23458

Prof. FELICE VINCI

UN NUOVO METODO D'INTERPOLAZIONE

Estratto dall'Annuario
del Regio Istituto Superiore
di scienze economiche e commerciali in Bari
(1922-23 e 1923-24)

MINISTERO DELLE FINANZE
DONO
1934

BARI
CASA EDITRICE CRESSATI
1924

BIBLIOTECA MINISTERO DELLE FINANZE
MISCELL. 174. h. 178

BIBLIOTECA ★ DEL MINISTERO DELLE FINANZE

UN NUOVO METODO D'INTERPOLAZIONE

Con la presente nota ci proponiamo di esporre un metodo nuovo e semplicissimo d'interpolazione approssimata.

Esso mira ad eliminare la necessità della scelta di un moltiplicatore o del raggruppamento in classe dei dati empirici, che — a prescindere da casi particolari — sono gli aspetti arbitrari, e non sempre scevri d'inconvenienti, dei metodi interpolatori sinora usati.

Esso, inoltre, può spesso sostituirsi con vantaggio ai noti metodi d'interpolazione esatta.

*
* *

1. — Assumendo, anzitutto, che la funzione interpolatrice sia:

$$f(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_r x^r, \qquad (1)$$

ricordiamo che si ha:

$$\begin{aligned}
&\Delta^0 f(x) = f(x)\\
&\Delta^1 f(x) = f(x+h) - f(x)\\
&\Delta^2 f(x) = \Delta^1 f(x+h) - \Delta^1 f(x)\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
&\Delta^r f(x) = \Delta^{r-1} f(x+h) - \Delta^{r-1} f(x)\\
&\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned}$$

dove h è una costante arbitraria, le differenze prima, seconda, ... $(r-1)^{ma}$ sono rispettivamente polinomi di grado $(r-1)^{mo}$, $(r-2)^{mo}$, ... 1^0, la differenza r^{ma} è una costante, e le differenze $(r+1)^{ma}$, $(r+2)^{ma}$,... sono nulle.

Ora, quando le coppie di dati empirici:

$$(x_1, y_1),\ (x_2, y_2),\ \ldots\ (x_n, y_n)$$

diano luogo ad un numero di condizioni tali da determinare una funzione razionale intera di grado più elevato della (1), al fine di interpolare la (1) tra codesti dati empirici sorge naturale l'idea di vincolare in qualche guisa le differenze finite della funzione a quelle degli y_i.

Conviene limitarsi alla considerazione delle differenze *finite*, sia per assicurare al metodo un carattere elementare, sia per non impli-

care l'ipotesi della continuità dei dati empirici e quindi procedimenti ed espedienti più o meno arbitrari.

Ora, ponendo le $r+1$ condizioni:

$$\Sigma\, \Delta^k (f(x_i) - y_i) = 0, \qquad (k = 0, 1, 2, \ldots r)$$

ossia:

$$\Sigma\, (\Delta^k f(x_i) - \Delta^k y_i) = 0,$$

od anche:

$$\left.\begin{aligned} &\Sigma\, f(x_i) = \Sigma\, y_i \\ &\Sigma\, \Delta f(x_i) = \Sigma\, \Delta y_i \\ &\Sigma\, \Delta^2 f(x_i) = \Sigma\, \Delta^2 y_i \\ &\cdots\cdots\cdots \\ &\Sigma\, \Delta^r f(x_i) = \Sigma\, \Delta^r y_i, \end{aligned}\right\} \qquad (2)$$

si ottiene un sistema lineare rispetto agli $r+1$ parametri di $f(x)$, il cui significato è evidente.

Mentre la condizione di uguaglianza della somma degli y_i alla somma dei valori corrispondenti della funzione interpolatrice col metodo delle somme si estende a gruppi parziali dei valori medesimi, e coi metodi dei minimi quadrati e del CAUCHY si estende alle altre somme che si ottengono applicando alle prime dei moltiplicatori; col metodo che proponiamo la condizione di uguaglianza si estende alle somme delle differenze finite non nulle della funzione interpolatrice, venendo in certa guisa a rispettare, per così dire, la fisionomia parabolica dei dati empirici.

Nell'ipotesi — alla quale per ora ci limitiamo — che gli x_i formino una progressione aritmetica, si ha sempre: $\Delta^{k+1} x^k{}_i = 0$ ed il sistema precedente si riduce a:

$$\left.\begin{aligned} a_0 n + a_1 \Sigma\, x_i + a_2 \Sigma\, x^2{}_i + \ldots + a_r \Sigma\, x^r{}_i &= \Sigma\, y_i \\ a_1 \Sigma\, \Delta x_i + a_2 \Sigma\, \Delta x^2{}_i + \ldots + a_r \Sigma\, \Delta x^r{}_i &= \Sigma\, \Delta y_i \\ a_2 \Sigma\, \Delta^2 x^2{}_i + \ldots + a_r \Sigma\, \Delta^2 x^r{}_i &= \Sigma\, \Delta^2 y_i \\ \cdots\cdots\cdots\cdots\cdots & \\ a_r \Sigma\, \Delta^r x^r{}_i &= \Sigma\, \Delta^r y_i, \end{aligned}\right\} \qquad (2\ bis)$$

