STIELTJES INTEGRALS IN MATHEMATICAL STATISTICS

By

J. SHOHAT

(Jacques Chokhate)

Introduction. Stieltjes integrals, introduced into analysis in 1894-5[1], play an increasingly important role not only in pure mathematics, but also in theoretical physics and in the theory of probability. In mathematical statistics, however, their use, it seems, still remains very limited. And yet, one of the most remarkable features of Stieltjes integrals is that they represent, as the case may be, an integral proper or a sum of an finite or an infinite number of *discrete* aggregates. Thus *the statistician is enabled to treat in a single formula a continuous, as well as a discontinuous distribution.* This means far more than a mere simplification of writing. In fact, since Stieltjes integrals have many properties in common with Riemann and Lebesgue definite integrals, we can use all known resources of the theory of definite integrals (mean-value theorem, various inequalities), and therefore readily obtain general results which, otherwise, require special (often complicated) proofs. The advantage of such a treatment is particularly evident in the theory of interpolation, approximation, and mechanical quadratures.

Hence, the object of this paper is to present a general exposition of the properties and applications of Stieltjes integrals. Many of the results stated below are well known[2], and the proofs may be omitted. Some results are believed to be new (for example, extension of Tchebycheff and Hölder inequalities) and may prove useful in mathematical statistics. We close, as an illustration, with the theory of interpolation, for here, even in recently published books, the continuous and discontinuous cases are treated *separately* while the underlying ideas are *identical.*

---

1. Stieltjes: (a) Recherches sur les fractions continues, Oeuvres, v. II, p. 402-559; (b) Correspondence d'Hermite et de Stieltjes, v. II, p. 272, where these integrals are first mentioned in a letter (No. 351) to Hermite under date of October 25, 1892.

2. (a) Hobson, The Theory of Functions of a Real Variable, 2d. ed. (1921), v. I, p. 506-16, 605-09; (b) O. Perron, Die Lehre von den Kettenbrüchen (1913), p. 362-69.

I. *Definition and general properties.* Let $f(x)$ be continuous and $\psi(x)$ be bounded monotonic non-decreasing on the finite interval (a, b) $(a < b)$ Then, as is well known, the following limits exist:

$$\left.\begin{aligned}\psi(x+0) &= \lim_{\epsilon \to 0} \left[\psi(x+\epsilon) - \psi(x) \right] \\ \psi(x-0) &- \lim_{\epsilon \to 0} \left[\psi(x-\epsilon) - \psi(x) \right]\end{aligned}\right\} \quad (a \leq x \leq b)$$

If x is a point of discontinuity of $\psi(x)$, $\psi(x+0) - \psi(x-0)$ (>0) is called "saltus" of $\psi(x)$ at this point. The number of such points is at most denumerably infinite; the points of continuity of $\psi(x)$ are, therefore, everywhere dense in (a, b) $\psi(x)$ is R- integrable, and so is $\psi(x)x^k$ $(k = 0, 1 \cdots\cdots)$. The Riemann-Stieltjes integral (of $f(x)$ with respect to $\psi(x)$ $\int_a^b f(x)d\psi(x)$) is defined as follows:

(S)

$$\int_a^b f(x)d\psi(x) = \lim_{\substack{n \to \infty \\ \max(x_{i+1} - x_i) \to 0}} \sum_{i=0}^{n-1} f(\xi_i)\left[\psi(x_{i+1}) - \psi(x_i)\right]$$

$$a = x_0 < x_1 < x_2 < \cdots\cdots < x_{n-1} < x_n = b$$

$$x_i \leq \xi_i \leq x_{i+1} \quad (i = 0, 1, \cdots\cdots n-1)$$

The existence of the right-hand limit can be easily established. The continuity of $f(x)$ is here sufficient, but not necessary[1].

In many phases of mathematical statistics the case of a continuous $f(x)$ is evidently the more important, although many problems arising in the theory of probability require applications of the discontinuous case.

From the very definition (S) one may obtain many properties of Stieltjes integrals in common with the ordinary definite integrals. Thus:

$$(1) \qquad \int_a^b d\psi(x) = \psi(b) - \psi(a)$$

1. (a) Hobson, 1-c; (b) T. Hildebrandt, On Integrals Related to and Extension of the Lebesque Integrals, Bulletin of the American Mathematical Society (2), V. 24 (1918), p. 177-202; (c) Lebesque, Leçon sur l'intégration, 2d ed. (1928), p. 252-313.

(2) $\int_a^c f\,d\psi + \int_c^b f\,d\psi = \int_a^b f\,d\psi$ $\quad (a < c < b)$

(3) $\int_a^b (f_1 \pm f_2)\,d\psi = \int_a^b f_1\,d\psi \pm \int_a^b f_2\,d\psi$

(4) $\int_a^b A f\,d\psi = A\int_a^b f\,d\psi$ $\quad (A = \text{Const.})$

(5) $\left|\int_a^b f\,d\psi\right| \leq \int_a^b |f|\,d\psi$

(6) $\int_a^b f\,d\psi = f(\xi)\int_a^b d\psi$ $\quad (a \leq \xi \leq b$: mean-value theorèm)

(7) $\int_a^b f_1\,d\psi \leq \int_a^b f_2\,d\psi$, if $f_1(x) \leq f_2(x)$ for $a \leq x \leq b$

(8) $\int_a^b \sum_{i=1}^{\infty} f_i\,d\psi = \sum_{i=1}^{\infty} \int_a^b f_i\,d\psi$

if $\sum_{i=1}^{\infty} f_i(x)$ converges uniformly in (a, b)

(9) $\int_a^b f\,d\psi = f\psi\Big|_a^b - \int_a^b \psi\,df$ $\quad$ (integration by parts)

(10) $\int_a^b f\,d\psi = \int_a^b f(x)p(x)dx$, if $\psi(x) = \int \phi(x)dx + c$

with $p(x) \geq o$ in (a, b).

(10-bis) $\int_a^b f\,d\psi = \int_a^b f(x)\psi'(x)\,dx$ $\quad$ if $\psi'(x)$ exists

and is R- integrable in (a, b).

Let $\psi(x)$ have only a *finite* number of points of increase in (a, b).

$$(a = x_0 <)x_1 < x_2 < \cdots\cdots < x_n (< x_{n+1} = b)$$

with the saltus σ_i at $x = x_i$ $(i = 1, 2 \cdots\cdots n)$, so that $\psi(x)$ remains constant $= \sum_{s=1}^{i} \sigma_s$ for $x_i \leq x < x_{i+1}$ and $\psi(b) \sum_{i=1}^{n} \sigma_i$. Such functions, called *stepwise* functions ("fonction en escalier"), prove very useful. Here

$$(11)\quad \int_a^b f\,d\psi = \sum_{i=1}^{n} \sigma_i\, f(x_i) \qquad \{\sigma_i = f(x_i+0) - f(x_i-0)\}$$

