La Rivista internazionale di Statistica METRON esce in fascicoli. Quattro fascicoli consecutivi costituiscono complessivamente un volume di 700-800 pagine.

METRON accoglie articoli originali di metodologia statistica e di applicazioni statistiche alle varie discipline, e rassegne o discussioni di risultati raggiunti col metodo statistico in diversi campi della scienza o tali da poter interessare il cultore della statistica. Pubblica altresì una bibliografia di tutte le opere e riviste ricevute in omaggio od in cambio.

Articoli e rassegne potranno essere scritti in italiano, francese, inglese o tedesco. I manoscritti in lingua francese, inglese o tedesca dovranno essere dattilografati.

La collaborazione non è retribuita. Gli autori riceveranno gratuitamente 25 estratti dei lavori pubblicati.

I manoscritti per la pubblicazione dovranno essere indirizzati al *Prof. Corrado Gini, R. Università di Roma — Istituto di Statistica,* oppure al membro del Comitato direttivo che rappresenta lo Stato a cui l'autore appartiene. Gli autori sono pregati di conservare copia del manoscritto inviato, poichè, nel caso che questo non venga pubblicato, la Direzione non ne garantisce la restituzione.

Al Prof. Corrado Gini dovranno pure essere indirizzate le richieste di cambi da parte di riviste o di altri periodici e ogni pubblicazione inviata in cambio od in omaggio.

Le richieste di abbonamenti, del pari che i versamenti, dovranno invece essere indirizzati alla *Amministrazione del « Metron »* presso l'*Istituto di Statistica della R. Università di Roma — Via delle Terme di Diocleziano,* 10.

Il prezzo di abbonamento per ciascun Volume è di **100 Lire italiane** e quello del fascicolo di **30 Lire italiane,** porto compreso.

La Revue Internationale de Statistique METRON paraît par livraisons. Quatre livraisons consécutives forment un volume de 700-800 pages.

METRON publie des articles originaux de méthodologie statistique et d'applications statistiques aux différentes disciplines, ainsi que des revues ou des discussions des résultats obtenus par la méthode statistique dans toutes les sciences ou bien intéressant les savants qui s'occupent de statistique.

METRON publie aussi une bibliographie de tous les ouvrages et revues reçues en hommage ou en échange.

Les articles et les revues pourront être écrites en français, en italien, en anglais ou en allemand. Les manuscrits en français, en anglais ou en allemand doivent être envoyés dactylographiés.

On enverra gratis aux auteurs 25 copies tirées à part de leurs travaux après publication.

On adressera les manuscrits pour la publication à *M. le Prof. Corrado Gini, Istituto di Statistica, R. Università di Roma (Italie),* ou bien au membre du comité de direction représentant le pays de l'auteur. On prie les auteurs de garder une copie du manuscrit qu'ils adressent à la Revue, car, en cas de non publication, la rédaction ne garantit pas de pouvoir le renvoyer.

Les demandes d'échange de la part des Revues et des autres périodiques, ainsi que toutes les publications envoyées en échange ou en hommage doivent aussi être adressées au Prof. Corrado Gini.

Les demandes de nouveaux abonnements, ainsi que tout payement, devront être adressés à l'*Administration du « Metron »* auprès de l'*Institut de Statistique de l'Université Royale de Rome — Via delle Terme di Diocleziano, 10, Roma, Italie.*

Le prix d'abonnement par volume est fixé à 100 Lires it. et le prix par fascicule est de 30 Lires it. frais d'envoi compris.

METRON

RIVISTA INTERNAZIONALE DI STATISTICA — REVUE INTERNATIONALE DE STATISTIQUE
INTERNATIONAL REVIEW OF STATISTICS — INTERNATIONALE STATISTISCHE ZEITSCHRIFT

Vol. XII - N. 2 **31-V-1935.**

SOMMARIO — SOMMAIRE — CONTENTS — INHALT

ROMA
AMMINISTRAZIONE DEL «METRON»
R. UNIVERSITÀ — ISTITUTO DI STATISTICA

ARTICOLI GIUNTI ALLA RIVISTA CHE VERRANNO PUBBLICATI NEI PROSSIMI NUMERI.

(*Secondo l'ordine d'arrivo*).

ARTIKEL, DIE AN DIE ZEITSCHRIFT ANGELANGT SIND UND WELCHE IN DEN NACHFOLGENDEN NUMMERN ERSCHEINEN WERDEN.

(*Nach der Reihenfolge des Eingangs*)

ARTICLES REÇUS PAR LA REVUE ET À PARAÎTRE PROCHAINEMENT.

(*D'après la date de reception*).

ARTICLES RECEIVED BY THE REVIEW WHICH WILL BE PUBLISHEED IN FUTURE ISSUES.

(*According to date of receipt*)

S. Koller — *Die Analyse der Abhängigkeitsverhältnisse in zwei Korrelationssystemen.*

E. Zwinggi — *Zur Frage des Beharrungszustandes.*

S. Kullback — *A Note on the Distribution of a certain partial belonging Coefficient.*

S. Kullback — *A Note on the multiple Correlation Coefficient.*

A. Linder — *" Wahrscheinlichkeitsansteckung " und Differenzengleichungen.*

Gli Autori degli articoli inviati per la pubblicazione nella Rivista, rinunciano in favore della medesima alla proprietà letteraria degli articoli stessi, qualora vengano pubblicati.

Les Auteurs des articles envoyés à la Revue, pour y être publiés, renoncent, en faveur de celle-ci, à la propriéte littéraire de leurs articles, s'ils sont acceptés.

The Authors of papers sent for publication in the Review are supposed to give up their copyright in favour of the Review if the papers are published.

Die Verfasser der zur Veröffentlichung in der Zeitschrift zugesandten Aufsätzs, werden, falls selbige veröffentlicht werden, auf ihre Verfasserrechte zu Gunsten der Zeitschrift verzichten müssen.

ROBERTO MOGNO

Di un metodo di interpolazione statistica

SOMMARIO

PARTE I.

METODO DELLE SINTESI.

§ 1. — Diremo, col GINI, *) che l'interpolazione è quel procedimento col quale, ad una data successione di dati, che esprimono valori, o somme di valori, assunti da una funzione y corrispondentemente a valori, o classi di valori, assunti da una variabile indipendente x, si sostituisce una successione di dati più completa, che può includere i dati della successione originaria, oppure può sostituirli tutti o in parte, con altrettanti dati teorici.

I procedimenti ai quali più spesso si ricorre, sono analitici, ed in tal caso, ha particolare importanza, la distinzione tra interpolazione matematica, in cui, cioè, la formula interpolatrice corrisponde ad una curva che passa per gli n punti del piano, e interpolazione statistica, in cui la formula interpolatrice corrisponde a una curva che passa

*) C. GINI. *Considerazioni su l'interpolazione e la perequazione delle serie statistiche.* «Metron», Vol. I, N. 3, 1921.

accanto agli n punti del piano ; in quest'ultimo caso, fissato il tipo, il grado della funzione e il numero k dei suoi parametri, vi sono molti metodi che insegnano a determinarli, in modo che i valori calcolati della funzione risultante dall'interpolazione soddisfino a certe condizioni o a certe altre. Il più semplice dei metodi di interpolazione è quello delle somme, che soddisfa alla condizione che, raggruppando gli n dati osservati in k classi, la somma dei valori che rientrano in ciascuna classe, riferita o no a valori contigui, sia uguale alla somma dei corrispondenti valori calcolati della funzione.

Analoga condizione è soddisfatta, nel caso che la variabile sia continua, dal metodo delle aree.

Nel presente lavoro, esporremo un metodo di interpolazione statistica, prendendo le mosse dal metodo delle aree e delle somme, quest'ultimo nel caso che le somme di cui interessa la coincidenza nella distribuzione teorica e nella effettiva, siano riferite a valori contigui.

§ 2. — Siano dati i valori :

$$\int_{a_1}^{a_2} y\,dx\,,\ \int_{a_2}^{a_3} y\,dx\,,\ \ldots\ \int_{a_n}^{a_{n+1}} y\,dx \qquad 1)$$

rappresentazioni continue delle somme dei valori che una funzione statistica y assume in corrispondenza ai valori di x compresi negli n intervalli :

$$(a_1\,,\ a_2)\,,\ (a_2\,,\ a_3)\,,\ \ldots\ (a_n\,,\ a_{n+1}) \qquad 2)$$

e sia :

$$F\,(x\,,\ H_1\,,\ H_2\,,\ \ldots\ H_k)\,,\ k \leq n \qquad 3)$$

la funzione interpolatrice scelta.

La determinazione dei k parametri della 3), eseguita col metodo delle aree, ci conduce a vari sistemi di valori a seconda del modo col quale si suddivide in k intervalli il campo di variabilità della x.