dove è:

$$\begin{vmatrix} n & \Sigma\, x_i & \Sigma\, x^2{}_i & \ldots & \Sigma\, x^r{}_i \\ 0 & \Sigma\, \Delta x_i & \Sigma\, \Delta x^2{}_i & \ldots & \Sigma\, \Delta x^r{}_i \\ 0 & 0 & \Sigma\, \Delta^2 x^2{}_i & \ldots & \Sigma\, \Delta^2 x^r{}_i \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & & \Sigma\, \Delta^r x^r{}_i \end{vmatrix} = n \Sigma\, \Delta x_i \ \Sigma\, \Delta^2 x^2{}_i \ \ldots \Sigma\, \Delta^r x^r{}_i = 0.$$

Pertanto, ricavate dai dati empirici le sommatorie del sistema, si potrà procedere rapidamente al calcolo numerico dei parametri, a cominciare da a_r.

2. — La semplicità dei procedimenti di calcolo è ancora maggiore di quella che a prima giunta potrebbe sembrare:

a) Anzitutto, essendo gli x_i in progressione aritmetica di ragione h, le somme delle potenze successive di essi possono ricavarsi rispettivamente dai primi tre, quattro, ... termini della nota formula ricorrente:

$$x_{n+1}^{s+1} = x_1^{s+1} + n\,h^{s+1} + (s+1)\,h^s \sum x_i + \\ + \binom{s+1}{2} h^{s-1} \sum x_i^2 + \ldots + (s+1) \sum x_i^r , \qquad (3)$$

attribuendo successivamente ad s i valori: 1, 2, ... r.

D'altra parte, possiamo scrivere la nuova formula:

$$\sum x_i^s = x_1^s + (x_1 + h)^s + (x_1 + 2h)^s + \ldots + (x_1 + (n-1)h)^s =$$

$$= n\,x_1^s + s\left(\sum_{i=1}^{n-1} i\right) x_1^{s-1} h + \binom{s}{2}\left(\sum_{i=1}^{n-1} i^2\right) x_1^{s-2} h^2 + \ldots + \left(\sum_{i=1}^{n-1} i^r\right) h^s, \quad (4)$$

la quale si presta meglio della (3) ai calcoli numerici, anche perchè le formule per il calcolo di $\sum_{i=1}^{n-1} i$, $\sum_{i=1}^{n-1} i^r$, ... sono notissime e possono comunque ricavarsi successivamente e con molta facilità dalla (3).

In particolare si ha:

$$\left.\begin{aligned} \sum x_i &= n\,x_1 + \frac{n(n-1)}{2} h, \\ \sum x_i^2 &= n\,x_1^2 + 2\,\frac{n(n-1)}{2} x_1 h + \frac{n(n-1)(2n-1)}{6} h^2, \\ \sum x_i^3 &= n\,x_1^3 + 3\,\frac{n(n-1)}{2} x_1^2 h + 3\,\frac{n(n-1)(2n-1)}{6} x_1 h^2 + \\ &\quad + \frac{n^2(n-1)^2}{4} h^3, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned}\right\} \quad (5)$$

delle quali il calcolo è rapidissimo, specie semplificando e mettendo in ognuna in evidenza i fattori comuni a gruppi di termini convenientemente scelti.

b) Inoltre, le somme delle differenze prime, seconde, ... r^{me} potranno pure calcolarsi con molta rapidità, ricordando che, data una qualsiasi successione di numeri:

$$\alpha_1, \alpha_2, \dots \alpha_n ,$$

sussiste la formula:

$$\Delta^k \alpha_i = \alpha_{k+i} - k\,\alpha_{k+i-1} + \binom{k}{2}\alpha_{k+i-2} - \dots + (-1)^k \alpha_i ,$$

dove per $h = 1, 2, \dots r$, i primi due, tre, ... $r+1$ termini dànno rispettivamente le differenze prima, seconda, ... r^{ma} di α_i. Infatti, in base ad essa possiamo scrivere la nuova formula:

$$\begin{aligned}\Sigma\Delta^k \alpha_i &= \Delta^{k-1}\alpha_{n-k+1} - \Delta^{k-1}\alpha_1 = \\ &= \alpha_n - (k-1)\alpha_{n-1} + \binom{k-1}{2}\alpha_{n-2} \dots + (-1)^{k-1}\alpha_{n-k+1} - \\ &\quad - \alpha_k + (k-1)\alpha_{k-1} - \binom{k-1}{2}\alpha_{k-2} + \dots - (-1)^{k-1}\alpha_1 = \\ &= (\alpha_n - \alpha_k) - (k-1)(\alpha_{n-1} - \alpha_{k-1}) + \binom{k-1}{2}(\alpha_{n-2} - \alpha_{k-2}) - \\ &\quad - \dots + (-1)^{k-1}(\alpha_{n-k+1} - \alpha_1) ,\end{aligned}$$

od anche:

$$\begin{aligned}\Sigma\Delta^k \alpha_i = (\alpha_n \pm \alpha_1) - (k-1)(\alpha_{n-1} \pm \alpha_2) + \binom{k-1}{2}(\alpha_{n-2}) \pm \alpha_3) - \\ - \dots \mp (\alpha_{n-k+1} \pm \alpha_k) .\end{aligned}$$

Quest'ultima formula — in cui dovrà adottarsi il segno posto in alto o in basso, secondochè k sia rispettivamente pari o dispari — abbrevia molto i calcoli numerici.

