The historical development of the use of generating functions in probability theory

Autor(en): **Seal, H.L.**

Objekttyp: **Article**

Zeitschrift: **Mitteilungen / Vereinigung Schweizerischer Versicherungsmathematiker = Bulletin / Association des Actuaires Suisses = Bulletin / Association of Swiss Actuaries**

Band (Jahr): **49 (1949)**

PDF erstellt am: **28.05.2024**

Persistenter Link: https://doi.org/10.5169/seals-555164

The Historical Development of the Use of Generating Functions in Probability Theory

by *H. L. Seal*, Toronto

Introduction

An important part of probability theory consists of the derivation of the probability distribution of the sum of n random variables, each of which obeys a given probability law, and the development of asymptotic forms of these distributions valid for increasing n. Probability generating functions owe their dominant position to the simplifications they permit in both problems. Their employment to obtain the successive moments of a probability distribution and to solve the difference equations of probability theory is ancillary to their chief use in connexion with sums of random variables.

A didactic exposition of the use of generating functions in probability theory might easily be made to parallel the historical development of these functions. This circumstance will be clearly perceived in the following historical sketch of the use of probability generating functions from their origin with De Moivre to their present-day wide application under different guises.

The generating function of a discrete law

Although the theoretical frequencies of the various possible totals obtained in throws with two and three ‚perfect' dice had been known from at least the time of Cardan (Todhunter, 1865), it was De Moivre who gave the problem its first algebraic solution. Considering with De Moivre (1730) a generalized die with k faces any one of which is equally likely to appear when the die is thrown, we ask the probability that with n such dice a total of x points will be thrown.

Without any essential change in the problem we may restate it in modern terminology thus: to derive the probability distribution of the random variable formed by adding n random variables each subject to the same discrete rectangular probability law with k equidistant variate values.

De Moivre expresses his solution in terms of the favourable «chances» (combinations) and the total possible «chances». The latter number k^n in the problem under consideration. On the other hand, the total array of chances on any one of the dice may be represented algebraically by $t + t^2 + t^3 + \ldots + t^k$ the index of any t representing the number of points on the corresponding ‚face'. As De Moivre says this progression «may represent all the Chances of one Die: this being supposed, it is very plain that in order to have all the Chances of two such Dice, this Progression ought to be raised to its Square, and that to have all the Chances of three Dice, the same Progression ought to be raised to its Cube; and universally, that if the number of Dice be expressed by n, that Progression ought accordingly to be raised to the Power of n» [1]). With this preliminary De Moivre proceeds to the algebraic solution mentioned.

The words quoted show that De Moivre attached to the discrete rectangular probability law a function designed to represent that probability distribution (namely, «all the Chances of one Die»). Although the title is quite modern we refer to this type of ‚image' function as a probability generating function and write it generally as $\psi(t)$. De Moivre, then, in effect defined $\psi(t)$, the probability generating function of the discrete rectangular law, by

$$\psi(t) \equiv \sum_{x=1}^{k} \frac{1}{k} t^r = \frac{t}{k} \cdot \frac{1-t^k}{1-t} \tag{1}$$

and stated it to be «very plain» [2]) that $\psi_n(t)$ the probability generating function of the sum of n such discrete rectangular variables, should

---

[1]) Quoted from the third edition of *The Doctrine of Chances*, 1756.

[2]) In 1777 Euler (1788) presented a paper to the Academy with the sole purpose of proving inductively that the probability of a variate x appearing as the sum of n random variables each distributed according to the law

$$p(x),\ x = x_1, x_2, \ldots x_k,$$

is given by the coefficient of t^x in $\left(\sum_{x_j} p(x_j)\, t^{x_j}\right)^n$.

be given by

$$\psi_n(t) \equiv \sum_{x=n}^{nk} p_n(x)\, t^x = [\psi(t)]^n = \frac{t^n}{k^n}(1-t^k)^n (1-t)^{-n}$$

$$= \frac{t^n}{k^n}\left[\sum_{\lambda=0}^{n}\binom{n}{\lambda}(-1)^\lambda t^{k\lambda}\right]\left[\sum_{\nu=0}^{\infty}\binom{n+\nu-1}{\nu} t^\nu\right]$$

$$= \frac{1}{k^n}\sum_{x=n}^{nk} t^x \sum_{j=0}^{m_1}(-1)^j\binom{n}{j}\binom{x-kj-1}{x-n-kj},$$

$$m_1 = \min.\left(n, \left[\frac{x-n}{k}\right]\right)$$

where $p_n(x)$ is the required probability of a total x with n dice, and may be read off from the last expression written, viz.

$$p_n(x) = \frac{1}{k^n}\sum_{j=0}^{m_1}(-1)^j\binom{n}{j}\binom{x-kj-1}{x-n-kj} \qquad (2)$$

De Moivre's contemporaries were fully seized of the value of the artifice thus introduced. An important application of it was made by Simpson in one of the essays of his *Miscellaneous Tracts*. Simpson's object was to consider mathematically the method «practised by *Astronomers*» of taking the mean of several observational readings «in order to diminish the errors arising from the imperfection of instruments and of the organs of sense». He supposes that at any one reading errors in excess or defect are symmetrically disposed and have assignable upper and lower limits.