If the number of points of increase is infinite

$$(a<)\, x_1 < x_2 < \cdots < x_n < \cdots, \quad \lim_{n\to\infty} x_n = b$$

$$(12)\quad \int_a^b f\,d\psi = \sum_{i=1}^{\infty} \sigma_i\, f(x_i)\,.$$

Conversely, any sum $\sum_{i=1}^{n} u_i\, v_i$ can be represented as a Stieltjes integral in infinitely many ways. Let us introduce n positive numbers $\sigma_1, \sigma_2, \cdots \sigma_n$ a certain interval (a, b), n points $(a<)\, x_1 < \cdots < x_n (< b)$ (the choice of x_i σ_i depends upon the nature of the problem involved), and a stepwise function $\psi(x)$ having at $x = x_i$ a saltus σ_i $(i = 1, 2, 3 \cdots n)$. Then, writing $u_i = \sigma_i\, w_i$, we may consider v_i, w_i as values taken respectively by some functions $f(x), \phi(x)$ at $x = x_i$ $(i = 1, 2, \cdots\cdots n)$. Hence,

$$(13)\quad \sum_{i=1}^{n} u_i\, v_i = \int_a^b f(x)\phi(x)\,d\psi(x)$$

Formulae (11-13) show clearly the use of Stieltjes integrals for the representation of sums of discrete aggregates.

$$(14)\quad \int_a^b f\,d\psi \geq 0, \quad \text{if } f(x) \geq 0 \text{ in } (a, b)$$

Here "=" takes place if and only if $\psi(x)$ has a finite or denumerably infinite number of points of increase in (a, b) (not everywhere dense) and $f(x)$ vanishes at all these points, for we exclude, of course, functions $f(x)$ which vanish at all points of continuity of $\psi(x)$ and therefore vanish identically in (a, b). *If $\psi(x)$ has infinitely many points of increase, while $f(x)$ vanishes in (a, b) only a finite number of times, without changing sign, then $\int_a^b f(x)\,d\psi(x) \neq 0$ and has the sign of $f(x)$.*

$$(15)\quad \int_a^b f(x)\, x^k\, d\psi(x) = 0 \qquad (k = 0, 1, \cdots n-1) \text{ implies: } f(x)$$

has at least n distinct roots inside (a, b) assuming that $\psi(x)$ has at least n points of increase[1].

1. This is a form of a theorem due to Perron (1-c, p. 368-69). If the number of such points is $m < n$, (15) shows only that $f(x)$ vanishes at all such points.

(16) $\int_a^b x^k d\psi(x) = 0 \quad (k = 0, 1, \cdots\cdots)$ implies:

$\psi(x)$ constant for $(a \leq x \leq b)$[1].

Since in the defiition (S) only the differences $\psi(x_{i+1}) - \psi(x_i)$ enter, it follows that a Stieltjes integral does not change its value if we replace $\psi(x)$ by $\psi(x)+c$. More precisely:

(17) $\int_a^b f\, d\psi_1 = \int_a^b f\, d\psi_2$

if the two monotonic non-decreasing functions $\psi_{1,2}(x)$ differ by an additive constant only at all points of continuity. Applying the mean-value theorem to $\int_x^{x+h} f(x) d\psi(x)$, we conclude:

(18) $F(x) = \int_a^x f(t) d\psi(t)$ is continuous at all points of continuity of

$\psi(x)$ and therefore, almost everywhere in (a, b).

(19) $\lim\limits_{h \to 0} \dfrac{F(x+h) - F(x)}{\psi(x+h) - \psi(x)} = f(x) \qquad (a \leq x \leq b)$

(20) $F'(x) = f(x)\, \psi'(x)$ at any point x, in (a, b), where $\psi'(x)$ exists.

One recognizes in (18-20) a generalization of the properties of the ordinary definite integral which is a special case, for $\psi(x) \equiv x$.

(21) $\phi\, t = \int_a^b f(x, t) d\psi(x)$ is continuous in $t (t_0 \leq t \leq t_1)$

if $f(x, t)$, continuous in x, is uniformly continuous with respect to $t (t_0 \leq t \leq t_1)$ for all values of x in $(a\ b)$. Moreover,

(22) $\dfrac{d\phi(t)}{dt} = \int_a^b \dfrac{\partial f(x, t)}{\partial t} d\psi(x)$

if $\dfrac{\partial f(x, t)}{\partial t}$ exists and is continuous in x and uniformly continuous in $t (a \leq x \leq b;\ t_0 \leq t \leq t_1)$.

---

1. If $\psi(x)$ has a finite number, n, of points of increase, then n such relations imply the same conclusion.

Notes. (i) The above results hold, with proper limitations and modifications, if $\psi(x)$ be of *bounded variation* in (a,b), for such a function can be represented as a difference of two monotonic non-decreasing functions $\psi_{1,2}(x)$ and we define in accordance with (S),

$$\int_a^b f\,d\psi = \int_a^b f\,d\psi_1 - \int_a^b f\,d\psi_2 \;.$$

(ii) In applications to probability and mathematical statistics $\psi(x)$ stands for the "cumulative law of distribution," so that

(23) $\psi(x)$ is monotonic non-decreasing from $\psi(a)=0$ to $\psi(b)=1$.

(24) For $(a \leq c < d \leq b)$ the integral $\int_c^d d\,\psi(x)$ $=$ probability $P:[c \leq x \leq d]$; $\int_a^b d\,\psi(x)=1$.

(25) $\int_a^b f(x)d\,\psi(x) = E(f)$, i. e., the expected value or mathematical expectation of $f(x)$.

Let $w(x)$ $f(x)$ be continuous in (a , b), and $\alpha(x)$ be of bounded variation. Then,

(26) $\psi(x)=\int_a^x w(x)d\,\alpha(x)$ is of bounded variation,[1]

$$\int_a^b f(x)d\,\psi(x) = \int_a^b f(x)w(x)d\alpha(x)$$

Given an infinite sequence of functions $\psi_n(x)$ $(n=1,\ 2,\ \cdots)$ of bounded variation in (a,b). If the total variation in (a, b) of all $\psi_n(x)$ does not exceed a fixed quantity M independent of n, and if, in addition, $\lim\limits_{n\to\infty} \psi_n(x) = \psi(x)$ exists for $a \leq x \leq b$, then[2]

(27) $\lim\limits_{n\to\infty} \int_a^b f(x)d\,\psi_n(x) = \int_a^b f(x)d\,\psi(x)$, for any continuous $f(x)$.

Notes. (i) (27) holds true if we know that $\lim\limits_{n\to\infty} \psi_n(x) = \psi(x)$ exists at all points of continuity of the sequence $\psi_n(x)$ and at $x=a\,,\,b$.

1. T. Carleman Leçons sur les équations intégrales singulières noyau réel et symétrique (Uppsala) (1923), p. 11-12.
2. Page 9 of preceding reference.

(ii) In applications to probability and statistics (27) is of great importance. In fact, consider $\psi_n(x)$ *as a sequence of variable laws of distribution approaching, as a limit, a certain fixed law of distribution* $\psi(x)$. Then(by (23), the total variation of any $\psi_n(x)$ in (a, b) is 1; (27) thus becomes applicable and shows that *under the said conditions the expected value of any continuous function in the variable law of distribution approached, as* $n \to \infty$ *its expected value in the limiting law of distribution.*

II. *Stieltjes Integrals Over an Infinite Interval.* We define

$$(28) \quad \int_a^\infty f\,d\psi = \lim_{x\to\infty}\int_a^x f\,d\psi \;; \qquad \int_{-\infty}^\infty f\,d\psi = \lim_{\substack{a\to-\infty\\ b\to\infty}} \int_a^b f\,d\psi$$

(similarly $\int_{-\infty}^b$), provided the right-hand limits exist as finite numbers. It is assumed that $\int_a^x f\,d\psi$, $\int_a^b f\,d\psi$ exist respectively for any finite $x > a$, and for any finite interval (a, b). For the existence of (28) it is necessary and sufficient that

$$(29) \quad \left|\int_{x_1}^{x_2} f\,d\psi\right| < \epsilon \quad \text{for } x \geq \text{ a certain number } x(\epsilon),$$

$\epsilon > 0$ — arbitrarily small.