Da essa si trae in particolare:

$$\left.\begin{aligned}\Sigma\Delta\,\alpha_i &= \alpha_n - \alpha_1 \\ \Sigma\Delta^2\alpha_i &= (\alpha_n + \alpha_1) - (\alpha_{n-1} + \alpha_2) \\ \Sigma\Delta^3\alpha_i &= (\alpha_n - \alpha_1) - 2(\alpha_{n-1} - \alpha_2) + (\alpha_{n-2} - \alpha_3) \\ \Sigma\Delta^4\alpha_i &= (\alpha_n + \alpha_1) - 3(\alpha_{n-1} + \alpha_2) + 3(\alpha_{n-2} + \alpha_3) - \\ &\quad - (\alpha_{n-3} + \alpha_4) \\ &\dots\dots\dots\dots\dots\dots\dots\end{aligned}\right\} \quad (6)$$

Codeste formule valgono qualunque sia la legge di successione degli elementi considerati, e quindi per $\alpha_i = y_i$ e per $\alpha_i = x_i^s$ dànno rispettivamente le somme delle differenze finite degli y_i e degli x_i^s.

Ma, riguardo a questi ultimi, poichè abbiamo ammesso che gli x_i siano in progressione aritmetica, possiamo scrivere:

$$\begin{aligned}
\Sigma \Delta^k x_i^s = &\left[(x_1 + (n-1)\,h)^s \pm x_1^s \right] - \\
&- (k-1) \left[(x_1 + (n-2)\,h)^s \pm (x_1 + h)^s \right] + \\
&+ \binom{k-1}{2} \left[(x_1 + (n-3)\,h)^s \pm (x_1 + 2\,h)^s \right] - \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
&\mp \left[(x_1 + (n-k)\,h)^s \pm (x_1 + (k-1)\,h)^s \right],
\end{aligned}$$

donde:

$$\left.\begin{aligned}
&\Sigma \Delta x_i = (n-1)h, \quad \Sigma \Delta x^2{}_i = (n-1)h(2x_1 + (n-1)h), \quad \Sigma \Delta x^3{}_i = (n-1)\,h\,(3x^2{}_1 + \\
&\qquad\qquad + 3x_1(n-1)h + (n-1)^2h^2), \quad \ldots \\
&\Sigma \Delta^2 x^2{}_i = 2(n-2)h^2, \qquad \Sigma \Delta^2 x^3{}_i = 3(n-2)h^2(2x_1 + (n-1)h), \ldots \\
&\qquad\qquad \Sigma \Delta^3 x^3{}_i = 6(n-3)h^3, \quad \ldots
\end{aligned}\right\} \quad (7)$$

3. — Se gli x_i siano uguali ai primi n numeri della serie naturale a cominciare da 0, chiamando m l'ultimo termine della serie, le formule precedenti porgono:

$$\left.\begin{aligned}
&n = m+1, \quad \sum_{x=0}^{m} x = \frac{m(m+1)}{2}, \quad \sum_{x=0}^{m} x^2 = \frac{m(m+1)(2m+1)}{6}, \quad \sum_{x=0}^{m} x^3 = \frac{m^2(m+1)^2}{4}, \quad \ldots \\
&\sum_{x=0}^{m} \Delta x = m, \qquad \sum_{x=0}^{m} \Delta x^2 = m^2, \qquad \sum_{x=0}^{m} \Delta x^3 = m^3, \quad \ldots \\
&\qquad\qquad \sum_{x=0}^{m} \Delta^2 x^2 = 2(m-1), \qquad \sum_{x=0}^{m} \Delta^2 x^3 = 3m(m-1), \ldots \\
&\qquad\qquad\qquad\qquad\qquad\qquad \sum_{x=0}^{m} \Delta^3 x^3 = 6(m-2), \quad \ldots
\end{aligned}\right\} \quad (8)$$

Ed il sistema (2 *bis*), nel caso più semplice dell'interpolazione di una costante, fornirà:

$$a_0 = \frac{1}{m+1} \Sigma y_i;$$

nel caso dell'interpolazione di una retta:

$$\left.\begin{aligned} a_1' &= \frac{1}{m}\,\Sigma\,\Delta\,y_i\,, \\ a_0' &= \frac{1}{m+1}\Sigma\,y_i - \frac{1}{2}\,\Sigma\Delta\,y_i; \end{aligned}\right\} \quad (9)$$

nel caso di una parabola quadratica:

$$\left.\begin{aligned} a_2'' &= \frac{1}{2(m-1)}\,\Sigma\Delta^2\,y_i, \\ a_1'' &= \frac{1}{m}\Sigma\Delta y_i - \frac{m}{2(m-1)}\Sigma\,\Delta^2\,y_i, \\ a_0'' &= \frac{1}{m+1}\Sigma\,y_i - \frac{1}{2}\Sigma\Delta y_i + \frac{m}{12}\Sigma\Delta^2\,y_i; \end{aligned}\right\} \quad (10)$$

e così via.