In the first of two propositions Simpson considers a discrete rectangular law with an odd number, $2h+1$, of variates centred about zero. The problem of obtaining the probability of a total error x arising as the sum of n individual errors is solved precisely in the manner of De Moivre though now $\psi(t)$ appears with a factor t^{-h} representing a removal of the origin to the centre of the distribution. Simpson finally obtains the probability that an error lies between $-z$ and z by summing the appropriate terms of the expression thus obtained.

In his second proposition Simpson assumes a discrete symmetric triangular probability law for the individual errors. He writes, effectively,

$$p(x) = \begin{cases} \dfrac{h+1-|x|}{(h+1)^2} & x = -h, -h+1, \ldots, -1, 0, 1, \ldots, h \\ 0 & x = \text{any other value} \end{cases}$$

and

$$\psi(t) \equiv \sum_{n=-h}^{h} p(x)\, t^x = \frac{t^{-h}}{(h+1)^2} \cdot \frac{(1-t^{h+1})^2}{(1-t)^2}$$

Hence

$$\psi_n(t) = \frac{t^{-nh}}{(h+1)^{2n}} (1-t^{h+1})^{2n} (1-t)^{-2n}$$

$$= \frac{1}{(h+1)^{2n}} \sum_{n=-nh}^{nh} t^x \sum_{j=0}^{m_2} (-1)^j \binom{2n}{j} \binom{2n+nh+x-\overline{h+1}\,j-1}{nh+x-\overline{h+1}\,j} \tag{3}$$

$$m_2 = \text{min.} \left(2n, \left[\frac{nh+x}{h+1}\right]\right)$$

Simpson notices that with the exception of the displacement factor [1]) t^{-nh} and the constant, the form of the generating function in the second proposition is similar to that of the first with $2n$ replacing n.

The probability of a total error lying between $\pm z$ (the limits included) when individual triangular variates have been combined is thus

$$1 - \frac{2}{(h+1)^{2n}} \sum_{x=-nh}^{-z-1} \sum_{j=0}^{m_2} (-1)^j \binom{2n}{j} \binom{2n+nh+x-\overline{h+1}\,j-1}{2n-1}, \tag{4}$$

$$z \leqslant nh$$

since, as Simpson points out, the distribution of the mean error, x/n, is symmetrical about zero and it is simpler to work with $-z$ than with z. Simpson carries out the summation of x arriving at

---

[1]) Our terminology not Simpson's.

$$1 - \frac{2}{(h+1)^{2n}} \sum_{j=0}^{m_3} (-1)^j \binom{2n}{j} \binom{2n + nh - z - \overline{h+1}\, j - 1}{2n} \tag{5}$$

$$m_3 = \text{min.} \left(2n, \left[\frac{nh - z - 1}{h+1}\right]\right)$$

and illustrates this result by putting $h = 5$, $n = 6$ and $z = 1$ and 2, respectively. He concludes that «the taking of the mean of a number of observations, greatly diminishes the chances for all the smaller errors, and cuts off almost all possibility of any large ones.»

Up to this point Simpson's analysis has been a straightforward application of De Moivre's technique to a new problem. The real advance due to Simpson lies in the corollary to the second proposition. Here for the first time a continuous (symmetric triangular) probability law is introduced. By making $h \to \infty$ in such a way that the range of variation of an individual error x remains within ± 1, the probability of a total error between $\pm y (0 \leqslant y \leqslant n)$ arising from the addition of n errors each subject to a continous triangular law centered on zero and extending one unit to the right and to the left, is given by [1])

$$1 - \lim_{\substack{h \to \infty \\ \frac{z}{h} \to y}} \frac{2}{(h+1)^{2n}} \sum_{j=0}^{m_3} (-1)^j \binom{2n}{j} \frac{1}{(2n)!} (n - 1 + \overline{n-j}\, \overline{h+1} - z)^{(2n)}$$

$$= 1 - \frac{2}{(2n)!} \sum_{j=0}^{m_4} (-1)^j \binom{2n}{j} (n - y - j)^{2n} \tag{6}$$

$$m_4 = \text{min.}\ (2n, [n - y])$$

Simpson did not attach a similar corollary to his first proposition but his observation on the interchangeability of $2n$ and n indicates that he could have written this down without further calculation. It is noted here that the probability of a total error $< | y |$ arising from the sum of n random variables each subject to a continuons rectangular probability law centered on zero and extending one unit in either direction, is

$$1 - \frac{2}{2^n n!} \sum_{j=0}^{m_5} (-1)^j \binom{n}{j} (n - y - 2j)^n \tag{7}$$

$$m_5 = \text{min.}\ (n, [n - y])$$

---

[1]) Although Simpson's formula is correctly stated he mistakenly wrote $nz/h = y$ in his three numerical examples.