Questo fatto, quando non vi sia alcuna ragione di preferenza per la scelta di una particolare suddivisione del campo di variabilità della x, rappresenta un inconveniente, tanto più grave, quanto più i dati 1) si allontanano da una curva regolare della forma che ha la 3).

Col metodo che qui esporremo, si ha, invece, il vantaggio di ottenere un unico sistema di equazioni che permette di determinare univocamente i k parametri della 3).

§ 3. — Supponiamo che la riduzione degli n intervalli 2) al numero di k, mediante unioni di intervalli contigui, si possa fare in un certo numero h di modi, ed indichiamo con:

$s_{1,1}$; $s_{2,1}$; .... $s_{k,1}$ i valori assunti da $\int y\,dx$ nei k intervalli relativi al 1° dei modi suddetti,

con: $s_{1,2}$; $s_{2,2}$; .... $s_{k,2}$ i valori assunti da $\int y\,dx$ nei k intervalli relativi al 2° dei modi suddetti,

con: $s_{1,3}$; $s_{2,3}$; .... $s_{k,3}$ i valori assunti da $\int y\,dx$ nei k intervalli relativi al 3° dei modi suddetti,

. . . . . 4)

con: $s_{1,h}$; $s_{2,h}$; .... $s_{k,h}$ i valori assunti da $\int y\,dx$ nei k intervalli relativi all'h^{mo} dei modi suddetti.

Per la determinazione dei parametri della 3), anzichè considerare solo i k valori s relativi ad uno di questi h modi, si considerano, invece, i k particolari valori:

$$\frac{\sum_{i=1}^{i=h} s_{1,i}}{\binom{n}{k}}\ *),\quad \frac{\sum_{i=1}^{i=h} s_{2,i}}{\binom{n}{k}},\quad \ldots\ \frac{\sum_{i=1}^{i=h} s_{k,i}}{\binom{n}{k}} \qquad 5)$$

che sintetizzano tutti i valori 4) e che, perciò, chiameremo sintesi k^{me} medie degli n valori 1) e per brevità indicheremo con:

$$Y_1\,,\ Y_2\,,\ \ldots .\ Y_k \qquad 6)$$

Se indichiamo ora con:

$$\overline{Y}_1\,,\ \overline{Y}_2\,,\ \ldots .\ \overline{Y}_k \qquad 7)$$

le sintesi k^{me} medie dei valori interpolati corrispondenti ai valori 1), basterà porre la sola condizione che le sintesi medie dei valori noti siano uguali alle corrispondenti sintesi teoriche, ossia che:

$$\begin{cases} Y_1 = \overline{Y}_1 \\ Y_2 = \overline{Y}_2 \\ \ldots\ldots\ldots \\ Y_k = \overline{Y}_k \end{cases} \qquad 8)$$

*) $\binom{n}{k}$ è la somma delle frequenze con le quali i valori 1) entrano nelle 5).

sistema di k equazioni che determina le k incognite :

$$H_1 , H_2 , H_3 , \ldots . H_k .$$

§ 4. — Il metodo esposto, si può ovviamente estendere al caso che i valori della funzione statistica siano dati in corrispondenza a un numero discreto di valori della variabile.

Basterà adattare al caso considerato il concetto di sintesi k^{ma}, e porre una condizione analoga alla 8).

Supposto infatti che siano :

$y_1 , y_2 , y_3 , \ldots . y_n$, 1') i valori che la funzione statistica y assume in corrispondenza ai punti :

$x_1 , x_2 , x_3 , \ldots . x_n$, 2') della variabile, interni rispettivamente agli intervalli :

$$(a_1 , a_2) \ ; \ (a_2 , a_3) \ ; (a_3 , a_n) , \ldots . (a_n , a_{n+1}),$$

e supposto, come prima, che la riduzione di essi al numero di k , eseguita mediante unioni di intervalli contigui, si possa fare in h modi diversi, indichiamo con :

$s_{1,1} \ ; \ s_{2,1} , \ldots . s_{k,1}$ le somme dei valori che y assume in corrispondenza ai punti 2') interni rispettivamente ai k intervalli relativi al 1° dei suddetti modi;

con : $s_{1,2} \ ; \ s_{2,2} , \ldots . s_{k,2}$ le analoghe somme relative al secondo dei suddetti modi;

. . . . . .

con : $s_{1,h} \ ; \ s_{2,h} , \ldots . s_{k,h}$ le analoghe somme relative all'h^{mo} dei suddetti modi.

Attribuendo alle s che compaiono nelle 5) del § precedente, questo nuovo significato, esse si diranno sintesi k^{me} medie degli n valori 1').

Per la determinazione dei parametri basterà porre la condizione già vista tra le sintesi dei valori noti e le sintesi teoriche.

§ 5. — Questo metodo, oltre al già accennato pregio della determinazione univoca dei parametri della funzione interpolatrice, ha anche quello della semplicità, perchè le formule date nella Parte IIª permettono il rapido calcolo delle sintesi. Ma, il pregio più notevole, che, quando non vi sia ragione alcuna di suddividere in un certo modo

particolare il campo di variabilità della x, rende il metodo preferibile a quello delle aree ed a quello delle somme, è di assicurare una migliore approssimazione, nella rappresentazione di tutti i dati, di quella che generalmente si otterrebbe applicando i metodi accennati.

Ciò è intuitivo, dal momento che alla determinazione dei parametri della 3) contribuiscono i valori s relativi a tutti gli h modi nei quali gli n intervalli 2) possono essere ridotti al numero di k, mediante unione di intervalli contigui.

Ma, a meglio lumeggiare la ragione di ciò, giova, a questo proposito, osservare che, quando la funzione interpolatrice scelta si può porre sotto la forma :

$$\varphi(y) = H_1 + H_2 x + H_3 x^2 + H_4 x^3 + \ldots . \qquad 9)$$

e quando, inoltre, i valori della variabile sono in progressione aritmetica, si ottengono, con questo metodo, per i parametri gli stessi valori che si otterrebbero, con calcoli più complicati, applicando il metodo dei minimi quadrati *) ; se, invece, i valori della variabile non sono in progressione aritmetica, i risultati si avvicinano tanto più a quelli che si otterrebbero col metodo dei minimi quadrati, quanto minore è la variabilità delle differenze prime dei valori della variabile.

Quest'ultimo risultato ci fa comprendere perchè, almeno quando la funzione interpolatrice è la 9), e quando, inoltre, il campo di variabilità della x sia diviso in intervalli uguali, applicando il metodo delle sintesi, anzichè quello delle aree, si ottiene, generalmente, un'approssimazione migliore della curva interpolatrice ai dati osservati.

Supposti, infatti, abbastanza piccoli ed uguali, gli intervalli in cui è suddiviso il campo di variabilità della x, si può porre con errore trascurabile :

$$\int_{a_1}^{a_2} y \, d\, x = (a_2 - a_1)\, y_{1,2} \qquad 10)$$

dove y è la funzione statistica considerata, e $y_{1,2}$ è un valore di y corrispondente ad un qualunque valore di x interno all'intervallo (a_1, a_2), ad esempio corrispondente al punto che lo bipartisce.

Eseguire, perciò, l'interpolazione col metodo delle sintesi, equivale praticamente, in questo caso, ad applicare il metodo dei minimi quadrati all'interpolazione dei valori :

---

*) Dimostrazione nella nota, alla fine di questa prima parte, § 8.

$$y_{1,2} \; ; \; y_{2,3} \; ; \; y_{3,4} \; ; \; \ldots . \; y_{n-1,n} ,$$

che la funzione statistica y assume in corrispondenza ai punti medi degli intervalli :

$$(a_1 , a_2) , (a_2 , a_3) , (a_3 , a_4) , \ldots . (a_{n-1} , a_n) .$$

§ 6. — Consideriamo ora il caso che la funzione interpolatrice scelta sia la :

$$y = H_1 + H_2 x + H_3 x^2 + \ldots . H_k x^{(k-1)} \qquad 11)$$

il sistema 8) diventa :

$$\left\{ \begin{array}{l} Y_1 = H_1 + H_2 X_1 + H_3 X_1^{(2)} + H_4 X_1^{(3)} + \ldots . H_k X_1^{(k-1)} \\ Y_2 = H_1 + H_2 X_2 + H_3 X_2^{(2)} + H_4 X_2^{(3)} + \ldots . H_k X_2^{(k-1)} \\ \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \\ Y_k = H_1 + H_2 X_k + H_3 X_k^{(2)} + H_4 X_k^{(3)} + \ldots . H_k X_k^{(k-1)} \end{array} \right. \qquad 12)$$

nel quale le : Y_1 , Y_2 , ... Y_k , sono le sintesi k^{me} medie dei valori noti, che si possono calcolare facilmente con le formule 6), 20) 24) 25) 26) 27) (Parte II[a]), e nel quale le :

$X_1 , X_2 , \ldots . X_k$ sono le sintesi k^{me} medie dei valori della variabile,

le : $X_1^{(2)} , X_2^{(2)} , \ldots . X_k^{(2)}$ sono le sintesi k^{me} medie dei quadrati dei valori della variabile,

. . . . . .

le : $X_1^{(k-1)} , X_2^{(k-1)} , \ldots . X_k^{(k-1)}$ sono le sintesi k^{me} medie delle potenze $(k-1)^{me}$ dei valori della variabile.