Così procedendo, le formule successive per il calcolo dei parametri contengono sempre tutti i termini delle corrispondenti formule precedenti; e quindi, ove non sia prefissato il grado della funzione interpolatrice e si voglia procedere per approssimazioni successive, non occorrerà in ogni stadio ricominciare i calcoli da capo, ma si potranno utilizzare completamente quelli già eseguiti negli stadî precedenti dell'interpolazione.

Si noti che in un dato stadio si potrebbero utilizzare anche gli scarti, che già si fossero calcolati tra gli y_i ed i valori corrispondenti desunti dalla funzione interpolatrice.

Infatti, chiamando: $a_o^{(s)}$, $a_1^{(s)}$, ... $a_s^{(s)}$ i parametri già ottenuti, e:

$$d_i^{(s)} = y_i - a_o^{(s)} - a_1^{(s)} x_i - \ldots - a_s^{(s)} x_i^{(s)}$$

gli scarti calcolati in base ad essi, lo stadio successivo fornirà:

$$\begin{aligned} a_o^{(s+1)} &= a_o^{(s)} + \varepsilon_o\,, \\ a_1^{(s+1)} &= a_1^{(s)} + \varepsilon_1\,, \\ &\cdots\cdots\cdots \\ a_s^{(s+1)} &= a_s^{(s)} + \varepsilon_s\,, \\ a_{s+1}^{(s+1)}&, \end{aligned}$$

e si avrà:

$$d_i^{(s+1)} = d_i^{(s)} - \varepsilon_0 - \varepsilon_1 x_i - \ldots - \varepsilon_s x_i^s - a_{s+1}^{(s+1)} x_i^{(s+1)}. \qquad (11)$$

4. — Nel caso che $m+1$, ossia il numero delle coppie di dati empirici, sia uguale al numero dei parametri da calcolare, si otterrà una parabola passante non più *tra*, ma *per* i punti dati; e le formule del numero precedente — insieme alle successive, che si potrebbero aggiungere senza difficoltà — condurranno agli stessi risultati di quelle notissime del LAGRANGE e del NEWTON. Esse forniranno ancora un altro metodo per far passare una parabola di r^{mo} ordine per $r+1$ punti; metodo che, peraltro, presenta una maggior semplicità dei due citati, oltrechè una portata più generale almeno nel senso che si può seguire anche per un numero di punti maggiore di $r+1$, nel qual caso fornirà una parabola approssimata a quei punti sotto le condizioni che abbiamo illustrato.

Si voglia, ad es., far passare una parabola del secondo ordine per i tre punti di coordinate: (0;8), (1;10), (2;15).

In base alle (6), essendo nel nostro caso: $\alpha_{n-1} = \alpha_2 = 10$, avremo anzitutto:

$$\Sigma y_i = 8 + 10 + 15 = 33,$$

$$\Sigma \Delta y_i = 15 - 8 = 7,$$

$$\Sigma \Delta^2 y_i = 23 - 10 \times 2 = 3,$$

e poscia, in forza delle (10):

$$a_2'' = \frac{3}{2} = 1{,}5$$

$$a_1'' = \frac{7}{2} - 3 = 0{,}5$$

$$a_0'' = \frac{33}{3} - \frac{7}{2} + \frac{1}{2} = 8.$$

Per $x = 0, 1, 2$, la funzione:

$$f(x) = 8 + 0{,}5x + 1{,}5x^2$$

assumerà i valori assegnati: 8, 10, 15.

Se invece, vogliamo interpolare tra quei tre punti una retta, in forza delle (9) avremo:

$$a_1' = \frac{7}{2} = 3{,}5$$

$$a_0' = \frac{33}{3} - \frac{7}{2} = 7{,}5,$$

e per $x = 0, 1, 2$, la funzione:

$$f(x) = 7{,}5 + 3{,}5x$$

assumerà valori obbedienti alle due condizioni che le somme di essi e delle loro differenze prime siano uguali a quelle corrispondenti calcolate sui valori assegnati: 8, 10, 15.

Ed infatti si ha:

x	$f(x)$	$\Delta f(x)$
0	7,5	3,5
1	11	3,5
2	14,5	
	33	7

Come si vede, col metodo esposto le due teorie dell'interpolazione esatta ed approssimata vengono a differire soltanto per il numero delle somme di differenze empiriche, che si possono uguagliare alle corrispondenti somme teoriche.

Interpolare tra una successione di $r+1$ coppie di numeri una funzione parabolica di r^{mo} grado significherà uguagliare le somme delle prime $r+1$ differenze empiriche (a cominciare da $\sum \Delta^0 y_i = \sum y_i$) alle somme corrispondenti calcolate sulla funzione; mentre interpolare tra quella successione funzioni paraboliche di $(r-1)^{mo}$, $(r-2)^{mo}$, ... 2^0, 1^0 grado significherà porre le condizioni di uguaglianza rispettivamente tra le somme delle prime r, $r-1$, ... 3, 2 differenze empiriche e teoriche.