One sees readily that

(30) $\int_{-\infty}^\infty f\,d\psi$ exists, if $\int_{-\infty}^\infty d\psi$ does, and if $f(x)$ is bounded for all real values of x. The first of these conditions is satisfied if $\psi(x)$ is a law of distribution. We notice that any $\int_a^b f\,d\psi$ can be written as $\int_{-\infty}^\infty f\,d\psi$, if we agree to take $\psi(x) = \psi(a)$, $\psi(b)$ respectively for $x \leq a$, $\geq b$.

The formulae given above hold, in general, for infinite limits as well, *with the exception of those which require a double limiting process*, like 8, 21, 27, etc., where ordinarily additional precautions must be taken in the form of certain assumptions specifying the behaviour of $\psi(x)$ and of other functions involved at infinity. Thus, (8) is not valid in general for $(a, b) = (-\infty, \infty)$, and requires a more detailed discussion. Furmulae 21, 22 hold true if we assume, for example, the uniform boundedness and continuity with respect to t of the functions involved for all x in $(-\infty, \infty)$, and also the existence of $\int_{-\infty}^\infty d\psi(x)$, i. e. definite values for $\psi(\pm\infty)$.

Formula 17 deserves *special attention*: *in general, it is not true* *for an infinite interval*, as was shown by Stieltjes[1].

III. *Approximate Evaluation of Stieltjes Integrals.* In practice, as in statistical computations, we evaluate $\int_a^b f\,d\psi$ approximately, replacing it by the right-hand member of (S), for a certain chosen n. The question arises regarding the error r_n of such an approximation. Let $\omega(a)$ represent the modulus of continuity of $f(x)$, i e.

$$(31)\qquad |f(x)-f(y)| \leq \omega(\delta) \text{ for } |x-y| \leq \delta \,(a \leq x,\ y \leq b)$$

Then, if $x_{i+1} - x_i < h$ in (S) for $i = 0, 1, \cdots\cdots n-1$, we have

$$r_n = \int_a^b f\,d\psi - \sum_{i=0}^{n-1} f(\xi_i)\left[\psi(x_{i+1}) - \psi(x_i)\right]$$
$$= \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} \left[f(x) - f(\xi_i)\right] d\psi(x)$$

$$(32)\qquad |r_n| \leq \omega(h) \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} d\psi = \omega(h)\left[\psi(b) - \psi(a)\right]$$
$$\{h = max.\,(x_{i+1} - x_i)\ ;\ i = 0, 1, \cdots\cdots\cdots n-1\}$$

(32) *answers the above question for any continuous* $f(x)$.

Special Case: Lipschitz condition[2]:

$$(33)\qquad |f(x)-f(y)| \leq \lambda\,|x-y| \qquad (a \leq x,\ y \leq b\ ;$$
$$|r_n| \leq \lambda h\left[\psi(b)-\psi(a)\right] \qquad \lambda = const.)$$

In (32, 33) *we replace* h *by* $h/2$, *if* (ξ_i) *in* (S) *is, as usual, the mid-point of the interval* (x_i, x_{i+1}) $(i = 0, 1, \cdots\cdots n-1)$.

It must be noticed, however, that the above considerations are

---

1. (1. c. p. 73, p. 505-06. (17) is closely related to the so-called "Moments-problem": find a monotonic non-decreasing function $\psi(x)$ in (a, b) with infinitely many points of increase, if all its moments $\gamma_k = \int_a^b x^k d\psi(x)$ $[k = 0, 1 \cdots]$ are given. *This problem, for* (a, b) *infinite, may be "indetermined," i. e. it may admit infinitely many solutions, while it is always "determined," for finite* (a, b). Stieltjes gives the following example:
$$\int_0^\infty x^k[1+\lambda \sin(x^{1/4})]e^{-x^{1/4}}\,dx = \int_0^\infty x^k e^{-x^{1/4}}\,dx$$
$[\lambda$ ∴ constant, $k = 0, 1, \cdots\cdots]$ and $\psi(x) = 0 = \int_0^x [1+\lambda \sin x^{1/4}]e^{x^{1/4}}$ is monotonic non-decreasing in $0, \infty$, if $|\lambda| \leq 1$.

2. If $f'(x)$ exists for $a \leq x \leq b$, then $\lambda^{-\infty}$ can be taken equal to max. $|f'(x)|$ in (a, b). If $f(x)$ is given graphically, λ can be found roughly as the maximum of the absolute va[illegible] ...ne slope in (a, b).

not workable in general on an infinite interval, for here, in place of (31), we ordinarily have the more complicated relation

$$|f(x)-f(y)| \leq \omega(x,y,\delta) \qquad (|x-y| \leq \delta)$$

where $\omega(x,y,\delta) \to \infty$ with x, y (ex.: $f(x)=x^2$). Thus here, in order to obtain an inequality for the error, we must add to the right member of (32), where a, b are finite numbers properly chosen, two more terms—the upper limits of $\left|\int_{-\infty}^{a} f\,d\psi\right|$ and $\left|\int_{b}^{\infty} f\,d\psi\right|$, which we obtain by means of a suitable hypothesis concerning the behavior of $f(x)$, $\psi(x)$ at infinity.

IV. *Tchebycheff and Hölder Inequalities for Stieltjes Integrals*[1]. Hereafter $\psi(x)$ stands for a monotonic non-decreasing function defined on a certain interval (a, b), finite or infinite. Let $f_i(x)$, $\phi_i(x)\,[i=1,2,\cdots n]$ be continuous on (a,b)[2]. Then we have the following *fundamental transformation*:

$$(34)\qquad \begin{vmatrix} \int_a^b f_1\phi_1 d\psi & \int_a^b f_1\phi_2 d\psi & \cdots & \int_a^b f_1\phi_n d\psi \\ \int_a^b f_2\phi_1 d\psi & \cdots & \cdots & \int_a^b f_2\phi_n d\psi \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \int_a^b f_n\phi_1 d\psi & \cdots & \cdots & \int_a^b f_n\phi_n d\psi \end{vmatrix}$$

$$= \frac{1}{n!}\underbrace{\int_a^b \cdots \int_a^b}_{(n \text{ times})} \begin{vmatrix} f_1(x_1) \cdots f_1(x_n) \\ \vdots \\ f_n(x_1) \cdots f_n(x_n) \end{vmatrix} \cdot \begin{vmatrix} \phi_1(x_1) \cdots \phi_1(x_n) \\ \vdots \\ \phi_n(x_1) \cdots \phi_n(x_n) \end{vmatrix} \prod_{i=1}^{n} d\psi(x_i) .$$

The proof is very simple for $n=2$, for we can write

$$\int_a^b u(x)d\psi(x) \cdot \int_a^b v(x)d\psi(x)$$

$$\text{as} \int_a^b \int_a^b u(x_1)v(x_2)\, d\psi(x_1)\, d\psi(x_2)$$

and it may readily be extended to any n. Formula (34) yields many

1. Cf. my Note: Jacques Chokhate, Sur les intégrales de Stieltjes, Comptes Rendus, v. 189 (1929), p. 618-20.

2. In case $\psi(x)$ has a finite number of points of increase in (a, b), we require only definite values of all $f_i(x)$, $\phi_i(x)$ at these points.

interesting results by a proper choice of $n,\ f_i\ ,\ \phi_i$.