Quando gli n valori della x sono i primi n numeri naturali, le sintesi delle potenze dei valori della variabile, si possono esprimere, come già sappiamo (Parte II[a]), in funzione di n, ed in tal caso, per il calcolo delle sintesi teoriche conviene usare le formule 7), 15), 16), 43), 44), 45) [Parte II[a]].

Ad esempio, se la funzione interpolatrice è la retta di equazione :

$y = H_1 + H_2 x$, (H_1 e $H_2 \neq 0$), il sistema che determina i parametri, applicando le 7) (Parte II^a), diventa :

$$\begin{cases} Y_1 = H_1 + \dfrac{(n+1)}{3} H_2 \\ Y_2 = H_1 + \dfrac{2\,(n+1)}{3} H_2 \end{cases} \qquad 13)$$

Dal quale si ottiene :

$$H_1 = 2\,Y_1 - Y_2 \qquad 14)$$

$$H_2 = \frac{3\,(Y_2 - Y_1)}{n+1} \qquad 15)$$

Le quali ultime due formule, danno per H_1 ed H_2 gli stessi valori che si otterrebbero col metodo dei minimi quadrati.

Se la funzione interpolatrice è la :

$y = H_1 + H_2 x + H_3 x^2$, applicando le 15) e 16) (P. II^a), il sistema che determina i parametri diventa :

$$\begin{cases} Y_1 = H_1 + \dfrac{n+1}{4} H_2 + \dfrac{(n+1)\left(n - \frac{1}{2}\right)}{10} H_3 \\ Y_2 = H_1 + \dfrac{2\,(n+1)}{4} H_2 + \dfrac{(n+1)\,(3n+1)}{10} H_3 \\ Y_3 = H_1 + \dfrac{3\,(n+1)}{4} H_2 + \dfrac{(n+1)\left(6n + \frac{9}{2}\right)}{10} H_3 \end{cases}$$

dalle quali si ottiene :

$$H_1 = 3\,Y_1 - 3\,Y_2 + Y_3 \qquad 16)$$

$$H_2 = \frac{(12n+14)\,Y_1 - (20n+20)\,Y_2 + (8n+6)\,Y_3}{-(n+1)\,(n+2)} \qquad 17)$$

$$H_3 = \frac{-10\,Y_1 + 20\,Y_2 - 10\,Y_3}{-(n+1)\,(n+2)}\,. \qquad 18)$$

Analogamente, se la funzione interpolatrice è la :

$$y = H_1 + H_2 x + H_3 x^2 + H_4 x^3 , \qquad \text{applicando le :}$$

43), 44), 45), (P. II^a), si ha :

$$H_1 = 4\ Y_1 - 6\ Y_2 + 4\ Y_3 - Y_4 \tag{19}$$

$$H_2 = \left\{ \begin{array}{l} \dfrac{-(6n^2 + 18n + 14)\ Y_1 + (14n^2 + 37n + 27)\ Y_2}{(n+1)(n+2)(n+3)/5} + \\ + \dfrac{-(11n^2 + 25n + 18)\ Y_3 + (3n^2 + 6n + 5)\ Y_4}{(n+1)(n+2)(n+3)/5} \end{array} \right. \tag{20}$$

$$H_3 = \frac{(4n+5)\ Y_1 - (11n+12)\ Y_2 + (10n+9)\ Y_3 - (3n+2)\ Y_4}{(n+1)(n+2)(n+3)/15} \tag{21}$$

$$H_4 = \frac{-Y_1 + 3\ Y_2 - 3\ Y_3 + Y_4}{(n+1)(n+2)(n+3)/35} . \tag{22}$$

Analoghe formule si potrebbero ricavare per la determinazione dei parametri di una parabola di ordine qualunque, formule che ci danno gli stessi valori dei parametri che si otterrebbero con calcoli più laboriosi applicando il metodo dei minimi quadrati. Quando, invece, gli n valori della variabile non sono in progressione aritmetica, il calcolo delle sintesi teoriche si eseguisce più agevolmente applicando le formule : 4), 12), 13), 14), 28), 29), 30), 31) (P. II). Terminiamo con qualche interpolazione, eseguita col metodo delle sintesi, allo scopo di metterne in evidenza i vantaggi offerti.

§ 7. Esempio I. — Nel censimento del 31 Dicembre 1871, sono riportati i dati sulla distribuzione per età della popolazione italiana, di anno in anno.

Questi dati, riuniti in gruppi quadriennali di età, sono indicati nel seguente prospetto, a partire dal gruppo 27-30.

Gruppi di età	Ammontare dei censiti (migliaia)	U_x
27–30	1765	U_{27}
31–35	1327	U_{31}
35–39	1429	U_{35}
39–43	1412	U_{39}
43–47	1162	U_{43}
47–51	1301	U_{47}
51–55	817	U_{51}
55–59	756	U_{55}
59–63	834	U_{59}
63–67	550	U_{63}

Supponiamo che la distribuzione per età della popolazione possa essere rappresentata bene da una funzione della forma :

$$F(x) = H_1 + H_2 x + H_3 x^2 + H_4 x^3.$$

Allora, per determinare i quattro parametri della funzione interpolatrice scelta, basterà uguagliare le sintesi quarte dei valori osservati U_x, alle corrispondenti sintesi teoriche.

Le sintesi quarte delle U_x, già calcolate nel § 5 della P. II^a sono :

$$Y_1 = \frac{318.478}{210} = 1516{,}5619 \; ; \; Y_2 = \frac{266.262}{210} = 1267{,}91 \; ,$$

$$Y_3 = \frac{214.218}{210} = 1020{,}09 \; ; \; Y_4 = \frac{154.778}{210} = 737{,}04 \; .$$

Fissando, per semplicità, l'origine all'età 47, ed assunto il quadriennio ad unità di misura delle ascisse, i valori teorici corrispondenti alle U_x, sono :

$$\int_{-5}^{-4} F(x)\, dx = H_1 - 4{,}5\, H_2 + 20{,}33\, H_3 - 92{,}25\, H_4$$

$$\int_{-4}^{-3} F(x)\,dx = H_1 - 3{,}5\,H_2 + 12{,}33\,H_3 - 43{,}75\,H_4$$

$$\int_{-3}^{-2} F(x)\,dx = H_1 - 2{,}5\,H_2 + 6{,}33\,H_3 - 16{,}25\,H_4$$

$$\int_{-2}^{-1} F(x)\,dx = H_1 - 1{,}5\,H_2 + 2{,}33\,H_3 - 3{,}75\,H_4$$

$$\int_{-1}^{0} F(x)\,dx = H_1 - 0{,}5\,H_2 + 0{,}33\,H_3 - 0{,}25\,H_4$$

$$\int_{0}^{1} F(x)\,dx = H_1 + 0{,}5\,H_2 = 0{,}33\,H_3 + 0{,}25\,H_4$$

$$\int_{1}^{2} F(x)\,dx = H_1 + 1{,}5\,H_2 + 2{,}33\,H_3 + 3{,}75\,H_4$$

$$\int_{2}^{3} F(x)\,dx = H_1 + 2{,}5\,H_2 + 6{,}33\,H_3 + 16{,}25\,H_4$$

$$\int_{3}^{4} F(x)\,dx = H_1 + 3{,}5\,H_2 + 12{,}33\,H_3 + 43{,}75\,H_4$$

$$\int_{4}^{5} F(x)\,dx = H_1 + 4{,}5\,H_2 + 20{,}33\,H_3 + 92{,}25\,H_4\,.$$

Le sintesi quarte *) di questi dieci valori, uguagliate alle corrispondenti sintesi delle U_x, danno il sistema lineare di quattro equazioni con quattro incognite :

$$\begin{cases} H_1 - 3{,}3\,H_2 + 12{,}73\,H_3 - 51{,}62\,H_4 = 1516{,}56 \\ H_1 - 1{,}1\,H_2 + 3{,}93\,H_3 - 9{,}0357\,H_4 = 1267{,}91 \\ H_1 + 1{,}1\,H_2 + 3{,}93\,H_3 + 9{,}0357\,H_4 = 1020{,}09 \\ H_1 + 3{,}3\,H_2 + 12{,}73\,H_3 + 51{,}62\,H_4 = 737{,}04 \end{cases}$$

dal quale si ricava :

$$H_1 = 1151{,}7 \qquad H_2 = -101{,}5 \qquad H_3 = -1{,}96 \qquad H_4 = -1{,}35\,.$$

Dall'espressione generale dell'integrale definito della $F(x)$ fra x ed $x + h$:

---

*) § 6 e § 5 Parte IIª.

$$\int_x^{x+h} F(x)\,dx = H_1\{(x+h) - x\} + \frac{H_2}{2}\{(x+h)^2 - x^2\} +$$

$$+\frac{H_3}{3}\{(x+h)^3 - x^3\} + \frac{H_4}{4}\{(x+h)^4 - x^4\}$$, possiamo ora ottenere il numero delle persone di età compresa fra x ed $x+h$.