5. — Tornando all'ipotesi che gli x_i siano comunque in progressione aritmetica, notiamo che in alcuni casi semplici non è neppur necessario l'impiego di tutte le formule generali, di cui al n. 2.

Se ad es. si volesse interpolare la parabola del 2^0 ordine tra i punti di coordinate: (2;8), (4;7), (6;15), escludendo dai primi membri del sistema (2 *bis*) i termini successivi a quelli in x_i^2 ed applicando solo le (6), avremmo immediatamente:

$$n=3,\ \Sigma x_i=2+4+6=12,\ \Sigma x_i^2=4+16+36=56,\ \Sigma y_i=8+7+15=30,$$
$$\Sigma\Delta x_i=6-2=4,\ \Sigma\Delta x_i^2=36-4=32,\ \Sigma\Delta y_i=15-8=7,$$
$$\Sigma\Delta^2 x_i^2=40-16\times 2=8,\ \Sigma\Delta^2 y_i=23-7\times 2=9.$$

E quindi:

$$a_2''=\frac{9}{8}=1{,}125$$

$$a_1''=\frac{7}{4}-1{,}125\frac{32}{4}=-7{,}25$$

$$a_0''=\frac{30}{3}+7{,}25\frac{12}{3}-1{,}125\frac{56}{3}=18.$$

La funzione:

$$f(x)=18-7{,}25\,x+1{,}125\,x^2$$

per $x=2,\ 4,\ 6$ assumerà i valori assegnati: 8, 7, 15.

Per interpolare, invece, una retta tra quei punti avremmo:

$$a_1'=\frac{7}{4}=1{,}75$$

$$a_0'=\frac{30}{3}-1{,}75\frac{12}{3}=3,$$

e la funzione:

$$f(x)=3+1{,}75\,x$$

assicurerà l'uguaglianza delle somme dei valori e delle differenze prime degli y_i con quelle corrispondenti teoriche.

Un esempio concreto: Si voglia interpolare una funzione parabolica di secondo grado tra le seguenti otto coppie di dati, ricavati da un censimento italiano:

Età	Persone di età $> x$
x	y
30	12.062
35	10.113
40	8.283
45	6.556
50	4.914
55	3.521
60	2.245
65	1.290
	48.984

Qui trovano conveniente applicazione le formule (5), (6) e (7), e si ha:

$$n=8,\ \Sigma x_i=95\times\frac{8}{2}=380,\ \Sigma x_i^2=8\times900+56(30\times5+\frac{15}{6}25)=19.100,\ \Sigma y_i=48.984,$$

$$\Sigma\Delta x_i=35,\quad \Sigma\Delta x_i^2=35(60+35)=3.325,\quad \Sigma\Delta y_i=1.290-12.062=-10.772,$$

$$\Sigma\Delta^2 x_i^2=12\times25=300,\quad \Sigma\Delta^2 y_i=13.352-12.358=994.$$

Quindi, in forza delle (2 *bis*):

$$a_2''=\frac{994}{300}=3,3133$$

$$a_1''=\frac{-10.772}{35}-3,3133\frac{3.325}{35}=-622,535$$

$$a_0''=\frac{48.984}{8}+622,535\frac{380}{8}-3,3133\frac{19.100}{8}=27.782,91.$$

Dalla funzione:

$$f(x)=27.782,91-622,535x+3,3133x^2$$

si ottiene:

x	$f(x)$	$y - f(x)$
30	12.088,83	— 26,83
35	10.052,98	+ 60,02
40	8.182,79	+ 100,21
45	6.478,27	+ 77,73
50	4.939,41	— 25,41
55	3.566,22	— 45,22
60	2.358,69	— 113,69
65	1.316,83	— 26,83
	48.984,02	

Si verifica, inoltre, che:

$$\Sigma\Delta f(x) = 1.316,83 - 12.088,83 = -10.772$$

$$\Sigma\Delta^2 f(x) = 13.405,66 - 12.411,67 = 993,99.$$

Adunque, con l'approssimazione consentita dai decimali trascurati nei calcoli, le somme dei valori e delle differenze prime e seconde della funzione coincidono con quelle empiriche. L'ultima colonna della tavola precedente contiene gli scarti tra i valori empirici e quelli teorici, i quali dimostrano la buona approssimazione raggiunta con la parabola quadratica. Lo scarto quadratico medio è:

$$\sqrt{\frac{\Sigma(y_i - f(x_i))^2}{n}} = 67,8,$$

ossia l'1, 1 % della media degli y_i.

Il MORTARA (*Lezioni*, p. 211) ha già interpolato tra quei dati la stessa funzione, seguendo il metodo dei minimi quadrati, ed ha ottenuto uno scarto quadratico medio più piccolo (50,7), ciò che era da attendersi dalla condizione stessa del metodo.