Examples: (i) $n=2\ ;\ f_1\equiv\phi_1\ ,\ f_2\equiv\phi_2$

$$(35)\qquad \int_a^b f_1^2\,d\psi\int_a^b f_2^2\,d\psi-\Big(\int_a^b f_1 f_2\,d\psi\Big)^2=$$

$$\tfrac{1}{2}\int_a^b\!\!\int_a^b\begin{vmatrix} f_1(x_1) & f_1(x_2)\\ f_2(x_1) & f_2(x_2)\end{vmatrix}^2 d\psi(x_1)\,d\psi(x_2)\ \geq\ 0 .$$

Schwartz inequality—("=" only if f_1 and f_2 are linearly dependent.

(ii) $n=2\ ;\ f_2\equiv\phi_2\equiv 1$. Write f,ϕ in place of $f_1,\ \phi_1$:

$$(36)\qquad \int_a^b f\phi\,d\psi\cdot\int_a^b d\psi-\int_a^b f\,d\psi\cdot\int_a^b\phi\,d\psi$$

$$=\tfrac{1}{2}\int_a^b\!\!\int_a^b\big[f(x)-f(y)\big]\big[\phi(x)-\phi(y)\big]\,d\psi(x)\,d\psi(y)$$

$$\int_a^b d\psi\cdot\int_a^b f\phi\,d\psi \gtrless \int_a^b f\,d\psi\cdot\int_a^b\phi\,d\psi$$

Tchebycheff inequality (derived by him for the special case $d\psi=dx$), *where* f , ϕ *are any two functions both varying monotonically in* $(a\,,b)$, *either in the same sense* (*sign* $>$ *in* (36) *or in the opposite sense* (*sign* $<$). *In* (34-37) *we may replace* $d\psi(x)$ *by* $p(x)dx\big[p(x)\geq 0\ \ in(a,\ b)\big]$.[1]

(iii) $f_i(x)=x^{i-1},\ \ \phi_i(x)=F(x)\,x^{i-1}\big[i=1,2,\cdots\cdots n\big]$:

$$\Delta_n=\begin{vmatrix}\int_a^b F\,d\psi & \int_a^b Fx\,d\psi & \cdots\cdots & \int_a^b Fx^{n-1}d\psi\\ \int_a^b Fx\,d\psi & \cdots\cdots & \cdots & \int_a^b Fx^{n}\,d\psi\\ \vdots & & & \vdots\\ \int_a^b Fx^{n-1}d\psi & \cdots\cdots & \cdots & \int_a^b Fx^{2n-2}d\psi\end{vmatrix}$$

$$=\tfrac{1}{n!}\underbrace{\int_a^b\cdots\cdots\int_a^b}_{(n\ \text{times})}\prod_{i=1}^{n}F(x_i)\,d\psi(x_i)\prod_{\substack{i,j=1\\ i\neq j}}^{n}(x_i-x_j)^2 .$$

1. Cf. E. Fischer, Ueber den Hadamardschen Determinontensatz, Archiv für Mathematik and Physik (3), v. 13 (1908), p. 32-49, where (34) is derived for the particular case $d\psi(x)\equiv dx$

The determinant Δ_n plays an important role in the theory of orthogonal Tchebycheff polynomials (see below). Formula (37) gives an upper limit for Δ_n:

$$(38)\qquad |\Delta_n| < \frac{1}{n!}(b-a)^{n(n-1)} M^n \Big[\int_a^b d\psi(x)\Big]^n$$
$$\big[M = \max.\ |F(x)| \text{ in } (a,\ b)\big].$$

Applying (13) to the above formulae, we get:

$$(39)\qquad \Big(\sum_{i=1}^n u_i v_i\Big)^2 \leq \sum_{i=1}^n u_i^2 \sum_{i=1}^n v_i^2$$

—*Cauchy inequality* (from (34))

$$(40)\qquad n\sum_{i=1}^n a_i b_i \gtrless \sum_{i=1}^n a_i \sum_{i=1}^n b_i$$

$$(41)\qquad \frac{\sum_{i=1}^n u_i v_i}{\sum_{i=1}^n v_i} \gtrless \frac{\sum_{i=1}^n u_i w_i}{\sum_{i=1}^n w_i} \quad (v_i\,,\ w_i\ > 0)^1$$

Formulae 40, 41 follow from (36) by means of (13), with $\sigma_i = 1$ (in (40)), $\sigma_i = w_i$ (in (41)) $[\ i = 1, 2, \cdots n\ ]$. The sequences $\{(a_i),\ (b_i)\}$, $\{(u_i),\ (\frac{v_i}{w_i})\}$ are assumed to be either increasing or decreasing, the sign $\gtrless$ being chosen as in (36). Thus *all these (and many similar) inequalities have the same origin-formula* (34). Applying (13) to Hölder-Minkowski inequalities[2].

$$(42)\qquad \sum_{i=1}^n |a_i\ b_i| \leq \Big\{\sum_{i=1}^n |a_i|^s\Big\}^{1/s} \cdot \Big\{\sum_{i=1}^n |b_i|^{s/(s-1)}\Big\}^{(s-1)/s} \qquad (s>1)$$

$$\Big\{\sum_{i=1}^n |a_i + b_i|^s\Big\}^{1/s} \leq \Big\{\sum_{i=1}^n |a_i|^s\Big\}^{1/s} + \Big\{\sum_{i=1}^n |b_i|^s\Big\}^{1/s} \qquad (s>1)$$

we get:

$$(43)\qquad \int_a^b |f\phi|\,d\psi \leq \Big\{\int_a^b |f|^s d\psi\Big\}^{1/s} \cdot \Big\{\int_a^b |\phi|^{\frac{s}{s-1}}\ d\psi\Big\}^{(s-1)/s} \quad (s>1)$$

---

1. Cf. l. c. p. 73, l-b, pp. 142, 143, 146, 194.

2. F. Rietz, Ueber Systeme integrierbarer Funktionen, Mathematische Annalen, v. 69 (1911), pp. 449-497; p. 456.

(44) $$\left\{\int_a^b |f+\phi|^s d\psi\right\}^{1/s} \le \left\{\int_a^b |f|^s d\psi\right\}^{1/s} + \left\{\int_a^b |\phi|^s d\psi\right\}^{1/s} (s \ge 1).$$

Formula (43), with $\phi \equiv 1$, $s = {}^{s_2}/_{s_1} > 1$ and f replaced by $|f|^{s_1}$, yields:

(45) $$\left\{\int_a^b |f|^{s_1} d\psi\right\}^{1/s_1} \le \left\{\int_a^b |f|^{s_2} d\psi\right\}^{1/s_2} \cdot \left\{\int_a^b d\psi\right\}^{\frac{s_2-s_1}{s_1 s_2}}$$
$$(s_2 > s_1 > 0).$$

The applications of the above inequalities to the theory of probability and mathematical statistics are many. A few illustrations follow:

(i) Consider even moments $u_{2s} = \int_{-a}^{a} x^{2s} f(x)dx = 2\int_0^a x^{2s} f(x)dx$ of a continuous unimodal symmet distribution over a finite interval $(-a, a)$. Here (36) gives (with $s_1 <$, $d\psi(x) = f(x)dx$, $2\int_0^a f(x)dx = 1$).