Eseguendo il calcolo per gli intervalli annuali di età fra 37 e 57 anni, otteniamo i valori riportati nella colonna 2) del seguente prospetto, nel quale confrontiamo questi risultati con i dati del censimento del 1871 e con i risultati che si ottengono applicando il metodo delle aree del Cantelli *).

Classi di età $x-(x+1)$	y calcolate (Metodo delle sint)	y osservate	y calc. — y oss. met. sint. Δ_1	y calcolate (met. delle aree)	y calc. — y oss. met. aree Δ_2
1)	2)	3)	4)	5)	6)
37–38	343	309	+ 34	351	+ 42
38–39	336	340	— 4	345	+ 5
39–40	329	234	+ 95	339	+ 105
*40–41	323	634	— 311	332	— 302
41–42	316	237	+ 79	326	+ 89
42–43	310	307	+ 3	319	+ 12
43–44	304	242	+ 62	313	+ 71
44–45	297	274	+ 23	306	+ 32
*45–46	291	367	— 76	300	— 67
46–47	285	280	+ 5	293	+ 13
47–48	278	248	+ 30	286	+ 38
48–49	272	281	— 9	280	— 1
49–50	265	207	+ 58	273	+ 66
*50–51	258	565	— 307	266	— 299
51–52	251	198	+ 53	259	+ 61
52–53	244	242	+ 2	251	+ 9
53–54	236	179	+ 57	244	+ 65
54–55	229	198	+ 31	236	+ 38
*55–56	220	228	— 8	228	0
56–57	212	197	+ 15	220	+ 23
			$\Sigma\|\Delta_1\| = 1262$		$\Sigma\|\Delta_2\| = 1338$

*) F. P. CANTELLI, *Sull'adattamento di curve ad una serie di misure o di operazioni*, Roma, 1905.

Quest'esempio mostra che, anche nel caso che vi sia ragione di preferenza per la scelta di una particolare suddivisione nel campo di variabilità della x, applicando il metodo delle sintesi si ottengono dei risultati più soddisfacenti di quelli ottenuti col metodo delle aree. Dalla precedente tabella si constata, infatti, che la somma dei valori assoluti degli scostamenti dei valori osservati dai valori calcolati col metodo delle sintesi, è più piccola della analoga somma che si ottiene quando l'interpolazione è eseguita col metodo delle aree; inoltre, solo in corrispondenza alle età rotonde, (segnate con l'asterisco) i $|\Delta_1|$ hanno un valore superiore ai corrispondenti $|\Delta_2|$, mentre, in corrispondenza alle altre sedici classi di età, succede l'opposto. Si ha, cioè, che la serie 2) si avvicina di più a quei dati che si ha ragione di ritenere più corrispondenti al vero e rappresenta, perciò, meglio della serie 5), la distribuzione per età della popolazione.

Risultati anche migliori di questi, relativamente a quelli che si otterrebbero col metodo delle aree, si otterrebbero applicando il nuovo metodo, quando non vi sia ragione di preferenza per una particolare suddivisione del campo di variabilità della x.

ESEMPIO II. — Si consideri la somma del risparmio italiano alla fine di ciascun anno, dal 1881 al 1911.

Anni x	Ammontare del risparmio (milioni); y	Anni x	Ammontare del risparmio (milioni); y	Anni x	Ammontare del risparmio (milioni); y	Anni x	Ammontare del risparmio (milioni); x
1881	979	1889	1.757	1897	2.199	1905	3.491
1882	1.041	1890	1.792	1898	2.273	1906	3.862
1883	1.151	1891	1.820	1899	2.400	1907	4.350
1884	1.304	1892	1.928	1900	2.512	1908	4.703
1885	1.420	1893	1.977	1901	2.621	1909	5.043
1886	1.602	1894	1.975	1902	2.790	1910	5.369
1887	1.662	1895	2.072	1903	2.990	1911	5.588
1888	1.726	1896	2.109	1904	3.257	—	—

Supposto che l'andamento del fenomeno possa essere rappresentato con buona approssimazione mediante una funzione della forma: $y = H_1 + H_2 x + H_3 x^2 + H_4 x^3$, determiniamone i parametri con questo nuovo metodo, per mettere in evidenza la semplicità dei calcoli occorrenti.

Posto nelle : 19), 20), 21), 22) del § 6 [P. I^a], $n = 31$, si ha :

$$H_1 = 4\ Y_1 - 6\ Y_2 + 4\ Y_3 - Y_4$$

$$H_2 = \frac{-6338\ Y_1 + 14.628\ Y_2 - 11364\ Y_3 + 3074\ Y_4}{7180,8}$$

$$H_3 = \frac{129\ Y_1 - 353\ Y_2 + 319\ Y_3 - 95\ Y_4}{2393,6}$$

$$H_4 = \frac{-Y_1 + 3\ Y_2 - 3\ Y_3 + Y_4}{1025,8285}$$

Calcolando le Y col metodo esposto nella Parte II° § 5, si ottiene :

$$Y_1 = \frac{46.622.835}{31.465} = 1481,7363 \qquad Y_2 = \frac{63.364.623}{31.465} = 2013,8129$$

$$Y_3 = \frac{86.152.507}{31.465} = 2738,0424 \qquad Y_4 = \frac{127.697.795}{31.465} = 4058,4075$$

Sostituendo questi valori nelle formule precedenti, si ottiene :

$$H_1 = +737,83 \qquad H_2 = +198,75709$$

$$H_3 = -13,3042 \qquad H_4 = +0,39381$$

e l'equazione della parabola interpolatrice è :

$y = +737,83 + 198,75709\ x - 13,3042\ x^2 + 0,39381\ x^3$, che è quella stessa che si ottiene, con calcoli molto più laboriosi, applicando il metodo dei minimi quadrati.

ESEMPIO III°. — Consideriamo la distribuzione dei redditi delle persone fisiche in Sassonia, nell'anno 1908 :

Reddito (x) in marchi	Numero (y) delle persone con reddito superiore ad x
—	—
2.200	175.842
8.300	22.792
26.000	4.822
54.000	1.634
100.000	572
250.000	101
500.000	17
1.000.000	2

Poichè la relazione tra y ed x è espressa abbastanza bene da una formula del tipo :
$y = H_1 x^{-H_2}$ 1) a partire da un certo reddito minimo > di zero, adottiamo come formula interpolatrice la : $y = \frac{y}{\log y} \log H_1 - \frac{y}{\log y} \times (\log x) H_2$ 2) che si ricava dalla 1).

I valori teorici, corrispondenti agli otto valori dati di y, sono :

$$\begin{aligned}
33.524{,}88 \log H_1 &- 112.054{,}22 H_2 \\
5.230{,}19 \log H_1 &- 20.497{,}53 H_2 \\
1.309{,}18 \log H_1 &- 5.779.99 H_2 \\
508.52 \log H_1 &- 2.406{,}51 H_2 \qquad 3) \\
207{,}44 \log H_1 &- 1.037{,}20 H_2 \\
50{,}39 \log H_1 &- 272{,}00 H_2 \\
13{,}82 \log H_1 &- 78{,}75 H_2 \\
6{,}64 \log H_1 &- 39{,}84 H_2
\end{aligned}$$

uguagliando le sintesi seconde dei valori y, alle sintesi seconde dei valori 3) si ottiene il sistema :

$$\begin{cases} 46.674{,}2333 = 9.179{,}0730 \log H_1 - 31.654{,}168 H_2 \\ 1.141{,}5666 = 352{,}8407 \log H_1 - 1.517{,}907 H_2 \end{cases}$$

risolvendo il quale, si ottiene :

$$\log H_1 = 10{,}26771 \qquad H_2 = 1{,}5029206$$

La funzione interpolatrice è :

$y = (1{,}8522944)\ 10^{10} \cdot x^{-1{,}5029206}$, dalla quale si ottengono i valori y interpolati, corrispondenti ai valori y dati :

Valori interpolati	Valori osservati
$y_1 = 175.515$	175.842
$y_2 = 23.857$	22.792
$y_3 = 4.289$	4.822
$y_4 = 1.430$	1.634
$y_5 = 561$	572
$y_6 = 143$	101
$y_7 = 50$	17
$y_8 = 18$	2

che sono abbastanza prossimi ai valori osservati.

