^2 equations must be solved. For our triangular matrices, we need only solve n equations for one column. To each triangular matrix $L(f)$ there corresponds a sequence $\{L_k(f)\}$, which is in fact the first column of the former. Similarly to $L(f)^{-1}$, there corresponds $\{\tilde{L}_k(f)\}$; the elements of the latter are defined by the n equations

$$L_0(f)\tilde{L}_0(f) = 1$$

$$L_0(f)\tilde{L}_1(f) + L_1(f)\tilde{L}_0(f) = 0$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\sum_0^n L_k(f)\tilde{L}_{n-k}(f) = 0$$

But these are precisely the equations involved in the formal inversion of any linear operator system of the form

$$\sum_0^\infty c_k h^k y = z \tag{6}$$

where h is an operator which commutes with a constant, and for which $h^0 = 1$. z is a known function and y unknown. Thus h may be such operators as

$$x,\ d/dx,\ xd/dx,\ E,\ \Delta.$$

A particular solution of (6) is given by the formal expansion

$$y = \sum_0^\infty \bar{c}_j h^j z$$

where the $\bar{c}$'s bear the same relationship to the c's as do the $\bar{L}$'s to the L's.

Such "reciprocal" sequences appear in many branches of applied mathematics. In particular, they arise in the inversion of a power series. If formally,

$$W(\alpha) = \sum_0^\infty S_k \alpha^k$$

then

$$\frac{1}{W(\alpha)} = \sum_0^\infty \bar{S}_k \alpha^k.$$

Thus, to any triangular matrix with the sequence property, we can formally associate a function $W(\alpha)$ as well as a sequence of numbers. The calculus of multiplication of our triangular matrices clearly "corresponds" to the calculus of multiplication of functions, i.e. if the triangular matrices $T_1, T_2, \cdots T_n$ and $W_1(\alpha), W_2(\alpha), \cdots W_n(\alpha)$ correspond, and $T_n = T_1 \cdot T_2 \cdots T_{n-1}$; then

$$W_n(\alpha) = W_1(\alpha) W_2(\alpha) \cdots W_{n-1}(\alpha).$$

Also, $1/W_i(\alpha)$ corresponds to T_i^{-1}.

7. Moment generating functions. If only for the above reasons and no others, we should be tempted to consider the function formally defined by

$$\sum_0^\infty L_k(f) \alpha^k.$$

But this is precisely the expression for the familiar moment generating function, m. g. f.

$$M(\alpha; f) = \int_{-\infty}^{\infty} e^{\alpha x} f(x)\, dx = \sum_0^\infty L_k(f) \alpha^k.$$

In this way, the method of triangular matrices joins the method used by Aitken for the Type A family. If

$$F(x) \sim \sum_0^\infty a_i f^i(x),$$

and we formally equate moment generating functions of each side, we get

$$M(\alpha; F) = M(\alpha; f) \sum_0^\infty (-1)^i a_i \alpha^i, \tag{7}$$

by means of the Aitken identity (4). Thus $(-1)^i a_i$ equals the coefficient of α^i in the formal expansion of

$$\frac{M(\alpha;F)}{M(\alpha;f)} = M(\alpha;F)M(\alpha;f)^{-1}.$$

Our relationship (2) follows immediately from (7); and by Taylor's expansion in α of $M(\alpha;f)^{-1}$, the identity (3) is quickly realized.

For many problems, the reciprocal of the m. g. f. of $f(x)$ is itself a simple function; to that our triangular equations may be inverted without solving linear equations. Thus where $F(x) = f(x+b)$, we immediately verify Taylor's expansion by use of familiar properties of the m. g. f. under shift of origin.

8. Finite difference expansions. Corresponding to integration by parts, we have the formula

$$\sum_{-\infty}^{\infty} W_i \nabla^k V_i = (-1) \sum_{-\infty}^{\infty} \Delta W_i \nabla^{k-1} V_i = (-1)^2 \sum_{-\infty}^{\infty} \Delta^2 W_i \nabla^{k-2} V_i, \quad \text{etc.,}$$

provided "high contact" properties are assumed. ∇ and Δ are receding and advancing differences respectively. Recalling the familiar property of "reduced factorial" polynomials, ${}^k x$, we have

$$\begin{aligned} \sum_{-\infty}^{\infty} {}^j x \nabla^k f(x) &= (-1)^k \sum_{-\infty}^{\infty} {}^{j-k} x f(x) && j \geqq k \\ &= 0 && j < k, \end{aligned}$$

or

$$\begin{aligned} Q_j(\nabla^k f) &= (-1)^k Q_{j-k}(f) && j \geqq k \\ &= 0 && j < k, \end{aligned}$$

where

$$Q_j(g) = \sum_{-\infty}^{\infty} \frac{x(x-1)(x-2)\cdots(x-j+1)}{j!} g(x).$$

In the expansion

$$F(x) \sim a_0 f(x) + a_1 \nabla f(x) + a_2 \nabla^2 f(x) + \cdots,$$

the a's obey laws identical to (2) and (3) where reduced factorial moments are substituted for the reduced L moments, and the f. m. g. f.

$$\sum_{-\infty}^{\infty} f(x)(1+\alpha)^x,$$

for the ordinary m. g. f.