Se si volesse procedere all'interpolazione della parabola cubica, basterebbe calcolare le espressioni:

$$\Sigma x_i^3,\ \Sigma\Delta x_i^3,\ \Sigma\Delta^2 x_i^3,\ \Sigma\Delta^3 x_i^3,\ \Sigma\Delta^3 y_i.$$

Utilizzando anche parte dei calcoli già eseguiti, si otterrebbero rapidamente i nuovi parametri, e — in base alle differenze tra i parametri corrispondenti — potrebbe anche applicarsi la formula (11) per il calcolo dei nuovi scarti. E così di seguito.

Si noti che quando — come nell'esempio scelto — le differenze tra i successivi y_i non presentino variazioni di segno, l'uguaglianza tra le somme empiriche e quelle teoriche delle differenze prime assicurerà altresì l'uguaglianza tra il cosidetto *campo di variabilità* dei valori teorici ed empirici.

6. — Se gli x_i non formassero una progressione aritmetica, le differenze singole non presenterebbero più le proprietà, già sommariamente ricordate al n. 1, ma: *a*) il metodo non perderebbe valore, perchè non è legato a codeste proprietà; *b*) sarebbe sempre possibile calcolare speditamente, a mezzo delle (6), le somme delle differenze successive che compaiono nel sistema:

$$\left.\begin{aligned} a_0 n + a_1 \Sigma x_i + a_2 \Sigma x_i^2 + \ldots + a_r \Sigma x_i^r &= \Sigma y_i \\ \cdots\cdots\cdots\

si avrebbe:

$$a_2'' = \frac{1}{\alpha'}(\Sigma\Delta^2 y_i - \alpha\Sigma\Delta y_i),$$

$$a_1'' = \frac{\Sigma\Delta y_i}{\Sigma\Delta x_i} - a_2''\frac{\Sigma\Delta x_i^2}{\Sigma\Delta x_i},$$

$$a_0'' = \frac{\Sigma y_i}{n} - a_1''\frac{\Sigma x_i}{n} - a_2''\frac{\Sigma x_i^2}{n}.$$

Per l'interpolazione della parabola cubica, ponendo ancora:

$$\frac{\Sigma\Delta^3 x_i}{\Sigma\Delta x_i} = \beta, \quad \Sigma\Delta^3 x_i^2 - \beta\Sigma\Delta x_i^2 = \beta',$$

si avrebbe:

$$a_3''' = \frac{\alpha'(\Sigma\Delta^3 y_i - \beta\Sigma\Delta y_i) - \beta'(\Sigma\Delta^2 y_i - \alpha\Sigma\Delta y_i)}{\alpha'(\Sigma\Delta^3 x_i^3 - \beta\Sigma\Delta x_i^3) - \beta'(\Sigma\Delta^2 x_i^3 - \alpha\Sigma\Delta x_i^3)},$$

$$a_2''' = \frac{1}{\alpha'}(\Sigma\Delta^2 y_i - \alpha\Sigma\Delta y_i) - a_3'''\frac{1}{\alpha'}(\Sigma\Delta^2 x_i^3 - \alpha\Sigma\Delta x_i^3),$$

$$a_1''' = \frac{\Sigma\Delta y_i}{\Sigma\Delta x_i} - a_2'''\frac{\Sigma\Delta x_i^2}{\Sigma\Delta x_i} - a_3'''\frac{\Sigma\Delta x_i^3}{\Sigma\Delta x_i},$$

$$a_0''' = \frac{\Sigma y_i}{n} - a_1'''\frac{\Sigma x_i}{n} - a_2'''\frac{\Sigma x_i^2}{n} - a_3'''\frac{\Sigma x_i^3}{n},$$

e così di seguito.

Donde si rileva che, col metodo di eliminazione, le formule successive per il calcolo dei parametri contengono tutti i coefficienti delle corrispondenti formule precedenti; e che, quindi, procedendosi per approssimazioni successive, si possono utilizzare anche in questo caso parte dei calcoli eseguiti negli stadî precedenti dell'interpolazione e — in base alle differenze tra i parametri corrispondenti — applicare pure la formula (11) per il calcolo degli scarti successivi.

In questo caso, però, il metodo sarebbe particolarmente raccomandabile nell'interpolazione approssimata.

7. — Anche nell'interpolazione delle funzioni non paraboliche, che siano o si possano ridurre a forme lineari rispetto ai parametri (o rispetto a funzioni numericamente determinate di essi), non è da escludere che codesto metodo possa dare buoni risultati e nella pratica statistica sia talvolta da preferire ai noti metodi d'interpolazione approssimata.

La facilità dei calcoli resterebbe notevolissima, anche perchè potrebbero sempre usarsi le formule (6), sostituendo in esse ad α_i una funzione numericamente determinata di x_i o di y_i.

Se si volesse, ad es., interpolare la funzione:

$$f(\theta) = a_0 + a_1 \, sen\, \theta + a_2 \, cos\, \theta,$$

il sistema, lineare rispetto ai parametri,:

$$\begin{aligned} a_0 n + a_1 \Sigma sen\, \theta_i + a_2 \Sigma cos\, \theta_i &= \Sigma y_i \\ a_1 \Sigma \Delta sen\, \theta_i + a_2 \Sigma \Delta cos\, \theta_i &= \Sigma \Delta y_i \\ a_1 \Sigma \Delta^2 sen\, \theta_i + a_2 \Sigma \Delta^2 cos\, \theta_i &= \Sigma \Delta^2 y_i \end{aligned}$$

permetterebbe di calcolare facilmente le somme delle differenze finite a mezzo delle (6), e potrebbe anche fornire risultati soddisfacenti.