§ 8. NOTA. — Dimostriamo, ora, che quando la funzione interpolatrice y si può porre sotto la forma:

$$\varphi(y) = H_1 + H_2 x + H_3 x^2 + H_4 x^3 + \ldots . \quad (H \neq 0) \qquad 1)$$

e quando, inoltre, i valori della variabile sono in progressione aritmetica, i parametri H, calcolati col metodo delle sintesi, sono gli stessi di quelli ottenuti applicando il metodo dei minimi quadrati.

Cominciamo dal caso dell'interpolazione con una retta:

$$y = H_1 + H_2 x . \qquad 2)$$

Col metodo delle sintesi si ha: [14) e 15) P. I],

$$H_1 = 2\,Y_1 - Y_2 \qquad 3') \qquad H_2 = \frac{3\,(Y_2 - Y_1)}{n+1} \qquad 3'') \qquad 3)$$

Col metodo dei minimi quadrati si ha invece:

$$H_1 = \frac{2}{n\,(n-1)}\left[(2n+1)\,[y_i] - 3\,[x_i\,y_i]\right] \qquad 4')$$

$$H_2 = \frac{3}{n+1}\left[\frac{2\,(n+1)\,[y_i] - 4\,[x_i\,y_i]}{n\,(1-n)}\right]. \qquad 4'') \qquad 4)$$

Dimostriamo prima che il secondo membro della 3') è uguale al secondo membro della 4').

Ricordando le 4) P. II[a], si ha:

$$2\,Y_1 - Y_2 = \left\{ \begin{array}{l} \dfrac{2\,(n-1)\,y_1 + 2\,(n-2)\,y_2 + 2\,(n-3)\,y_3 + \ldots 2\,y_{n-1}}{n\,(n-1)/2} + \\ + \dfrac{-(n-1)\,y_n - (n-2)\,y_{n-1} - \ldots - 3\,y_4 - 2\,y_3 - y_2}{n\,(n-1)/2} \end{array} \right.$$

ossia:

$$2\,Y_1 - Y_2 = \left\{ \begin{array}{l} \dfrac{(2n-2)\,y_1 + (2n-5)\,y_2 + (2n-8)\,y_3 + \ldots .}{n\,(n-1)/2} + \\ + \dfrac{(2n-3n+4)\,y_{n-1} + (2n-3n+1)\,y_n}{n\,(n-1)/2} \end{array} \right. \qquad 5)$$

Il secondo membro della 4') è :

$$\left\{ \frac{(2n+1)\,y_1 - 3\,y_1 + (2n+1)\,y_2 - 3\cdot 2\,y_2 + (2n+1)\,y_3}{n\,(n-1)/2} + \right.$$

$$\left. + \frac{-3\cdot 3\,y_3 - \ldots\ldots (2n+1)\,y_n - 3\cdot n\,y_n}{n\,(n-1)/2} \right. \qquad \text{ossia :}$$

$$\frac{(2n-2)\,y_1 + (2n-5)\,y_2 + (2n-8)\,y_3 + \ldots (2n-3n+1)\,y_n}{n\,(n-1)/2} \qquad 6)$$

che è uguale al 2° membro della 5).

Dimostriamo ora che il secondo membro della 3'') è uguale al secondo membro della 4'').

Infatti, ricordando le 4) [P. II^a], si ha :

$$\frac{3}{n+1}(Y_2 - Y_1) = \frac{3}{n+1}\left\{ \left\{ \frac{(n-1)\,y_n + (n-2)\,y_{n-1} + \ldots\ldots y_2}{n\,(n-1)/2} + \right.\right.$$

$$\left.\left. + \frac{-(n-1)\,y_1 - (n-2)\,y_2 \ldots\ldots - y_{n-1}}{n\,(n-1)/2} \right\} \right.$$

ossia, supposto n pari :

$$\frac{3}{n+1}(Y_2 - Y_1) = \frac{3}{n+1}\left\{ \frac{(n-1)\,(y_n - y_1) + (n-3)\,(y_{n-1} - y_2) + \ldots \left(y_{\frac{n}{2}+1} - y_{\frac{n}{2}}\right)}{n\,(n-1)/2} \right\} \qquad 7)$$

D'altra parte, il secondo membro della 4'') è :

$$\frac{3}{n+1}\left\{ \frac{(n-1)\,y_1 + (n-3)\,y_2 + (n-5)\,y_3 + \ldots - (n-3)\,y_{n-1} - (n-1)\,y_n}{n\,(n-1)/2} \right\}$$

ossia, supposto n pari, uguale a :

$$\frac{3}{n+1}\left\{ \frac{(n-1)\,(y_n - y_1) + (n-3)\,(y_{n-1} - y_2) + \ldots \left(y_{\frac{n}{2}+1} - y_{\frac{n}{2}}\right)}{n\,(n-1)/2} \right\} \qquad 8)$$

che è uguale al secondo membro della 7)

Passiamo ora al caso della parabola di secondo ordine :$y = H_1 + H_2\,x + H_3\,x^2$ e dimostriamo, prima che la 18) [P. I^a], è uguale alla 9')

di questa Nota. Infatti poichè :

$$-10\,Y_1 = \frac{60}{n(n-1)(n-2)}\left\{-\frac{(n-1)(n-2)}{2}y_1 - \frac{(n-2)(n-3)}{2}y_2 - \right.$$

$$-\frac{(n-3)(n-4)}{2}y_3 - \frac{(n-4)(n-5)}{2}y_4 - \ldots - \frac{4\cdot 3}{2}y_{n-4} -$$

$$\left. -\frac{3\cdot 2}{2}y_{n-3} - \frac{2\cdot 1}{2}y_{n-2}\right\}$$

$$-10\,Y_3 = \frac{60}{n(n-1)(n-2)}\left\{-\frac{2\cdot 1}{2}y_3 - \frac{3\times 2}{2}y^4 - \ldots - \right.$$

$$-\frac{(n-4)(n-5)}{2}y_{n-3} - \frac{(n-3)(n-4)}{2}y_{n-2} -$$

$$\left. -\frac{(n-2)(n-3)}{2}y_{n-1} - \frac{(n-1)(n-2)}{2}y_n\right\}$$

$$+20\,Y_2 = \frac{60}{n(n-1)(n-2)}\left\{+2\cdot(n-2)\cdot 1\cdot y_2 + 2(n-3)\cdot 2\,y_3 + \right.$$

$$+2(n-4)\cdot 3\,y_4 + \ldots + 2\cdot 3(n-4)\,y_{n-3} +$$

$$\left. +2\cdot 2\cdot(n-3)\,y_{n-2} + 2\cdot 1(n-2)\,y_{n-1}\right\}$$

sommando membro a membro, e dividendo poi ambo i membri per $-(n+1)(n+2)$ si ha :

$$H_3 = \frac{60}{-(n-2)(n-1)n\cdot(n+1)(n+2)} \times \sum_{i=1}^{i=n}\left\{-\frac{(n-i)(n-i-1)}{2} - \right.$$

$$\left. -\frac{(i-1)(i-2)}{2} + 2(n-i)(i-1)\right\}\cdot y_i \quad \text{ossia :}$$

$$H_3 = \frac{60}{-(n-2)(n-1)n(n+1)(n+2)} \times \sum_{i=1}^{i=n}\left\{-3i^2 + \right.$$

$$\left. +3(n+1)i - \frac{(n+1)(n+2)}{2}\right\}\cdot y_i \,. \qquad 9)$$

Col metodo dei minimi quadrati, l'espressione di H_3 è :

$$H_3 = \frac{60}{-(n-2)(n-1)n(n+1)(n+2)} \left\{ -3[y_i x_i^2] + \right.$$

$$\left. + 3(n+1)[y_i x_i] - \frac{(n+1)(n+2)}{2}[y_i] \right\}, \quad \text{ossia :} \qquad 9')$$

$$H_3 = \frac{60}{-(n-2)(n-1)n(n+1)(n+2)} \sum_{i=1}^{i=n} \left\{ -3i^2 + \right.$$

$$\left. + 3(n+1)i - \frac{(n+2)(n+1)}{2} \right\} y_i \quad \text{che è uguale alla 9).}$$