(46) $$u_{2s} = \int_{-a}^{a} x^{2s} f(x)dx < \frac{a^{2s}}{2s+1}$$
$$\left[s = 1, 2, \cdots\cdots; \; f(x) \equiv f(-x)\right].$$

(ii) If ξ denotes an arbitrary constant, take in (42) $f(x) = x - \xi$, $\psi(x) =$ *law of distribution of* x *over* (a, b), so that $\int_a^b d\psi(x) = 1$. We get:

(47) $$v_{s_1}^{1/s_1} \le v_{s_2}^{1/s_2} \quad \text{for } s_1 < s_2$$
$$\left(v_s^{1/s} \equiv \left[\int_a^b |x-\xi|^s \, d\psi\right]^{1/s}\right)$$

Hence, in any distribution over any interval the quantity $v_s = \left[\int_a^b |x-\xi|^s \, d\psi(x)\right]^{1/s}$ *increases with* s *for any constant* ξ and, in particular,

$$u_{2s}^{1/2s} = \left[\int_a^b x^{2s} \, d\psi\right]^{1/2s} \qquad \textit{also if } a \ge 0,$$
$$u_s'^{1/s} = \left[\int_a^b f \, d\psi\right]^{1/s}.\ [1]$$

1. Paul Levy, Calcul des probabilités (Paris 1925), p. 157-58.

(iii) Apply (36) to the functions $f(x)$, $\phi(x)$ *both monotonic in* (a, b), $\psi(x)$ the same as in (ii):

(48) $E(f\phi) \gtrless E(f)E(\phi)$ (for the choice of $\gtrless$ see (36))[1].

The same formula (36) gives for *any function* $f(x)$[2]

(49) $E(f^n) > \{E(f)\}^n \quad (n = 2, 3, \cdots)$[2]

Formula (45) gives with the same $\psi(x)$:

(50) $\{E(|f|^{s_1})\}^{1/s_1} \leq \{E(|f|^{s_2})\}^{1/s_2} \quad (s_1 < s_2)$.

V. *Application of Stieltjes Integrals to Some Minimum-Problems. Given a number* $m \geq 1$, *M finite points* $x_1 \leq x_2 \leq \cdots x_M$, *M positive quantities* $\sigma_1, \sigma_2, \cdots \sigma_M$, *and a function* $f(x)$ *with well determined values* $f(x_i)(i = 1, 2, \cdots M)$. *Find a polynomial* $P_n(x)$, *of degree not exceeding* $n (\leq M-2)$, *minimizing the expression* $\sum_1^M \sigma_i |f(x_i) - P_n(x_i)|^m$. *Discuss the behavior of* $P_n(x)$ *for* $m \to \infty$. We introduce a finite interval (a, b), containing in its interior all points x_i and a monotonic non-decreasing step-wise function $\psi(x)$ with the above properties (saltus σ_i at $x = x_i$, etc.; see p. 75). Then our problem can be formulated as follows: *Find a polynomial* $P_n(x)$ *of degree not exceeding* n, *minimizing the integral*

$$\int_a^b |f(x) - P_n(x)|^m \, d\psi(x) \quad [m \geq 1].$$

Here the advantage of Stieltjes integrals is clearly evident, for the latter problem has been discussed by G. Polya[3], D. Jackson[3] and the writer[3]. We know that a solution always exists and is unique for $m > 1$. The behavior of $P_n(x)$, when either or both m and n in

1. G. Bohlman. Formulierung und Begründung Zweier Hulfssätze der Mathematischen Statistik. Mathematische Annalen, v. 74 (1913), pp. 341-442; p. 374-75.

2. In fact, (36) holds, with sign $>$, if $f(x) - f(y)$ and $\phi(x) - \phi(y)$ have the same sign for any x, y in a, b, which, of course, is true for $\phi(x) \equiv f(x)$.

3. (a) G. Polya, Sur un algorithme toujours convergent . . ., Comptes Rendus, v. 157 (1913) p. 840-43. (b) D. Jackson, On the Convergence of certain polynomial and trigonometric approximations. Transactions of the American Mathematical Society, v. 22 (1921), p. 158-66. (c) Idem, Note on the Convergence of Weighted Trigonometric Series, Bulletin of the American Mathematical Society, v. 29 (1923), p. 259-63. (d) J. Shohat, On the Polynomial and Trigonometric Approximation, Mathematische Annalen, v. 103 (1929), p. 157-75.

crease indefinitely, has also been discussed by the above writers. It was found that, *if* $f(x)$ *be continuous in* (a, b), *then for* n *fixed and* $m \to \infty$, $P_n(x)$ *approaches uniformly in* (a, b) *the polynomial* $\Pi_n(x)$, *of degree* $\leq n$, *of the best approximation (in Tchebycheff sense*[1]) to $f(x)$ *,provided,* $\psi(x)$ *has infinitely many points of increase everywhere dense in* (a, b). *Furthermore,* $\left[\int_a^b |f(x)-P_n(x)|^m d\psi(x)\right]^{1/m}_{m\to\infty}$ *the best approximation* $E_n(f) = \max. |f(x) - \Pi_n(x)|$ *for* $a \leq x \leq b$.

This result has been supplemented by the writer (in a paper which will appear elsewhere), who showed that *the above result holds if* $\psi(x)$ *has a finite number* $M (\geq n+2)$ *points of increase.* $\Pi_n(x)$ *representing here the polynomial (of degree* $\leq n$) *giving the best approximation to* $f(x)$ *on the aggregate of the said points of increase of* $\psi(x)$. The following cases are of special interest.

(a) $n = 0$, *i. e. find a constant* X_m *minimizing the sum* $\sum_{i=1}^{M} \sigma_i |f(x_i) - X_m|^m$.

Very simple considerations show that the best approximation to $\{f(x_i)\}$ $(i = 1, 2, \cdots\cdots)$ by means of a constant is $E_0(f) = \frac{1}{2}|f(x_l) - f(x_k)|$, $f(x_l)$, $f(x_k)$ *being respectively the largest and the smallest of the* $f(x_i)$. *so that* $|f(x_l) - f(x_k)|$ *is the largest possible, and the "constant of the best approximation" is* $N_0 = \frac{1}{2}\left[f(x_k) + f(x_l)\right]$. Thus here

$$
(51) \qquad
\begin{aligned}
&\lim_{m\to\infty} X_m = \frac{f(x_k) + f(x_l)}{2} \\
&\lim_{m\to\infty} \left\{ \sum_{i=1}^{m} \sigma_i \left[f(x_i) - X_m\right]^m \right\}^{1/m} \\
&= \left|\frac{f(x_k) - f(x_l)}{2}\right| = \max. \left|\frac{f(x_i) - f(x_j)}{2}\right| (i, j = 1, 2, \cdots n)
\end{aligned}
$$

$$
(52) \qquad
\begin{aligned}
&f(x_1) < f(x_2) < \cdots\cdots < f(x_M) \text{ implies :} \\
&\lim_{m\to\infty} X_m = \frac{f(x_M) + f(x_1)}{2} \\
&\lim_{m\to\infty} \left\{ \sum_{i=1}^{m} \sigma_i |f(x_i) - X_m|^m \right\}^{1/m} = \frac{f(x_M) - f(x_1)}{2}
\end{aligned}
$$

1. That is: $E_n(f) \equiv \max. |f(x) - \Pi_n(x)| \leq \max. |f(x) - G_n(x)|$ $(a \leq x \leq b)$ where $G_n(x)$ is an arbitrary polynomial of degree $\leq n$, equality implying necessarily: $G_n \equiv \Pi_n$.

and the limiting results do not depend on $x_2, x_3, \cdots x_{n-1}$ As an illustration $f(x)=x^{2k+1}$ may serve, or, more generally, $f(x)=\sum_i A_i x^{2k_i+1}$ (all $A_i > 0$; all k_i and K are positive integers or zero)[1].