The generating function of a continuous law

In an article published (apparently) in 1773 Lagrange restated Simpson's two propositions except that he placed his origins away from the centres of the distributions of the individual errors, and directed his attention to the probability of a total variate (error) lying between $-z_1$, and z_2 instead of between $\pm z$. He also used Simpson's limiting process to obtain the distribution laws of the sums of n variates from continuous rectangular and triangular laws of individual error. As Todhunter (*l. c.*) points out, there are a number of algebraic slips in Lagrange's work, but from the modern point of view his mathematical technique is considerably in advance of Simpson's more pedestrian approach.

However, the real advance of Lagrange's memoir consists in his generalization of generating functions to apply directly to continuous probability distributions. Lagrange argued, by direct analogy with the discrete case, that if

$$\psi(t) = \int_{-\infty}^{\infty} t^x p(x)\, dx \tag{8}$$

then

$$\psi_n(t) \equiv \int_{-\infty}^{\infty} t^x p_n(x)\, dx = \left[\int_{-\infty}^{\infty} t^x p(x)\, dx\right]^n \tag{9}$$

where the meaning to be attributed to $\psi_n(t)$ and $p_n(x)$ will be clear from the similar relations of the discrete case. In order to carry out the calculations indicated in these relations Lagrange provided what is, in effect, the first dictionary of Laplace transforms. This dictionary was effectively as follows, where $P_m(x)$ represents an arbitrary polynominal in x of the mth degree.

No.	$p(x)$		$\psi(t) = \int_{-\infty}^{\infty} t^x p(x)\,dx$
1	$\sum_{j=1}^{k} p_j(x)$		$\sum_{j=1}^{k} \psi_j(t)$
2	$P_m(x)$ 0	$\left.\begin{matrix} 0 \leqslant x \leqslant a \\ -\infty < x < 0,\ x > a \end{matrix}\right\}$	$\sum_{j=0}^{m} (-1)^j \dfrac{t^a P_m^{(j)}(a) - P_m^{(j)}(0)}{(\log_e t)^{j+1}}$
3	$P_m(x)\, e^{-\alpha x}$ 0	$\left.\begin{matrix} 0 \leqslant x \leqslant a \\ -\infty < x < 0,\ x > a \end{matrix}\right\}$	$\sum_{j=0}^{m} (-1)^j \dfrac{t^a e^{-a\alpha} P_m^{(j)}(a) - P_m^{(j)}(0)}{(\log_e t - \alpha)^{j+1}}$
4	x^m 0	$\left.\begin{matrix} 0 \leqslant x \leqslant a \\ -\infty < x < 0 \end{matrix}\right\}$	$\dfrac{m!}{(-\log_e t)^{m+1}}$ $0 < t < 1;\ m = 1, 2, 3, \ldots$
5	0 $(-x)^m$	$\left.\begin{matrix} 0 < x < \infty \\ -\infty < x \leqslant 0 \end{matrix}\right\}$	$\dfrac{m!}{(\log_e t)^{m+1}}$ $t > 1;\ m = 1, 2, 3, \ldots$

No.	$p(x)$	$\psi(t) = \int_{-\infty}^{\infty} t^x p(x)\, dx$
6	$\left.\begin{array}{ll} \sum_{j=0}^{m_6} A_j (x-j)^m & 0 \leqslant x < \infty,\ m_6 = \min.(h,\ [x]) \\ 0 & -\infty < x < 0 \end{array}\right\}$	$\frac{m!}{(-\log_e t)^{m+1}} \sum_{j=0}^{h} A_j t^j \quad 0 < t < 1;\ h, m = 1, 2, 3, \dots$
7	$\left.\begin{array}{ll} 0 & 0 < x < \infty \\ \sum_{j=0}^{m_7} A_j (-x-j)^m & -\infty < x \leqslant 0,\ m_7 = \min.(h,\ [-x]) \end{array}\right\}$	$\frac{m!}{(\log_e t)^{m+1}} \sum_{j=0}^{h} A_j t^{-j} \quad t > 1;\ h, m = 1, 2, 3, \dots$
8	$\left.\begin{array}{ll} \sum_{j=0}^{m_6} A_j e^{\beta(x-j)} (x-j)^m & 0 \leqslant x < \infty \\ 0 & -\infty < x < 0 \end{array}\right\}$	$\frac{m!}{(-\log_e t - \beta)^{m+1}} \sum_{j=0}^{h} A_j t^j \quad 0 < t < 1;\ h, m = 1, 2, 3, \dots$
9	$\left.\begin{array}{ll} 0 & 0 < x < \infty \\ \sum_{j=0}^{m_7} A_j e^{\beta(x-j)} (-x-j)^m & -\infty < x \leqslant 0 \end{array}\right\}$	$\frac{m!}{(\log_e t - \beta)^{m+1}} \sum_{j=0}^{h} A_j t^{-j} \quad t > 1;\ h, m = 1, 2, 3, \dots$

Lagrange points out that α of No. 3 may be purely imaginary so that the cases $P_m(x)$ cos bx and $P_m(x)$ sin bx are included. He stated that other functions $p(x)$ than those included above have generating functions $\psi(t)$ obtainable only by approximation.