Dimostriamo ora che la 17), [P. I^a], è uguale alla 10') di questa Nota. Infatti poichè :

$$(12n+14)\,Y_1 = \frac{3}{-n(n-1)(n-2)} \Big[(12n+14)(n-1)(n-2)\,y_1 +$$

$$+ (12n+14)(n-2)(n-3)\,y_2 + (12n+14)(n-4)(n-5)\,y_3 +$$

$$+ (12n+14)(n-4)(n-5)\,y_4 + \ldots (12n+14)\,2\,.\,1\,.\,y_{n-2} \Big]$$

$$(8n+6)\,Y_3 = \frac{3}{-n(n-1)(n-2)} \Big[(8n+6)\,.\,1\,.\,2\,.\,y_3 +$$

$$+ (8n+6)\,.\,3\,.\,2\,.\,y_4 + \ldots (8n+6)(n-4)(n-3)\,y_{n-2} +$$

$$+ (8n+6)(n-3)(n-2)\,y_{n-1} + (8n+6)(n-2)(n-1)\,y_n \Big]$$

$$-(20n+20)\,Y_2 = \frac{3}{-n(n-1)(n-2)} \Big[-(40n+40)(n-2)\,.\,y_2 -$$

$$- (40n+40)(n-3)\,.\,2\,y_3 - (40n+40)(n-4)\,.\,3\,y_4 - \ldots -$$

$$- (40n+40)(n-2)\,y_{n-1} \Big]$$

sommando membro a membro e dividendo poi ambo i membri per $(n+2)\,(n+1)$ si ha :

$$H_2 = \frac{3}{(n+2)(n+1)n(n-1)(n-2)} \sum_{i=1}^{i=n} \Big\{ -(12n+14)(n-i)(n-i-1) -$$

$$- (8n+6)(i-1)(i-2) + (40n+40)(n-i)(i-1) \Big\} y_i \quad \text{ossia :}$$

$$H_2 = \frac{3}{(n+2)(n+1)n(n-1)(n-2)} \sum_{i=1}^{i=n} \Big\{ -6(12n+1)(n+2)(n+1) + \\ + 4(8n+11)(2n+1)\,i - 60(n+1)\,i^2 \Big\} y_i \qquad 10)$$

Col metodo dei minimi quadrati si ha invece:

$$H_2 = \frac{3}{(n+2)(n+1)n(n-1)(n-2)} \Big[-6(2n+1)(n+2)(n+1)[y_i] + \\ + 4(8n+11)(2n+1)[y_i x_i] - 60(n+1)[y_i x_i^2] \Big] \quad \text{ossia:} \qquad 10')$$

$$H_2 = \frac{3}{(n+2)(n+1)n(n-1)(n-2)} \sum_{i=1}^{i=n} \Big\{ -6(12n+1)(n+2)(n+1) + \\ + 4(8n+11)(2n+1)\,i - 60(n+1)\,i^2 \Big\} y_i \quad \text{che è uguale alla 10).}$$

Dimostriamo infine che la 16) [P. I^a], è uguale alla 11') di questa Nota. Infatti, poichè:

$$3\,Y_1 = \frac{3}{n(n-1)(n-2)} \Big\{ 3(n-1)(n-2)\,y_1 + 3(n-2)(n-3)\,y_2 + \\ + 3(n-3)(n-4)\,y_3 + 3(n-4)(n-5)\,y_4 + \ldots + \\ + 3\cdot 4\cdot 3\,y_{n-4} + 3\cdot 3\cdot 2\,y_{n-3} + 3\cdot 2\cdot 1\cdot y_{n-2} \Big\}$$

$$Y_3 = \frac{3}{n(n-1)(n-2)} \Big\{ 2\cdot 1\cdot y_3 + 2\cdot 3\cdot y_4 + \ldots + (n-4)(n-5)\,y_{n-3} + \\ + (n-3)(n-4)\,y_{n-2} + (n-3)(n-2)\,y_{n-1} + (n-1)(n-2)\,y_n \Big\}$$

$$-3\,Y_2 = \frac{3}{n(n-1)(n-2)} \Big\{ -6\cdot(n-2)\,y_2 - 6(n-3)\cdot 2\,y_3 - \\ - 6(n-4)\cdot 3\cdot y_4 - \ldots - 6(n-4)\,3\cdot y_{n-3} - \\ - 6(n-3)\cdot 2\,y_{n-2} - 6(n-2)\,y_{n-1} \Big\}$$

sommando membro a membro avremo :

$$H_1 = \frac{3}{n\,(n-1)\,.\,(n-2)} \sum_{i=1}^{i=n} \Big\{ 3\,(n-i)\,(n-i+1) + (i-1)\,(i-2) - \\ - 6\,(n-i)\,(i-1) \Big\}\, y_i = \frac{3}{n\,(n-1)\,(n-2)} \sum_{i=1}^{i=n} \Big\{ 10\,i^2 - 6\,i - \\ - 12\,n\,i + 3\,n^2 + 3\,n + 2 \Big\}\, y_i \,. \qquad 11)$$

Col metodo dei minimi quadrati si ha invece :

$$H_1 = \frac{3}{n\,(n-1)\,(n-2)} \Big[10\,[y_i\,x_i^2] - 6\,(2\,n+1)\,[y_i\,x_i] + \\ + (3\,n^2 + 3\,n + 2)\,[y_i] \Big] = \frac{3}{n\,(n-1)\,(n-2)} \sum_{i=1}^{i=n} (10\,i^2 - 6\,i - \\ - 12\,n\,i + 3\,n^2 + 3\,n + 2)\,y_i \quad \text{che è identica alla 11).} \qquad 11')$$

Questa stessa dimostrazione si può estendere facilmente al caso che la funzione interpolatrice sia una curva parabolica di ordine n con tutti i coefficienti diversi da zero.

Se la funzione interpolatrice scelta è, poi, della forma :

$y^n = H_1 + H_2\,x + H_3\,x^2 + \ldots$. dove n è un qualunque numero reale, le dimostrazioni date per il caso di $n = 1$, possono estendersi al caso di n qualunque, sostituendo y^n ad y nelle espressioni che danno le H secondo i due metodi, ed anche in questo caso si trova che i parametri determinati con i due metodi sono identici, sempre però che i valori della variabile siano in progressione aritmetica.

Analoga dimostrazione si può fare se la funzione interpolatrice può porsi sotto la forma : $\varphi\,(y) = H_1 + H_2\,x + H_3\,x^2 + \ldots$.

PARTE II.

Le sintesi.

§ 1. — Data la definizione di sintesi K^{me} (Parte I^a § 2) degli n elementi :

$$y_1\,,\ y_2\,,\ y_3\,,\ \ldots\,.\ y_n\,, \qquad 1)$$

diamone, ora, l'espressione algebrica in funzione di quegli elementi 1) che le costituiscono.

Cominciamo dal caso più semplice di $K = 2$; applicando la data definizione, si ha che il numeratore della : $\frac{\sum_{i=1}^{i=h} s_{1,i}}{\binom{n}{2}}$ è la somma di $h = n - 1$ valori s, cioè si ha che :

$$\frac{\sum_{i=1}^{i=n-1} s_{1,i}}{\binom{n}{2}} = \frac{y_1 + (y_1 + y_2) + (y_1 + y_2 + y_3) + \ldots (y_1 + y_2 + y_3 + \ldots y_{n-1})}{\binom{n}{2}} \quad 2)$$

e, analogamente, che :

$$\frac{\sum_{i=1}^{i=n-1} s_{2,i}}{\binom{n}{2}} = \frac{y_n + (y_n + y_{n-1}) + (y_n + y_{n-1} + y_{n-2}) + \ldots (y_n + y_{n-1} + \ldots y_2)}{\binom{n}{2}} \quad 3)$$

che si possono anche scrivere sotto la forma :

$$Y_1 = \frac{(n-1)\, y_1 + (n-2)\, y_2 + \ldots . y_{n-1}}{\binom{n}{2}}$$

$$Y_2 = \frac{(n-1)\, y_n + (n-2)\, y_{n-1} + \ldots . y_2}{\binom{n}{2}} . \quad 4)$$

Osservando che : $(n-1) + (n-2) + \ldots . 1 = \binom{n}{2}$, le sintesi seconde ci appaiono come medie ponderate dei valori

$$y_1 , \; y_2 , \; \ldots . \; y_{n-1}$$

e rispettivamente dei valori :

$$y_n , \; y_{n-1} , \; \ldots . \; y_2$$

quando siano :

$$(n-1) , \; (n-2) , \; \ldots . \; 1$$

i pesi corrispondenti.

Se ora si osserva che le 2) e 3) sono le somme delle somme succes-

sive prime degli $n-1$ elementi: $y_1, y_2, \ldots y_{n-1}$ e rispettivamente degli $n-1$ elmenti: $y_n, y_{n-1}, \ldots y_2$, e se indichiamo per semplicità, queste somme con i simboli:

$$ {}_2S\, y_1, y_2, \ldots y_{n-1} \quad ; \quad {}_2S\, y_n, y_{n-1}, \ldots y_2 \qquad 5) $$

si ha:

$$ Y_1 = \frac{{}_2S\, y_1, y_2, \ldots y_{n-1}}{\binom{n}{2}} \quad ; \quad Y_2 = \frac{{}_2S\, y_n, y_{n-1}, \ldots y_2}{\binom{n}{2}} \; ; \qquad 6) $$

formule utili per calcolare le sintesi seconde degli n valori noti.