Per fare un esempio dei più semplici, adatteremo la funzione:

$$f(x) = a_0 x^{-a_1}$$

alla distribuzione di frequenze dei contribuenti con reddito superiore a lire 1000, inscritti nei ruoli della tassa di famiglia in 23 città italiane nel 1887.

Redditi in lire	Numero di contribuenti con redditi $> x$		
x	y	$log\, x$	$log\, y$
1.000	59.486	3	4,77441
2.000	26.968	3,30103	4,43085
4.000	9.766	3,60206	3,98972
7.000	4.264	3,84510	3,62982
10.000	2.397	4	3,37967
15.000	1.310	4,17609	3,11727
25.000	645	4,39794	2,80956
	104.836	26,32222	26,13130

Ponendo la funzione nella forma:

$$\log f(x) = \log a_0 - a_1 \log x,$$

si ha:

$n=7$, $\Sigma \log x_i = 26{,}32222$, $\Sigma \log y_i = 26{,}13130$

$\Sigma \Delta \log x_i = 4{,}39794 - 3 = 1{,}39794$, $\Sigma \Delta \log y_i = 2{,}80956 - 4{,}77441 = -1{,}96485$

e quindi:

$$a_1 = \frac{-1{,}96485}{1{,}39794} = -1{,}4055$$

$$\log a_0 = \frac{26{,}13130}{7} + 1{,}4055 \frac{26{,}32222}{7} = 9{,}0182$$

$$\log f(x) = 9{,}0182 - 1{,}4055 \log x.$$

Ne segue:

$$f(x) = 1.042.800.000\, x^{-1{,}4055}.$$

Le prime due colonne della tavola seguente riportano i valori di $\log f(x)$ e gli scarti di essi dai logaritmi dei valori osservati:

Metodo delle differenze		Metodo del Cauchy	
$\log f(x)$	$\log y - \log f(x)$	$\log f(x)$	$\log y - \log f(x)$
4,8017	— 0,0273	4,8354	— 0,0610
4,3786	+ 0,0522	4,3990	+ 0,0318
3,9555	+ 0,0342	3,9624	+ 0,0273
3,6139	+ 0,0159	3,6101	+ 0,0197
3,3962	— 0,0165	3,3854	— 0,0057
3,1487	— 0,0314	3,1301	— 0,0128
2,8369	— 0,0273	2,8085	— 0,0011
26,1316	— 0,1025	26,1309	— 0,0795
	+ 0,1023		+ 0,0799

Le altre due colonne contengono i risultati ottenuti dal Benini (*Principi*, pag. 185) col metodo del Cauchy.

Passando dai logaritmi ai numeri, si ha:

Metodo delle differenze		Metodo del CAUCHY	
$f(x)$	$y - f(x)$	$f(x)$	$y - f(x)$
63.120	— 3.634	68.450	— 8.964
23.910	+ 3.058	25.060	+ 1.908
9.026	+ 740	9.170	+ 596
4.111	+ 153	4.075	+ 189
2.490	— 93	2.428	— 31
1.408	— 98	1.350	— 40
687	— 42	643	+ 2

donde si rileva il fatto, già noto, che la funzione scelta si adatta male ai dati empirici in corrispondenza ai redditi più piccoli, ma si constata altresì che i nostri risultati potrebbero anche preferirsi a quelli ottenuti col metodo del CAUCHY.

Gli è che l'uguaglianza tra le somme empiriche e quelle teoriche delle differenze prime viene ad attenuare la divergenza fra i valori estremi di $f(x)$, uguagliandola a quella empirica.

Analogamente, se si volesse ad es. interpolare la funzione:

$$f(x) = a_0 \, x^{-a_1} \, e^{\frac{-a_2}{x}},$$

sempre in base ai logaritmi volgari avremmo:

$$\log f(x) = \alpha - a_1 \log x - \beta \frac{1}{x},$$

dove:

$$\alpha = \log a_0,$$
$$\beta = a_2 \log e = 0{,}4343 \, a_2 .$$

Ne risulterebbe il sistema:

$$\alpha_n - a_1 \Sigma \log x_i - \beta \Sigma \frac{1}{x_i} = \Sigma \log y_i$$

$$- a_1 \Sigma \Delta \log x_i - \beta \Sigma \Delta \frac{1}{x_i} = \Sigma \Delta \log y_i$$

$$- a_1 \Sigma \Delta^2 \log x_i - \beta \Sigma \Delta^2 \frac{1}{x_i} = \Sigma \Delta^2 \log y_i$$

donde potrebbero facilmente ricavarsi i valori di β, a_1, α, e quindi anche di a_2 e a_0.