Four examples of the use of the new type of probability generating function close the memoir. They are based, respectively, on the following continuous probability laws:

$$\frac{1}{h_1+h_2} \qquad -h_1 \leqslant x \leqslant h_2; \qquad 0 \quad x < -h_1,\ x > h_2 \tag{I}$$

$$\frac{l-|x|}{l^2} \qquad -l \leqslant x \leqslant l; \qquad 0 \quad |x| > l \tag{II}$$

$$\frac{3}{4}a^{-3}(a^2-x^2) \qquad -a \leqslant x \leqslant a; \qquad 0 \quad |x| > a \tag{III}$$

$$\frac{1}{2}\cos x \qquad -\frac{\pi}{2} \leqslant x \leqslant \frac{\pi}{2}; \qquad 0 \quad |x| > \frac{\pi}{2} \tag{IV}$$

We illustrate Lagrange's procedure by considering (III) which he calls the law «la plus simple et la plus naturelle qu'on puisse imaginer».

We have in this case

$$\psi(t) = \int_{-a}^{a} t^x \frac{3}{4} a^{-3}(a^2-x^2)\,dx = \int_0^a t^x \frac{3}{4} a^{-3}(a^2-x^2)\,dx + \int_0^a t^{-x} \frac{3}{4} a^{-3}(a^2-x^2)\,dx$$

$$= \frac{3}{2} a^{-2} \left\{ \frac{t^a + t^{-a}}{(\log_e t)^2} - \frac{a^{-1}(t^a - t^{-a})}{(\log_e t)^3} \right\}$$

by no. 2 of the dictionary. Hence

$$\psi_n(t) = [\psi(t)]^n = \frac{3^n}{2^n} a^{-2n} \sum_{j=0}^{n} (-1)^j \binom{n}{j} a^{-j} \frac{(t^a + t^{-a})^{n-j}(t^a - t^{-a})^j}{(\log_e t)^{2n+j}}$$

$$= \frac{3^n}{2^n} a^{-2n} \sum_{j=0}^{n} (-1)^j \binom{n}{j} a^{-j} \sum_{\lambda=n-2\left[\frac{n}{2}\right]}^{n}{}^{(2)} \left\{ \frac{A_{\lambda j}}{(\log_e t)^{2n+j}} t^{-\lambda a} + \frac{A_{\lambda j}}{(-\log_e t)^{2n+j}} t^{\lambda a} \right\}$$

where
$$A_{\lambda j} = \sum_{\nu=0}^{\frac{n-\lambda}{2}} (-1)^{j+\nu} \binom{n-j}{\frac{n-\lambda}{2}-\nu} \binom{j}{\nu},$$

except that A_{0j}, if it occurs, is to be given half the value ascribed by this formula. The «(2)» affixed to the last summation sign on p. 217 indicates that every second term is to be summed. Using the dictionary inversely (Nos. 6 and 7)

$$p_n(|x|) = \frac{3^n}{2^n} a^{-2n} \sum_{j=0}^{n} (-1)^j \binom{n}{j} a^{-j} \frac{1}{(2n+j-1)!} \sum_{\lambda=n-2\left[\frac{n}{2}\right]}^{m_8}{}^{(2)} A_{\lambda j} (|x| - \lambda a)^{2n+j-1} \quad (10)$$

$$m_8 = \text{min.} \left(n, \left[\frac{|x|}{a}\right]\right)$$

Although this is not mentioned by Lagrange, according to his own dictionary the above transformations hold only when a is a positive integer: actually this limitation is unnecessary and results from the primitive methods of integration utilized by that author.

It will be noticed that the inversion illustrated above involves an interesting device which Lagrange himself did not justify. It was assumed that the integral representing $\psi_n(t)$ had an infinite upper limit whereas in fact the distribution law involved has a finite upper and lower bound. The justification is that since we are not concerned with values of x outside the limits $\pm na$ the integral may be completed arbitrarily.