Quando, poi, gli n valori 1) sono i primi n numeri naturali, le sintesi seconde assumono la forma semplicissima:

$$ X_1 = \frac{n+1}{3} \qquad X_2 = \frac{2(n+1)}{3} \qquad 7) $$

formule che permettono il calcolo immediato delle sintesi teoriche seconde, quando la funzione interpolatrice scelta è la:

$$ y = H_1 + H_2 x \; (H_1 \text{ e } H_2 \neq 0) \, . $$

§ 2. — Passiamo ora al caso di $K = 3$. Avremo:

$$ \sum_{i=1}^{i=h} s_{1,i} = C_1 y_1 + C_2 (y_1 + y_2) + C_3 (y_1 + y_2 + y_3) + \ldots . $$

$$ C_{n-2} (y_1 + y_2 + \ldots . y_{n-2}) \qquad 8) $$

dove: C_1 è uguale al numero dei valori s che costituiscono ciascuna delle sintesi seconde degli $n-1$ elementi: $y_2, y_3, y_4, \ldots . y_n$, cioè: $C_1 = n-2$, analogamente: C_2 è uguale al numero dei valori s che costituiscono ciascuna delle sintesi seconde degli $n-2$ elementi: $y_3, y_n, \ldots . y_n$, cioè: $C_2 = n-3$, e così via, fino a C_{n-2} che è uguale ad 1.

Avremo quindi che: $\sum_{i=1}^{i=h} s_{1,i}$ è la somma di

$$ h = (n-2) + (n-3) + \ldots . 1 = \frac{(n-1)(n-2)}{2} = \binom{n-1}{n-3} \qquad 9) $$

valori s, cioè :

$$\sum_{i=1}^{i=\frac{(n-1)(n-2)}{2}} s_{1,i} = (n-2)\, y_1 + (n-3)\,(y_1 + y_2) +$$

$$+ (n-4)\,(y_1 + y_2 + y_3) + \ldots . 1 \,.\, (y_1 + y_2 + y_3 + \ldots . y_{n-2}), \quad 10)$$

analogamente :

$$\sum_{i=1}^{i=\frac{(n-1)(n-2)}{2}} s_{3,1} = (n-2)\, y_n + (n-3)\,(y_n + y_{n-1}) +$$

$$+ (n-4)\,(y_n + y_{n-1} + y_{n-2}) + \ldots 1\,(y_n + y_{n-1} + y_{n-2} + \ldots y_3), \quad 11)$$

e poichè il numero degli elementi 1) che ciascuna di queste somme contiene è :

$$(n-2) \,.\, 1 + (n-3) \,.\, 2 + (n-4) \,.\, 3 + \ldots + 2\,(n-3) +$$

$$+ 1 \,.\, (n-2) = \binom{n}{3} \qquad \text{si ha :}$$

$$Y_1 = \frac{\frac{(n-1)\,(n-2)}{2}\, y_1 + \frac{(n-2)\,(n-3)}{2}\, y_2 + \ldots . \frac{2\,.\,1}{2}\, y_{n-2}}{\binom{n}{3}} \quad 12)$$

$$Y_3 = \frac{\frac{(n-1)\,(n-2)}{2}\, y_n + \frac{(n-2)\,(n-3)}{2}\, y_{n-1} + \ldots . \frac{2\,.\,1}{2}\, y_3}{\binom{n}{3}} \quad 13)$$

Inoltre, poichè ogni elemento 1) compare nelle 3 sintesi terze complessivamente un numero $h = \frac{(n-1)\,(n-2)}{2}$ di volte, si ha :

$$Y_2 = \frac{(n-2)\, y_2 + 2\,(n-3)\, y_3 + 3\,(n-4)\, y_4 + \ldots . (n-3) \,.\, 2\, y_{n-2} + (n-2)\, y_{n-1}}{\binom{n}{3}} \quad 14)$$

Le 12), 13), 14) sono utili per il calcolo delle sintesi terze teoriche ; quando, poi, i valori 1) sono i primi n numeri naturali, queste formule assumono la forma semplicissima :

$$X_1 = \frac{n+1}{4} \quad ; \quad X_2 = \frac{2\,(n+1)}{4} \quad ; \quad X_3 = \frac{3\,(n+1)}{4} \qquad 15)$$

mentre le sintesi terze dei quadrati dei primi n numeri naturali assumono la forma :

$$X_1^{(2)} = \frac{(n+1)\left(n-\frac{1}{2}\right)}{10} \; ; \; X_2^{(2)} = \frac{(n+1)\,(3n+1)}{10} \; ;$$

$$X_3^{(2)} = \frac{(n+1)\left(6n+\frac{9}{2}\right)}{10} \qquad 16)$$

formule, queste, che rendono immediato il calcolo delle sintesi terze teoriche, quando la funzione interpolatrice scelta è la :

$$y = H_1 + H_2\,x + H_3\,x^2 \; , \; (H_1 \; , \; H_2 \; , \; H_3 \; , \neq 0)$$

Osservando ora che i numeratori delle 12) 13) e 14) si possono porre rispettivamente sotto la forma :

$$\begin{array}{l} 1\,y_1 + \\ 2\,y_1 + 1\,y_2 + \\ 3\,y_1 + 2\,y_2 + 1\,y_3 + \\ 4\,y_1 + 3\,y_2 + 2\,y_3 + y_4 \\ \cdots\cdots\cdots\cdots\cdots \\ (n-2)\,y_1 + (n-3)\,y_2 + (n-4)\,y_3 + (n-5)\,y_4 + \ldots . y_{n-2} \end{array} \qquad 17)$$

$$\begin{array}{l} 1\,y_n + \\ 2\,y_n + 1\,y_{n-1} + \\ 3\,y_n + 2\,y_{n-1} + 1\,y_{n-2} + \\ 4\,y_n + 3\,y_{n-1} + 2\,y_{n-2} + 1\,y_{n-3} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ (n-2)\,(y_n + (n-3)\,y_{n-1} + (n-4)\,y_{n-2} + (n-5)\,y_{n-3} + \ldots . y_3 \end{array} \qquad 18)$$

$$\begin{array}{l} y_2 + \\ y_2 + 2\,y_3 + \\ y_2 + 2\,y_3 + 3\,y_4 + \\ y_2 + 2\,y_3 + 3\,y_4 + 4\,y_5 + \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ y_2 + 2\,y_3 + 3\,y_4 + 4\,y_5 + \ldots . (n-2)\,y_{n-1} \end{array} \qquad 19)$$

che sono rispettivamente, le due prime, le somme delle somme successive seconde degli $n-2$ elementi :

$y_1 , y_2 , y_3 , \ldots . y_{n-2}$ e degli $n-2$ elementi :

$y_n , y_{n-1} , y_{n-2} , \ldots . y_3$, e la terza è la somma delle somme successive prime degli $n-2$ prodotti :

$1 . y_2 , 2\, y_3 , 3\, y_4 , \ldots . , (n-2)\, y_{n-1}$, somme che indicheremo rispettivamente con i simboli :

$${}_3S\, y_1 , y_2 , \ldots y_{n-2} \; ; \; {}_3S\, y_n , y_{n-1} , \ldots y_3 \; ; \; {}_2S\, y_2 , 2\, y_3 , 3\, y_4 , \ldots (n-2)\, y_{n-1} .$$

Potremo allora scrivere :

$$Y_1 = \frac{{}_3S\, y_1 , y_2 , \ldots . y_{n-2}}{\binom{n}{3}} \; ; \; Y_3 = \frac{{}_3S\, y_n , y_{n-1} , \ldots . y_3}{\binom{n}{3}} \; ;$$

$$Y_2 = \frac{{}_2S\, y_2 , 2\, y_3 , 3\, y_4 , \ldots . (n-2)\, y_{n-1}}{\binom{n}{3}} \qquad 20)$$

formule che rendono facile il calcolo delle sintesi terze dei valori noti.

Giova osservare che quando si siano calcolate Y_1 ed Y_3, non è necessario calcolare direttamente la Y_2, perchè essendo :

$$Y_1 + Y_2 + Y_3 = \frac{h \sum_{i=1}^{i=n} y_i}{\binom{n}{3}} = \frac{\frac{(n-1)\,(n-2)}{2} \sum_{i=1}^{i=n} y_i}{\binom{n}{3}} = \frac{3 \sum_{i=1}^{i=n} y_i}{n} \; , \qquad 21)$$

cioè, essendo la somma delle sintesi terze uguale al triplo della media aritmetica degli n valori dati, si avrà :

$$Y_2 = 3\, M - Y_1 - Y_2 \qquad 22)$$

dove con M si è indicata la media aritmetica dei valori 1).

§ 3. — Ricordando ora l'espressione generale delle somme successive K^{me} (K intero $\leq n$) dei primi n numeri naturali :

$$\binom{K+1}{0} , \binom{K+2}{1} , \binom{K+3}{2} , \ldots \binom{K+n-1}{n-2} , \binom{K+n}{n-1} \qquad 23) ,$$

possiamo generalizzare le formule 20), al caso delle sintesi h^{me} (h intero $\leq n$; $n \neq 1$) .