(b) $M=n+2$, *n arbitrary*. Here the writer showed (the paper will appear elsewhere):

$$(53)\ \lim_{m\to\infty} P_n(x)=\Pi_n(x)$$

$$=\frac{f_{n+1}+f_{n+2}}{2}-\frac{1}{2K}\sum_{i,j=1}^{n}(-1)^{i+j}K_{i,j}\frac{f_i-f_{i+2}}{x_i-x_{i+2}}\left(x^j_{n+1}+x^j_{n+2}-2x^j\right)$$

$$(54)\ \lim_{m\to\infty}\left[\sum_{i=1}^{n}\sigma_i\left|f_i-P_n(x_i)\right|^m\right]^{1/m}=E_n(f)$$

$$=\frac{1}{2}\left|f_k-f_{k+1}-\frac{1}{K}\sum_{i,j=1}^{n}(-1)^{i+j}\frac{f_i-f_{i+2}}{x_i-x_{i+2}}\left(x^j_k-x^j_{k+1}\right)\right|$$

$$k=1,\ 2,\cdots\cdots,M$$

$$f_i\equiv f(x_i)\quad (i=1,\ 2,\cdots M)$$

where $K, K_{i,j}$ stand respectively for the following determinant and its minors:

$$(55)\qquad K=\begin{vmatrix} \dfrac{x_1^2-x_3^2}{x_1-x_3} & \dfrac{x_1^3-x_3^3}{x_1-x_3} & \cdots & \dfrac{x_1^n-x_3^n}{x_1-x_3} \\ \dfrac{x_2^2-x_4^2}{x_2-x_4} & \cdots & \cdots & \dfrac{x_2^n-x_4^n}{x_2-x_4} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \dfrac{x_n^2-x_{n+2}^2}{x_n-x_{n+2}} & \cdots & \cdots & \dfrac{x_n^n-x_{n+2}^n}{x_n-x_{n+2}} \end{vmatrix}$$

We proceed now to show the application of Stieltjes integrals to interpolation. This must be preceded by a discussion of

VI. *Orthogonal Tchebycheff Polynomials. Theorem. Any*

---

1. Cf. D. Jackson, Note on the Median of a Set of Numbers, Bulletin of the American Mathematical Society, v. 22 (1920), p. 160-64, where the above results have been obtained for the particular case $f(x)=x$.

function $\psi(x)$, *monotonic non-decreasing on* (a, b) *—finite or infinite, and having all moments* $\gamma_k = \int_a^b x^k d\psi(x)$ $(k=0, 1, \cdots)$ *with* $\gamma_0 > 0$ *generates a sequence of polynomials* $\{\phi_n(x)\}$ *of degree* $n = 0, 1, \cdots$ *uniquely determined by the relations*[1]: $\int_a^b \phi_m \phi_n d\psi = 0$. $(m \neq n\ ;\ m, n = 0, 1, \cdots)$ *equivalent to* $\int_a^b x^k \phi(x) d\psi(x) = 0$ $(k = 0, 1, \cdots n-1)$ $(n = 1, 2, \cdots\cdots)$.

Proof. Take $\phi_n(x) = x^n + f_{n-1} x^{n-1} \cdots f_1 x + f_0$. The above relations lead to the following set of equations:

$$
(56)\qquad
\begin{aligned}
&f_0 \gamma_0 + f_1 \gamma_1 + \cdots\cdots\cdots f_{n-1} \gamma_{n-1} + \gamma_n = 0\\
&f_0 \gamma_1 + f_1 \gamma_2 + \cdots\cdots\cdots f_{n-1} \gamma_n + \gamma_{n+1} = 0\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&f_0 \gamma_{n-1} + f_1 \gamma_n + \cdots\cdots\cdots f_{n-1} \gamma_{2n-2} + \gamma_{2n-1} = 0
\end{aligned}
$$

The determinant Δ_n of the coefficients γ_i is (see (37)):

$$
(57)\qquad \Delta_n = \frac{1}{n!} \underbrace{\int_a^b \cdots \int_a^b}_{(n\ \text{times})} \prod_{i=1}^{n} d\psi(x_i) \prod_{i,j=1,\ i \neq j}^{n} (x_i - x_j)^2 > 0
$$

which proves our statement. Add to (56) the identical relation $f_0 + f_1 x + \cdots\cdots\cdots + f_{n-1} x^{n-1} + (x^n - \phi_n) = 0$, and for $\phi_n(x)$ we obtain the following expression:

$$
(58)\qquad \phi_n(x) = \begin{vmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_n \\ \gamma_1 & \gamma_2 & \cdots & \gamma_{n+1} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{n-1} & \gamma_n & \cdots & \gamma_{2n-1} \\ 1 & x & \cdots & x^n \end{vmatrix} \Bigg/ \begin{vmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_2 & \cdots & \gamma_n \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{n-1} & \gamma_n & \cdots & \gamma_{2n-2} \end{vmatrix}
$$

Note. If $\psi(x)$ has in (a, b) only a finite number M of points of increase, then $\Delta_n \equiv 0$ *for* $n > M$, and $\phi_n(x)$ exists only for $n = 0, 1, \cdots M$ (See below (65), which in this case is a rational fraction).

The following table gives the most known and important Tchebycheff polynomials.

1. We disregard constant factors.

(a, b)	$d\psi(x) =$	$\phi(x)$ — polynomial of [Constant factors disregarded] [1]
finite	dx	Legendre: $\frac{d^n}{dx^n}\left[(x-a)^n(b-x)^n\right]$
finite	$(x-a)^{\alpha-1}(b-x)^{\beta-1}dx$ $(\alpha,\beta>0)$	Jacobi: $(x-a)^{1-\alpha}(b-x)^{1-\beta}\frac{d^n}{dx^n}\left[(x-a)^{n+\alpha-1}(b-x)^{n+\beta-1}\right]$
$(0, \infty)$	$x^{\alpha-1}e^{-kx}dx$ $(\alpha, K>0)$	Laguerre: $x^{1-\alpha}e^{kx}\frac{d^n}{dx^n}\left[x^{n+\alpha-1}e^{-kx}\right]$
$(-\infty, \infty)$	$e^{-kx^2}dx\ (K>0)$	Laplace-Hermite: $e^{kx^2}\frac{d^n}{dx^n}(e^{-kx^2})$

The polynomials $\phi_n(x)$ can be *normalized,* by multiplying by constant factors $a_n = 1 : \int_a^b \phi_n^2(x)d\psi$, so as to obtain *an orthogonal and normal* system of Tchebycheff polynomials $\{\phi_n(x) = a_n x^n + \cdots\}$ $(n = 0, 1, \cdots\cdots;\ a_n > 0)$

$$(59)\qquad \int_a^b \phi_m(x)\phi_n(x)d\psi = 0\ (m \neq n),\ = 1\ (m = n)$$

$$(m, n = 0, 1, 2 \cdots\cdots\cdots)$$

The following are some of the most important properties of $\phi_n(x)$.