Inversion in the discrete case

At the commencement of Ch. IV of Book II of his text book on probability theory Laplace (1812) provides a new treatment of De Moivre's problem of the addition of n random variables each rectangularly distributed over a finite number of equally spaced discrete points. He supposes that the individual probability laws consist of $2h+1$ points (h integral) of equal probability, the range extending from $-h$ to h and thus centering on zero. Replacing the t of De Moivre's generating function by e^{iu} the probability generating function of the distribution law for the sum of n discrete rectangular variates is

$$\psi_n(t) = \psi_n(e^{iu}) = \left[\sum_{x=-h}^{h} \frac{1}{2h+1} e^{ixu}\right]^n = \left(\frac{1}{2h+1}\right)^n \left[\sum_{j=-h}^{h} e^{iju}\right]^n$$

Laplace observes that $p_n(x)$, the coefficient of e^{ixu} in the expansion of the expression last written, is the probability of a total value x; it is obtained as the coefficient independent of x in

$$\left(\frac{1}{2h+1}\right)^n e^{-ixu}\left[\sum_{j=-h}^{h} e^{iju}\right]^n$$

Now since

$$\int_0^{\pi} e^{i(j-x)u}\,du = \begin{cases} 0 & j \neq x \\ \pi & j = x \end{cases}$$

this probability may be written

$$p_n(x) = \frac{1}{\pi}\left(\frac{1}{2h+1}\right)^n \int_0^{\pi} e^{-ixu}\left[\sum_{j=-h}^{h} e^{iju}\right]^n du$$

$$= \frac{1}{\pi}\left(\frac{1}{2h+1}\right)^n \int_0^{\pi} e^{-ixu}\left[\frac{e^{\left(h+\frac{1}{2}\right)iu} - e^{-\left(h+\frac{1}{2}\right)iu}}{e^{\frac{iu}{2}} - e^{-\frac{iu}{2}}}\right]^n du$$

$$= \frac{1}{\pi}\left(\frac{1}{2h+1}\right)^n \int_0^{\pi} e^{-ixu}\left[\frac{\sin\left(h+\frac{1}{2}\right)u}{\sin\frac{u}{2}}\right]^n du$$

$$= \frac{1}{\pi}\left(\frac{1}{2h+1}\right)^n \int_0^{\pi} \cos xu\left[\frac{\sin\left(h+\frac{1}{2}\right)u}{\sin\frac{u}{2}}\right]^n du \tag{11}$$

The reasoning used by Laplace in this example is quite general and he has thus derived an inversion formula for the generating function of a discrete probability law, viz.

if

$$\varphi(u) = \sum_x e^{iux}\,p(x) \tag{12}$$

then

$$p(x) = \frac{1}{\pi}\int_0^{\pi} e^{-ixu}\,\varphi(u)\,du \tag{13}$$

Although the preceding development was new in Laplace's 1812 text book this was not the first time he had used a somewhat similar device. In a paper (1810) published two years before his book appeared Laplace used much the same approach based on a continuous rectangular distribution; owing to the somewhat dubious limiting processes involved the derivation employed would not be acceptable nowadays. The argument was extended to cover the addition of n variates from any continuous probability law and an improved version of this development appears in Ch. IV (pp. 329—333) of his text book [1].

The interesting fact emerges from these references that in no case did Laplace use a probability generating function to derive an explicit form of probability law for the sum of n specified random variables. The result (11), for instance, was not intended to be integrated to arrive at the exact answer (given by (2) when $2h+1$ has replaced k and $x+nh$ written instead of x) but was deliberately left in the form of an integral because Laplace had previously (1785) obtained asymptotic forms for such an integral with increasing n. When Laplace required an explicit form for the probability law of the sum of n specified random variables he used an inductive method which he established in an earlier paper (1781) [2]. Comparing his method and the generating function method of Lagrange's (1773) article Laplace wrote: «sa méthode est très-ingénieuse et digne de son illustre auteur; mais la précédente a ... l'avantage d'être plus directe et plus générale, en ce qu'elle réduit la solution du Problème aux quadratures des courbes, quelle que soit la loi de facilité des erreurs des observations.» These remarks were made before Laplace had developed the artifice resulting in (12) and (13) and it is perhaps significant that apart from four articles on the application of probability theory to natural philosophy, astronomy, and geodesy (three of which were reproduced as Supplements in the third, and final, edition of Laplace's book, 1820) he made no further theoretical advances in this subject after the publication of his *Théorie analytique.*

---

[1] In the course of the demonstration Laplace in effect discovered the «moment generating» property of generating functions with $t = e^{iu}$.

[2] This method is reproduced in the Appendix, *post.*

It should be mentioned that the expression «fonctions génératrices» originated with Laplace in 1782 to denote such functions as

$$\psi(t) = \sum_x t^x f(x) \qquad \text{and} \qquad \mu(v) = \int x^v f(x)\, dx$$

These generating functions were used with great effect in the solution of difference and differential equations but Laplace never used the term in connexion with the synthesis of a probability distribution.

Inversion in the continuous case

The first quarter of the nineteenth century saw a number of contributions to the theory of functions which were of considerable importance in the establishment of an inversion formula for the generating function of a continuous probability law. It was possibly Fourier's prize paper of 1811 (not published until 1819/20) which led Gauss to the establishment of the pair of reciprocal relations

$$F(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itu} f(t)\, dt$$

and

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ixu} F(u)\, du$$

Unfortunately these formulae lay undiscovered until the end of the century in one of Gauss's notebooks which he completed in 1813. Their entry without comment under the title «Schönes Theorem der Wahrscheinlichkeitsrechnung» is indeed provocative.