La funzione precedente è il tipo V del PEARSON, che — adottato per le distribuzioni di frequenze dei morti, delle successioni ereditarie, dei patrimoni, ecc. — è stato da noi proposto, in base a considerazioni teoriche, per la rappresentazione analitica delle distribuzioni complete dei redditi globali (*Giornale degli economisti*, Novembre 1921, e *Statistica metodologica*, 1923, n. 103). Ma è chiaro che quella funzione dovrà presentare divergenze sistematiche dalle distribuzioni di frequenze di speciali categorie di redditi, in relazione allo speciale andamento di esse.

È stato, però, giustamente avvertito che le condizioni del metodo interpolatorio impiegato sarebbero in tal guisa rispettate dai logaritmi, ma non dai numeri. A quest'ultimo intento potrebbe seguirsi, anche col nostro metodo, l'artificio recentemente proposto dal MORTARA (*Lezioni*, p. 189) od anche quello del PARETO (*Giornale degli economisti*, Giugno 1908, p. 447).

8. — In generale, col metodo esposto potrebbero pure interpolarsi funzioni non lineari rispetto ai parametri a mezzo di quell'artificio notissimo, che suol seguirsi col metodo dei minimi quadrati.

Si scelgano arbitrariamente tante coppie di dati empirici, quanti sono i parametri della funzione scelta, e, con i metodi o gli artifici suggeriti dall'algebra, si calcolino provvisoriamente i valori dei parametri.

Codesti valori a'_s presenteranno una certa approssimazione a quelli definitivi a_s, che avremmo ottenuto, se avessimo potuto utilizzare tutte le coppie disponibili.

Ora, se noi possiamo *a priori* ammettere che le differenze $\varepsilon_s = a_s - a'_s$ siano piuttosto piccole, cosicchè ne siano trascurabili i quadrati e le potenze successive, sviluppando in serie del TAYLOR la funzione:

$$f(x,\ a'_0 + \varepsilon_0,\ a'_1 + \varepsilon_1,\ \ldots\ a'_r + \varepsilon_r),$$

potremo scrivere le condizioni:

$$\Sigma\ (\Delta^k f(x_i,\ a_0,\ a_1,\ \ldots\ a_r) - \Delta^k y_i) = 0, \qquad (k = 0,\ 1,\ 2,\ \ldots\ r)$$

nella forma:

$$\Sigma\left\{\Delta^k f(x_i, a'_0, a'_1, \ldots a'_r)+\Delta^k(\varepsilon_0\frac{\partial f}{\partial a'_0}+\varepsilon_1\frac{\partial f}{\partial a'_1}+\ldots+\varepsilon_r\frac{\partial f}{\partial a'_r})-\Delta^k y_i\right\}=0.$$

Ponendo:

$$\Delta^k d'_i=\Delta^k y_i-\Delta^k f(x_i, a'_0, a'_1, \ldots a'_r),$$

$$\frac{\partial f}{\partial a'_0}=A_i^{(0)},\quad \frac{\partial f}{\partial a'_1}=A_i^{(1)},\ \ldots\ \frac{\partial f}{\partial a'_r}=A_i^{(r)},$$

esse diventano:

$$\Sigma\,(\Delta^k(\varepsilon_0 A_i^{(0)}+\varepsilon_1 A_i^{(1)}+\ldots+\varepsilon_r A_i^{(r)})-\Delta^k d'_i)=0,$$

ossia:

$$\varepsilon_0\,\Sigma\Delta^k A_i^{(0)}+\varepsilon_1\,\Sigma\Delta^k A_i^{(1)}+\ldots+\varepsilon_r\,\Sigma\Delta^k A_i^{(r)}=\Sigma\Delta^k d'_i.$$

Codesto sistema lineare determinerà i valori ε_s, ossia le correzioni da apportare ai parametri provvisori, affinchè siano rispettate le condizioni del metodo.

Se gli ε_s non riusciranno piuttosto piccoli, non sarà giustificata l'ipotesi su cui codesto procedimento si fonda; ma si potranno assumere come nuovi parametri provvisori i valori corretti e ripetere il procedimento, in modo da pervenire a valori di ε_s piuttosto piccoli.

Non è da escludere che, interpolando in tal guisa la nota curva di GOMPERTZ-MAKEHAM tra una curva empirica di mortalità, possano ottenersi risultati soddisfacenti.

***

In base alla considerazione delle differenze finite, si potrebbero escogitare parecchi altri metodi interpolatorî: si potrebbe ad. es. pensare a render minima la somma dei quadrati degli scarti tra le differenze finite (algebriche) di un dato ordine, teoriche ed empiriche;

e, per ovviare al numero insufficiente di condizioni, che si verrebbero generalmente ad assumere, si potrebbero aggiungere altre condizioni convenienti, come l'uguaglianza della somma dei valori y_i a quelli teorici, ecc.

Ma noi crediamo che difficilmente si possa ancora costruire un metodo che, per semplicità di criterî, carattere elementare di procedimenti analitici e brevità di calcoli numerici, sia in grado di competere con quello che abbiamo sopra proposto; onde non ci resta che augurarci che a questi pregi si aggiunga sempre la bontà dei risultati, affinchè esso possa essere largamente usato nella pratica statistica.

Bari, aprile del 1924.

R. ISTITUTO SUPERIORE
di scienze econ. e comm.

Felice Vinci.

Biblioteca Storica
58685