(a) The roots of $\phi_n(x)$ are real, distinct and lie between a, b.

(b) If all integrals $\int^b x^n f(x)d\psi(x)$ exist $(n = 0, 1, \cdots)$, then, by (59), we have the formal development:

$$(60)\qquad f(x) \sim \sum_{i=0}^{\infty} A_i\phi_i(x), \qquad \left[A_i = \int_a^b f(x)\phi_i(x)d\psi(x)\right]^2$$

which, regardless of its convergence or divergence, has the following remarkable property: *any "section" ("Abschnitt") of* (60), *i. e. the polynomial* $P_n(x) = \sum_{i=0}^{n} A_i\phi_i(x)$, *obtained by taking its first* $n+1$ *terms* $(n = 0, 1, \cdots)$ *gives the best approximation to* $f(x)$ *in* (a, b), *in the sense of least squares, i. e. it minimizes the integral* $\int_a^b [f(x) - P_n(x)]^2 d\psi(x)$. *Moreover*

1. Cf. W. Romanowsky, Sur quelques classes nouvelles des polynomes orthogonaux, Comptes Rendus, v. 188 (1929), p. 1023-25, where new polynomials are discussed arising from Pearson's frequency curves of type IV, V, VI.

2. In the development $f(x) \sim \sum_{i=0}^{\infty} A_i\phi_i(x)$, where the $\phi_i(x)$ are not normalized, $A_i = \int_a^b f\phi_i d\psi : \int_a^b \phi_i^2 d\psi$.

$$(61)\quad \int_a^b [f-P_n]^2 d\psi = \min. \int_a^b [f(x)-G_n(x)]^2 d\psi(x) = \int_a^b f^2 d\psi - \sum_{i=0}^{n} A_i^2 ,$$

$G_n(x)=\sum_{i=0}^{n} g_i x^i$ denoting hereafter an *arbitrary* polynomial of degree $\leq n$. The proof is very simple. Write $G_n(x)$ as $\sum_{i=0}^{n} H_i \phi_i(x)$ with constant coefficient H_i, substitute this expression into $I=\int_a^b [f-G_n]^2 d\psi$, and write down the conditions of minima: $\frac{1}{2}\frac{\partial I}{\partial H_i} = 0$, which, by (59), lead to

$$H_i = \int_a^b f \phi_i \, d\psi = A_i \qquad (i = 0, 1, 2, \cdots n).$$

These coefficients A_i can be written down as linear combinations of the moments

$$(62)\qquad m_k = \int_a^b f(x) x^k d\psi(x) \quad (k=0, 1, \cdots\cdots).$$

Introduce the symbol

$$(63)\qquad \omega(G_n) = \sum_{i=0}^{n} m_i g_i$$

$$\left(G_n(x)=\sum_{i=0}^{n} g_i x^n;\ n=0, 1, \cdots ;\ g_i \text{ arbitrary}\right)$$

Then evidently,

$$(64)\qquad \begin{aligned} A_n &= \int_a^b f \phi_n \, d\psi = \omega(\phi_n) \\ f(x) &\sim \sum_{n=0}^{\infty} \omega(\phi_n)\, \phi_n(x) ; \end{aligned}$$

in other words, we have the following simple rule: *In the expression of $\phi_n(x)$ replace each power x^k by the corresponding moment m_k given in (62)($K= 0, 1, \cdots n$), and we obtain the coefficient A_n in (60)* $(n = 0, 1, \cdots\cdots)$.

(c) $\phi_n(x)$ are denominators of the successive convergents to the continued fraction

$$(65)\qquad \int_a^b \frac{d\psi(y)}{x-y} = \frac{\lambda_1 |}{|x-c_1} - \frac{\lambda_2 |}{|x-c_2} - \cdots$$

$$(\lambda_i\,(>0),\quad c_i \text{ - const.}) .$$

Historically, it was the aforesaid minimum property which has lead Tchebycheff to the discovery and investigation of the *general class or orthogonal polynomials corresponding to any monotonic non-decreasing function,* while before, only isolated special cases of such polynomials have been known (polynomials of Legendre, Jacobi, Laguerre, Laplace, Hermite). Tchebycheff found these polynomials in connection with

VII. *Least-squares Interpolation.* The problem can be formulated with Tchebycheff[1] as follows: *Given the values of a certain function $y=F(x)$ at $n+1$ real, distinct points $x_1, x_2, \cdots x_{n+1}$, with the corresponding weights σ_i [2]. Find its value at $x=X$, assuming for y the representation $a+bx+cx^2+\cdots+hx^m$, $(m\leq n)$ so that the errors of $F(x_i)$ $[i=1, 2, \cdots n+1]$ shall have the least possible influence on the required value $F(x)$.*

Using Stieltjes integrals (which greatly simplifies Tchebycheff's analysis), we are lead to the following solution:

$$(66)\qquad \begin{gathered} F(X)=P_m(X)=\sum_{k=0}^{m} A_k\,\phi_k(X) \\ \left[A_k=\int_a^b F(x)\,\phi_k(x)\,d\psi(x)=\sum_{i=1}^{n+1}\sigma_i\,F(x_i)\,\phi_k(x_i)\right], \end{gathered}$$

where $\psi(x)$ *is the stepwise function having at $x=x_i$ a saltus $\sigma_i (i=1, 2, \cdots n+1)$, (a, b) contains in its interior all points x_i, $\{\phi_n(x)\}$ are orthogonal and normal polynomials determined by* (59), *or, which is the same, denominators of the successive convergents to the continued fraction* (65) (we disregard constant factors), *which here reduces to*

$$(67)\qquad \sum_{i=1}^{n+1}\frac{\sigma_i}{x-x_i} = \frac{\lambda_1 |}{|x-c_1} - \frac{\lambda_2 |}{|x-c_2} - \cdots\cdots$$

1. Tchebycheff, (a) Sur les fractions continues, Journal des Mathematiques, (2), v. III (1858), p. 289-323; (b) On the least-squares interpolation, Collected Papers, v. I, p. 473-98; (c) On interpolation with equidistant ordinates, ibid., v. II, p. 219-42 (b, c, in Russian).

2. σ_i is inversely proportional to the mean-square error of $F(x_i)$.

We see that (66) *is nothing but the first* $m+1$ *terms of the development* (60). Hence, *Tchebecheff's solution* (66) *yields the minimum of* $\int_a^b [F(x)-P_n(x)]^2\, d\psi(x) = \sum_{i=1}^{n} \sigma_i [F(x_i)-P_n(x)]^2$. Moreover, for the mean-square error of (66), we get, by (59):

$$
\begin{aligned}
R^2 &= \int_a^b F^2 d\psi - \sum_{k=0}^{m} A_k^2 \\
&= \sum_{i+1}^{n+1} \sigma_i F^2(x_i) - \sum_{k=0}^{m} \left\{ \sum_{i=1}^{n+1} \sigma_i F(x_i)\,\phi_k(x_i) \right\}^2 .
\end{aligned} \tag{68}
$$

The name "least-squares interpolation" is thus fully justified, and we see the *complete identity between the two problems*: *least-squares interpolation and approximate representation of functions by series of Tchebycheff polynomials. Whether the data are discrete and in a finite number, or the form a continuous set, the underlying principles and the resulting formulae are identical, provided we use Stieltjes integrals.* There is no need to treat the two cases separately (as one finds even in recent books on this subject) and to introduce special symbols in the first case. Another very important feature of the above solution has been indicated by Tchebycheff: If we add one more term to the expression $a+bx+\cdots+hx^m$ assumed for $y=F(x)$, we need only add one more term to $P_m(x)$ above, *without changing the preceding ones* (compare with Lagrange interpolation formula!) Formula (68) enables one to find the number of terms necessary to attain a prescribed accuracy.