However, substantially the same relations, namely the sine and cosine pairs of transforms, were published by Cauchy in 1817 a year after Fourier's integral theorem, viz.

$$f(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} f(u)\, du \int_{0}^{\infty} \cos v(u-x)\, dv \quad {}^{1)}$$

had first appeared in print.

---

1) The inner integral is actually divergent and nowadays the theorem is written as

$$\frac{1}{2}\{f(x+0) + f(x-0)\} = \frac{1}{\pi} \int_{0}^{\infty} dv \int_{-\infty}^{\infty} f(u) \cos v(u-x)\, du$$

With the stage thus set it was not long before Poisson (1824) derived similar formulae for continuous probability laws. Writing

$$\varphi(u) = \int_a^b e^{iux} p(x)\, dx \tag{14}$$

where a or b may be infinite, he obtained, by what would now be called non-rigorous methods, the result

$$\int_{c-x}^{c+x} p(z)\, dz = \frac{1}{\pi} \int_{-\infty}^{\infty} \varphi(u)\, e^{-icu} \frac{\sin xu}{u} du \tag{15}$$

This relation may be obtained by formally integrating, between $c \pm x$, $(\sqrt{2\pi})^{-1}$ times the second of Gauss's two reciprocal relations given above: it is thus formally equivalent to the following relation which was not, however, written down explicitly by Poisson.

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixu} \varphi(u)\, du \tag{16}$$

Subsequent history

It is an extraordinary fact that although the theory of probability generating functions had achieved a considerable development by the end of the first quarter of the nineteenth century it was almost a hundred years before a synthesis of these results was made and further contributions to the subject published. The cause of this seems to have originated with Laplace's (perhaps personal) dislike of this artifice and the great weight of his authority with ninetenth century mathematicians.

In the analysis of pp. 10 and 11*ante* — which is a pattern of the only uses he made of e^{iu} — Laplace is never far from his own invention (see Appendix), the discontinuity factor. In fact Bessel (1838) re-wrote Laplace's general (Central Limit) theorem in a form which, when applied to the particular case we considered, would run as follows:

$$p_n(x) = \sum_{x_1, x_2, \ldots, x_n}' \left(\frac{1}{2h+1}\right)^n$$ where $\sum'$ represents a summation over the variables $x_1, x_2, \ldots x_n$ subject to $\sum_{j=1}^{n} x_j = x$

$$= \sum_{x_1=-h}^{h} \sum_{x_2=-h}^{h} \cdots \sum_{x_n=-h}^{h} \left(\frac{1}{2h+1}\right)^n \frac{\sin \pi (x_1 + x_2 + \ldots + x_n - x)}{\pi (x_1 + x_2 + \ldots + x_n - x)}$$

$$= \frac{1}{\pi} \int_0^{\pi} \prod_{j=1}^{n} \left\{ \sum_{x_j=-h}^{h} \frac{e^{iux_j}}{2h+1} \right\} e^{-iux} du, \quad \text{etc.}$$

A similar procedure was followed by Ellis (1849) and the close link between the discontinuity factor and generating function approaches was emphasized by Cauchy's four 1853 articles [1]).

Developments of the Central Limit theorem in articles and text-books written between Laplace's discovery of it and the first world war were almost the only occasions when techniques at all resembling probability generating functions were utilized. Some of these writers followed Laplace's introduction of $p_n(x)$ closely, others preferred the Bessel approach. Without giving exact references we mention Poisson, Galloway, De Morgan, Jullien, Laurent and Charlier as favouring the generating function approach, and Glaisher, Tchebychef, Sleshinski, Pizzetti, Liapounoff and Markoff, the discontinuity factor

---

[1]) Cauchy's contributions to the development of the technique of probability generating functions and even to the discovery of the properties of the probability law which bears his name, have often been exaggerated. In the first and second of the four papers cited Poisson's relation (15) is derived and used to find the probability distribution of a linear function of n equal variables each distributed as

$$\text{(I)} \quad \sqrt{\frac{k}{\pi}}\, e^{-kx^2} \qquad\qquad \text{(II)} \quad \frac{k}{2}\, e^{-k|x|} \qquad (n = 2 \text{ only})$$

It is further shown that the probability generating function (with $t = e^{-iu}$) of the law $\frac{k}{\pi} \frac{1}{1 + k^2 x^2}$ is $e^{-\frac{|u|}{k}}$ but no deductions are drawn about the sum of n such variables. Cauchy is thus less discerning than Poisson (1824) was before him for that author had derived the probability distribution of the sum of n variables each distributed according to $\frac{1}{\pi} \frac{1}{1 + x^2}$ and had pointed out that however great n may be the probability of the mean of n such variables lying between given limits is the same as that of an individual variable. Bienaymé (1853) made a similar observation. The last two of Cauchy's four papers are devoted to improving Laplace's «proof» of the Central Limit theorem for n equal components.

introduction. Strictly speaking the probability generating function did not reappear as an independent entity in probability theory until Poincaré (1896) devoted just over two pages of his text book to «characteristic functions», namely probability generating functions with $t = e^z$; and he added nothing new except the title.