9. Convergence. All of the above relationships are purely formal, without regard to convergence. The last is a difficult subject, and little discussed in the statistical literature, since applications of G-C series have been almost entirely concerned with empirical frequency curve fitting in which mathematical convergence does not enter. Actually in the scanty treatments of the subject there has arisen a confusion between the Type A *G-C* expansion, which equates moments, and the expansion of a function in orthogonal Hermite functions. These are not unrelated, but nevertheless they are distinct. This is well recognized in the purely mathematical literature, but hardlv at all in the literature of statistics and physics.

The series differ by an irremovable factor of 2. If the Type A functions are written as

$$H_i(x)e^{-x^2},$$

then the Hermite functions will take the form

$$H_i(x)e^{-\frac{1}{2}x^2},$$

where the H's are Hermite polynomials suitably normalized. Unfortunately the *G-C* series often diverges when the H series converges. Thus, the statistically interesting Cauchy distribution can be expanded in an H series; but since it possesses no finite higher moments, the *G-C* series cannot even be defined.

It is not hard to show that the *G-C* expansion of F in terms of a Type A function $f(x)$, is equivalent to an H expansion of $Ff^{-\frac{1}{2}}$ in terms of the H family $f^{\frac{1}{2}}$. It is sufficient for convergence in the mean of the last expansion that $Ff^{-\frac{1}{2}}$ be of integrable square or belong to L^2. This means that the *G-C* type A expansion will be valid if $Ff^{-\frac{1}{2}}$ is well behaved, not simply if F is well behaved. For F a histogram as is often the case in practise, no difficulties of convergence arise, although rapid convergence may be another matter. Nevertheless, many well behaved F's will not pass the more strict test. The reader is referred to the last five titles in the bibliography for mathematical discussions of this problem.

The above discussion holds only for the Type A expansion. There remains the very difficult problem of convergence conditions in the more general case. No immediate generalization suggests itself, except the application of the results of the "moment problem." However, this must be handled with delicacy, since the partial sums of the series may actually become negative over some range.

BIBLIOGRAPHY

References

[1] J. P. Gram, "Ueber die Entwicklung reeler Functionen in Reihen mittelst der Methode der Kleinsten Quadrate," *Journal für die Reine und Angewandte Mathematik*, Vol. 94 (1882), pp. 41–73.

[2] H. Bruns, *Wahrscheinlichkeitsrechnung und Kollektivmasslehre*, B. G. Teubner, Leipzig, 1906.

[3] C. V. L. Charlier, "Über die Darstellung willkürlicher Funktionen," *Arkiv för Matematik, Astronomi och Fysik (utgivet af k. svenska vetenskapsakademien)*, Vol. 2 (1905), 35 pp.

[4] H. C. Carver, Chapter VII, *Handbook of Mathematical Statistics*, edited by H. L. Rietz, Cambridge, Massachusetts, 1924.

[5] Emeterio Roa, "A number of new generating functions with application to statistics," 1923.

[6] A. C. AITKEN, *Statistical Mathematics*, London, 1939.
[7] V. ROMANOVSKY, "Generalisation of some Types of the Frequency Curves of Professor Pearson," *Biometrika*, Vol. 16 (1924), pp. 106–116.
[8] DUNHAM JACKSON, *Fourier Series and Orthogonal Polynomials*, (Carus Mathematical Monograph No. 6), The Mathematical Association of America, Oberlin, Ohio, 1941.
[9] E. H. HILDEBRANDT, "Systems of polynomials connected with the Charlier expansions and the Pearson differêntial and difference equations," *Annals of Math. Stat.*, Vol. II, 1931, pp. 379–439.

Further Literature

1. W. MYLLER-LEBEDEFF, "Die Theorie der Integralgleichungen in Anwendung auf einige Reihenentwicklungen," *Math. Annalen*, Vol. 64 (1907), pp. 388–416.
2. E. HILLE, "A class of reciprocal functions," *Annals of Math.*, Vol. 27 (1926), pp. 427–464. (Contains selected bibliography.)
3. M. H. STONE, "Developments in Hermite polynomials," *Annals of Math.*, Vol. 29 (1927), pp. 1–13.
4. W. E. MILNE, "On the degree of convergence of the Gram-Charlier series," *Trans. Amer. Math. Soc.*, Vol. 31 (1929), pp. 422–444.
5. *Bibliography on Orthogonal Polynomials*, National Research Council of the National Academy of Sciences, Washington, D. C., 1940.