Consider two special cases.

(a) *The ordinates are equidistant* $: x_{i+1}-x_i=h \; (i=1,2,\cdots n)$ *and all weights* σ_i *are equal* $(=1)$. Here Tchebycheff (1-c. 1-b. p. 91) gives very simple expressions for the polynomials $\phi_k(x)$, as well as for the coefficients A_k of (66):

$$
\begin{aligned}
\phi(z) &= \Delta^k\Big[\left(z+\tfrac{n-1}{2}\right)\left(z+\tfrac{n-3}{2}\right)\cdots\left(z+\tfrac{n-2k+1}{2}\right)\left(z-\tfrac{n+1}{2}\right)\left(z-\tfrac{n+3}{2}\right) \\
&\quad \cdots\left(z-\tfrac{n+2k-1}{2}\right)\Big] \quad k=0,1,2,\cdots; \; z=\frac{x-\frac{1}{2}(x_1+x_n)}{x_2-x_1}; \\
&\quad \Delta^k - k^{th} \text{ difference.}
\end{aligned} \tag{69}
$$

$$
\begin{aligned}
u(z) \equiv F(x) &= \frac{1}{n}\sum_{i=1}^{n} u_i \phi_0(z) + \frac{3}{n(n^2-1^2)}\sum_{i=1}^{n} \frac{i}{1}\cdot\frac{n-i}{1}\Delta u_i\, \phi_1(z) \\
&+ \frac{5}{n(n^2-1^2)(n^2-2^2)}\sum_{i=1}^{n} \frac{i(i+1)}{1\cdot 2}\cdot\frac{(n-i)(n-i-1)}{1\cdot 2}\Delta^2 u_i\, \phi_2(z) + \cdots \\
&\left[(u_i \equiv F(x_i) \right]
\end{aligned} \tag{70}
$$

(We have replaced $n+1$ in our above formulae by n). All $\phi_n(z)$ can be easily computed by means of the relations:

$$\phi_0(z)=\Delta^0 1=1 \quad ; \quad \phi_1(z)=2z\,,$$

$$(71) \qquad \phi_k(z)=2(2k-1)\phi_{k-1}(z)-(k-1)^2\left[n^2-(k-1)^2\right]\phi_{k-2}(z)$$

$$(k \geq 2)$$

(b) $m=1$, x_i *arbitrary* $(i=1,\ 2,\cdots n)$. We take in (67)

$$(72) \qquad \lambda_1=\int_a^b d\psi(x)=\sum_{i=1}^{n}\sigma_i\,.$$

We get now (by successive division, for ex.)

$$(73) \qquad c_1=\frac{\int_a^b x\,d\psi}{\int_a^b d\psi}=\frac{\gamma_1}{\gamma_0}$$

$$\left(\gamma_k=\int_a^b x^k d\psi(x)=\sum_{i=1}^{n}\sigma_i\,x_i^k\right)$$

$$\phi_0(x)=\frac{1}{\sqrt{\gamma_0}}$$

$$(74) \qquad \phi_1(x)=\frac{x_1-c_1}{\sqrt{\int_a^b (x-c_1)^2\,d\psi}}=\frac{\gamma_0\,x-\gamma_1}{\sqrt{\gamma_0^2\,\gamma_2-\gamma_1^2}}$$

$$(75) \qquad P_1(x)=A_0\,\phi_0(x)+A_1\,\phi_1(x)$$

$$=\sum_{i=1}^{n}\frac{\sigma_i\,y_i}{\gamma_0}+\sum_{i=1}^{n}\frac{\sigma_i\,y_i\,(\gamma_0\,x_i-\gamma_1)}{\gamma_0^2\,\gamma_2-\gamma_1^2}(\gamma_0\,x-\gamma_1)$$

$$\left[y_i\equiv F(x_i)\right]$$

$$R^2 \text{ (mean-square error) } =\sum_{i=1}^{n}\sigma_i\left[F(x_i)-P_n(x_i)\right]^2$$

$$(76)=\left(\sum_{i=1}^{n}\frac{\sigma_i\,y_i}{\sqrt{\gamma_0}}\right)^2+\left(\sum_{i=1}^{n}\frac{\sigma_i\,y_i\,(\gamma_0\,x_i-\gamma_1)}{\sqrt{\gamma_0^2\,\gamma_2-\gamma_1^2}}\right)^2$$

(See 68)

Let $\psi(x)$ *represent a law of distribution.* Then, $\gamma_0=1$, $\sqrt{\gamma_0^2\,\gamma_2-\gamma_1^2}=$ standard deviation σ and the above formulae become:

$$\phi_0(x)=1\,, \qquad \phi_1(x)=\frac{x}{\sigma} \tag{77}$$

$$P_1(x)=\frac{x\int_a^b F x\,d\psi}{\int_a^b x^2 d\psi} \tag{78}$$

$$=\frac{x\sum_{i=1}^{n}\sigma_i\,x_i\,y_i}{\sum_{i=1}^{n}\sigma_i\,x_i^2} \qquad \left[(y_i)=F(x_i)\right]$$

$$R^2=\int_a^b F^2 d\psi-\left(\int_a^b F\phi_1\,d\psi\right)^2 \tag{79}$$

$$=\sum_{i=1}^{n}\sigma_i\,y_i^2-\frac{\left(\sum_{i=1}^{n}\sigma_i\,x_i\,y_i\right)^2}{\sigma^2}$$

One recognizes in (78, 79) formulae quite similar to those for the line of regression of y on x and for the standard error of estimate of y. Introduce

$$\sigma_x^2=\int_a^b x^2 d\psi\;;\quad \sigma_y^2=\int_a^b y^2 d\psi\,,$$

$$r=\frac{\int_a^b x y\,d\psi}{\sqrt{\int_a^b x\,d\psi\cdot\int_a^b y\,d\psi}} \tag{80}$$

and our formulae become the classical ones:

$$P_1(x)=r\,\frac{\sigma_y}{\sigma_x}\,x\;;\quad R=\sigma_y\,(1-r^2)^{1/2}. \tag{81}$$

We thus obtained, using Stieltjes integrals, elegant, simple and easily memorizable formulae for σ_x, σ_y and for the coefficient of correlation r. Moreover, we see by inspection (Schwartz inequality) that $-1\leq r\leq 1$, equality attainable if and only if x and y are linearly dependent. We see also that *the theory of linear regression is but a very special case of the general theory — due to Tchebecheff — of least-squares interpolation.*

1. Cf. D. Jackson. The Elementary Geometry of Function Space, American Mathematical Monthly, v. 31 (1924), p. 461-71.

Paris (France).