The publication of Gauss's posthumous note mentioned earlier received notice by Hausdorff (1901) who made explicit use of probability generating functions with $t = e^{iu}$ without, however, giving them a title. The only novelty is the derivation of the distribution law of the linear function

$$\sum_{j=1}^{n} \left\{ \left(j - \frac{1}{2} \right) \pi \right\}^{-1} x_j$$

of the n variables $x_1, x_2, \ldots x_n$ each distributed according to the law

$$p(x) = \frac{1}{2} e^{-|x|} \qquad -\infty < x < \infty$$

The result is

$$p_n(x) = \left(e^{\frac{\pi x}{2}} + e^{-\frac{\pi x}{2}} \right)^{-1} \qquad -\infty < x < \infty$$

The first author to attempt the development of an independent theory of probability generating functions was Kameda (1916, 1925). To him is due the disinterment of the title ‚generating function' and the theorem which assimilates discrete and continuous laws for the purpose of inversion. Kameda was closely followed in time by von Mises (1919) who used Stieltjes integrals in the representation of the probability distribution, Soper (1922) who threw some of the results of the English statistical school into generating function form without, however, using an inversion formula, and Lévy (1925) whose work on the asymptotic behaviour of various types of probability law has become classic. None of these successive authors was aware of the work of the others; all their papers are readily accessible and need not be mentioned further here.

Appendix

Laplace's method of determining the probability distribution of the sum of n random variables

As mentioned in the body of this article Laplace (1781) devised a direct method of determining the probability distribution of the sum of n random variables each subject to the same probability distribution. Essentially his method is to write

$$p_n(x) = \int_{-\infty}^{\infty} p_{n-1}(x - z)\, p(z)\, dz$$

in the continuous case and he later (1810) extended this to include discrete distributions by means of the relation

$$p_n(x) = \sum_{j=0}^{\infty} p_{n-1}(x - j)\, p(j)$$

These inductive formulae are now one of the standard methods of deriving probability distributions of the sums of random variables (Kendall, 1945, Ch. 10) and of themselves, perhaps, of little interest. However, in the application of these relations to probability distributions of limited range Laplace made brilliant use of a discontinuity factor nearly fifty years before the supposed introduction of such factors by Libri in 1827 (Burckhardt, 1915).

The simplest example of Laplace's procedure is to be found in his 1810 article. The individual continuous random variables are assumed to be rectangularly distributed over the interval $(0, h)$ and a discontinuity factor ζ is introduced by writing

$$p(x) = \begin{cases} \dfrac{1}{h} & 0 \leqslant x \leqslant h \\[2ex] \dfrac{1 - \zeta^h}{h} & h < x < \infty \end{cases}$$

so that the range of the variable has become infinite; the final stage of the procedure is to write $\zeta = 1$.

Now

$$p_n(x) = \int_0^x p_{n-1}(x - z_1)\, p(z_1)\, dz_1$$

$$= \int_0^x p(z_1)\, dz_1 \int_0^{x-z_1} p_{n-2}(x - z_1 - z_2)\, p(z_2)\, dz_2$$

$$= \int_0^x p(z_1)\, dz_1 \int_0^{x-z_1} p(z_2)\, dz_2 \ldots \int_0^{x-z_1-z_2-\ldots-z_{n-2}} p(x - z_1 - z_2 - \ldots - z_{n-1})\, p(z_{n-1})\, dz_{n-1}$$

$$= \frac{1}{h^n}(1 - \zeta^h)^n \int_0^x dz_1 \int_0^{x-z_1} dz_2 \ldots \int_0^{x-z_1-z_2-\ldots-z_{n-2}} dz_{n-1} \qquad \text{(a)}$$

$$= \frac{1}{h^n}(1 - \zeta^h)^n \int_0^x dz_1 \int_0^{x-z_1} dz_2 \ldots \int_0^{x-z_1-z_2-\ldots-z_{n-3}} (x - z_1 - z_2 - \ldots - z_{n-2})\, dz_{n-2}$$

$$= \frac{1}{h^n}(1 - \zeta^h)^n \int_0^x dz_1 \int_0^{x-z_1} dz_2 \ldots \int_0^{x-z_1-z_2-\ldots-z_{n-4}} 1/2!\,(x - z_1 - z_2 - \ldots - z_{n-3})^2\, dz_{n-3}$$

$$= \frac{1}{h^n}(1 - \zeta^h)^n \frac{1}{(n-1)!} x^{n-1}$$

$$= \frac{1}{h^n} \frac{1}{(n-1)!} \sum_{j=0}^{n} (-1)^j \binom{n}{j} \zeta^{hj} x^{n-1} \qquad \text{(b)}$$

$$= \frac{1}{h^n} \frac{1}{(n-1)!} \sum_{j=0}^{\min.(n,[x])} (-1)^j \binom{n}{j} (x - hj)^{n-1} \qquad 0 \leqslant x \leqslant nh$$

Two steps in this derivation need explanation. At (a) it has been assumed that each of the n variables is making a contribution to x which exceeds h: that is to say all n variables are now measured from h instead of from zero. In fact either none, one, two, ... n of these variables falls in the $(0, h)$ portion of the range and the respective frequencies of these possibilities are provided by the numerical coefficients of the expansion of $(1 - \zeta^h)^n$. The term in ζ^{hj}, for instance, denotes that j of the variables have been given a variate value h in excess of the truth; ζ^{hj} thus acts on x in (b) to reduce it by hj.

The discrete analogue of the preceding development appears on pp. 253—256 of the *Théorie analytique*.

References

Bessel, F. W. (1838) Untersuchungen über die Wahrscheinlichkeit der Beobachtungsfehler. *Astron. Nachr.* xv, 369.

Bienaymé, J. (1853) Considérations à l'appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moindres carrés. *Comptes Rendus,* xxxvii, 309.

Burckhardt, H (1915) Trigonometrische Reihen und Integrale (bis etwa 1850). *Encyk. Math. Wissens.* Bd. II, I. Teil, 2. Hälfte, 825.

Cauchy, A. L. (1817) Sur une loi de reciprocité qui existe entre certaines fonctions. *Bull. philomat.* Paris, 121.

— (1853) Mémoire sur les coefficients limitateurs ou restricteurs.
Sur les résultats moyens d'observations de même nature, et sur les résultats les plus probables.
Sur la probabilité des erreurs qui affectent des résultats moyens d'observations de même nature.
Mémoire sur les résultats moyens d'un très grand nombre d'observations. *Comptes Rendus,* xxxvii, 150, 198, 264, 381.

De Moivre, A. (1711) De mensura sortis, seu, de probabilitate eventuum in ludis a casu fortuito pendentibus. *Philos. Trans.* xxvii, 213.

— (1730) *Miscellanea analytica de seriebus et quadraturis.* London.

Euler, L. (1788) Éclaircissemens sur le mémoire de Mr. de La Grange inseré dans le V^e volume des Mélanges de Turin, concernant la méthode de prendre le milieu entre les résultats de plusieurs observations etc. *Nova acta acad. sci. Petropol.* iii (1785) 289.

Fourier, J. (1816) Théorie de la chaleur. *Ann. chim. phys.* iii, 350.

Gauss, C. F. (1900) *Werke,* viii, 88.

Hausdorff, F. (1901) Beiträge zur Wahrscheinlichkeitsrechnung. *Ber. Verhand. königｇ. sächs. Gesell. Wissens. Leipzig, Math.-Phys. Cl.,* liii, 152.

Kameda, T. (1916) Theorie der erzeugenden Funktionen und ihre Anwendung auf die Wahrscheinlichkeitsrechnung. *Proc. Tôkyô. Math.-Phys. Soc.,* Ser 2, viii, 262, 336.

— (1925) Theory of generating functions and its application to the theory of probability. *Jour. Fac. Sci. Univ. Tokyo,* Sec. I, i, 1.

Kendall, M. G. (1945) *The advanced theory of statistics.* London.

Lagrange (1773) Mémoire sur l'utilité de la méthode de prendre le milieu entre les résultats de plusieurs observations; dans le quel on examine les avantages de cette méthode par le calcul des probabilités; et où l'on résoud différens problèmes relatifs à cette matière. *Miscellanea Taurinensia,* v, 167.

Laplace, P. S. de (1812) *Théorie analytique des probabilités.* Paris.

— (1781) Mémoire sur les probabilités. *Hist. Acad. Roy. Sci.* (1778), 227.

— (1810) Mémoire sur les approximations des formules qui sont fonctions de très-grands nombres, et sur leur application aux probabilités. *Mém. Sci. Math. Phys. Inst. France* (1809), 353.

— (1785) Mémoire sur les approximations des formules qui sont fonctions de très-grands nombres. *Hist. Acad. Roy. Sci.* (1782), 169.

Lévy, P. (1925) *Calcul des probabilités.* Paris.

Mises, R. von (1919) Die Fundamentalsätze der Wahrscheinlichkeitsrechnung. *Math. Zeitschr.* iv, 1.

Poincaré, H. (1896) *Calcul des probabilités.* Paris.

Poisson, S. D. (1824) Sur la probabilité des résultats moyens des observations. *Connaissance des Temps* (1827), 273.

Simpson, T. (1757) An attempt to show the advantage arising by taking the mean of a number of observations, in practical astronomy. *Miscellaneous Tracts* ... London.

Soper, H. E. (1922) *Frequency arrays.* Cambridge.

Todhunter, I. (1865) *A history of the mathematical theory of probability from the time of Pascal to that of Laplace.* Cambridge.