Supposto, ad esempio, h dispari, le h sintesi h^{me} si possono porre sotto la forma :

$$Y_1 = \frac{{}_hS\, y_1 , y_2 , \ldots y_{n-h+1}}{\binom{n}{h}} \; ; \; Y_h = \frac{{}_hS\, y_n , y_{n-1} , \ldots . y_h}{\binom{n}{h}} \qquad 24)$$

$$Y_2 = \frac{{}_{h-1}S\, y_2 , 2\, y_3 , 3\, y_4 , \ldots . (n-h+1)\, y_{n-h+2}}{\binom{n}{h}} \; ;$$

$$Y_{h-1} = \frac{{}_{h-1}S\, y_{n-1} , 2\, y_{n-2} , 3\, y_{n-3} , \ldots . (n-h+1)\, y_{h-1}}{\binom{n}{h}} \qquad 25)$$

$$Y_3 = \frac{{}_{h-2}S\, y_3 , 3\, y_4 , 6\, y_5 , \ldots . \binom{n-h+2}{n-h}\, x_{n-h+3}}{\binom{n}{h}} \; ;$$

$$Y_{h-2} = \frac{{}_{h-2}S\, y_{n-2} , 3\, y_{n-3} , 6\, y_{n-4} , \ldots . \binom{n-h+2}{n-h}\, y_{h-2}}{\binom{n}{h}} \qquad 26)$$

. . . . . .

$$Y_{\frac{h+1}{2}} = \frac{{}_{\frac{h+1}{2}}S \binom{h-1}{2}{0} \cdot y_{\frac{h+1}{2}} , \binom{h+1}{2}{1} \cdot y_{\frac{h+3}{2}} , \ldots \binom{2n-h-1}{2}{n-h} \cdot y_{n+\frac{1-h}{2}}}{\binom{n}{h}} \qquad 27)$$

nella quale ultima formula :

${}_{\frac{h+1}{2}}S$ indica la somma delle somme successive $\binom{h-1}{2}^{me}$ degli elementi :

$y_{\frac{h+1}{2}}$, $y_{\frac{h+3}{2}}$, ... $y_{n+\frac{1-h}{2}}$, moltiplicati rispettivamente per i numeri : $\binom{h-1}{2}{0}$, $\binom{h+1}{2}{1}$, ... $\binom{2n-h-1}{2}{n-h}$ che

sono le somme successive $\left(\frac{h-3}{2}\right)^{me}$ dei primi $n-h+1$ numeri naturali.

§ 4. — Ponendo in queste ultime $h=3$, si ritrovano le già note formule delle sintesi terze; dimostriamo ora che ponendo in esse $h=4$ si ottengono proprio le formule delle sintesi quarte degli n valori dati, cioè che:

$$\frac{{}_4S\, y_1, y_2, \ldots y_{n-3}}{\binom{n}{4}} = Y_1 \quad 28) \qquad \frac{{}_4S\, y_n, y_{n-1}, \ldots . y_4}{\binom{n}{4}} = Y_4 \quad 29)$$

$$\frac{{}_3S\, y_2, 2\,y_3, \ldots (n-3)\, y_{n-2}}{\binom{n}{4}} = Y_2 \qquad 30)$$

$$\frac{{}_3S\, y_{n-1}, 2\,y_{n-2}, \ldots (n-3)\, y_3}{\binom{n}{4}} = Y_3 . \qquad 31)$$

Dimostriamo prima la 28). Dalla definizione data di sintesi K^{me} degli n elementi y, si ha:

$$Y_1 = \frac{\sum_{i=1}^{i=k} S_{1,i}}{\binom{n}{k}} = \frac{C_1 y_1 + C_2 (y_1 + y_2) + C_3 (y_1 + y_2 + y_3) + \ldots C_{n-3} (y_1 + y_2 + \ldots . y_{n-3})}{\binom{n}{k}} \qquad 32)$$

nella quale, se $K=4$, C_1 è uguale al numero dei gruppi s che costituiscono ciascuna delle tre sintesi terze dgli $n-1$ elementi: $y_2, y_3, \ldots y_n$ cioè: $C_1 = \frac{(n-2)(n-3)}{2} = \binom{n-2}{n-4}$; C_2 è uguale al numero dei gruppi s che costituiscono ciascuna delle sintesi terze degli $n-2$ elementi: $y_3, y_4, \ldots . y_n$, cioè:

$$C_2 = \frac{(n-3)(n-4)}{2} = \binom{n-3}{n-5}$$

e analogamente :

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

$$C_{n-4} = \frac{3 \cdot 2}{2} = \binom{3}{1}$$

$$C_{n-3} = \frac{2 \cdot 1}{2} = \binom{2}{0}.$$

I coefficienti $C_1, C_2, \ldots . C_{n-3}$, della 32) non sono altro, quindi, che le somme successive prime dei primi $(n-3)$ numeri naturali.

E poichè h è, in questo caso, uguale alla somma :

$$\binom{n-2}{n-4} + \binom{n-3}{n-5} + \ldots . \binom{3}{1} + \binom{2}{0} = \binom{n-1}{n-4},$$

potremo scrivere :

$$Y_1 = \frac{\sum_{i=1}^{i=\binom{n-1}{n-4}} s_{1,i}}{\binom{n}{4}} = \left\{ \frac{\left[1 + 3 + 6 + \ldots \binom{n-2}{n-4}\right] y_1 + \left[1 + 3 + 6 + \ldots \right.}{\binom{n}{4}} + \right.$$

$$\left. + \frac{\binom{n-3}{n-5}\right] y_2 + \ldots \left[\binom{n-\{n-3+1\}}{n-\{n-3+3\}}\right)\right] y_{n-3}}{\binom{n}{4}} \qquad 33)$$

ossia :

$$Y_1 = \frac{\sum_{i=1}^{i=\binom{n-1}{n-4}} s_{1,i}}{\binom{n}{4}} = \frac{\binom{n-1}{n-4} y_1 + \binom{n-2}{n-5} y_2 + \binom{n-3}{n-6} y_3 + \ldots \binom{3}{0} y_{n-3}}{\binom{n}{4}} \qquad 34)$$

formula utile per calcolare le sintesi teoriche.

Osservando che :

$\binom{n-1}{n-4} + \binom{n-2}{n-5} + \binom{n-3}{n-6} + \ldots \binom{3}{0} = \binom{n}{4}$, la Y_1 non è altro che la media ponderata degli elementi :

$$y_1, y_2, y_3, \ldots . y_{n-3}$$

quando i pesi rispettivi sono:

$$\binom{n-1}{n-4}, \binom{n-2}{n-5}, \binom{n-3}{n-6}, \ldots \binom{3}{0}.$$

Poichè il numeratore della 33) può anche scriversi sotto la forma:

$$\begin{aligned}
&1 \,.\, y_1 \\
&3 \,.\, y_1 + 1\, y_2 \\
&6 \,.\, y_1 + 3\, y_2 + 1\, y_3 \\
&10\, y_1 + 6\, y_2 + 3\, y_3 + y_4 \\
&\ldots\ldots\ldots\ldots\ldots
\end{aligned}$$

$$\binom{n-2}{n-4} y_1 + \binom{n-3}{n-5} y_2 + \binom{n-4}{n-6} y_3 + \ldots \binom{2}{0} y_{n-3}$$

che è la somma delle somme successive terze degli $n-3$ valori $y_1, y_2, y_3, \ldots y_{n-3}$, potremo, infine, scrivere:

$$Y_1 = \frac{\sum_{i=1}^{i=\binom{n-1}{n-4}} s_{1,i}}{\binom{n}{4}} = \frac{{}_4S y_1, y_2, \ldots y_{n-3}}{\binom{n}{4}} \quad \text{c. v. d.} \qquad 35)$$

Con dimostrazione analoga possiamo dimostrare che:

$$Y_4 = \frac{\sum_{i=1}^{i=\binom{n-1}{n-4}} s_{4,i}}{\binom{n}{4}} =$$

$$= \frac{\binom{n-1}{n-4} y_n + \binom{n-2}{n-5} y_{n-1} + \binom{n-3}{n-6} y_{n-2} + \ldots \binom{3}{0} y_4}{\binom{n}{4}} \qquad 36)$$

e che:

$$Y_4 = \frac{{}_4S\, y_n, y_{n-1}, \ldots y_4}{\binom{n}{4}} \qquad 37)$$

che :

$$Y_2 = \frac{\sum_{i=1}^{i=\binom{n-1}{n-4}} s_{2,i}}{\binom{n}{4}} =$$

$$= \frac{\binom{n-2}{n-4} y_2 + \binom{n-3}{n-5} 2\, y_3 + \binom{n-4}{n-6} 3 \cdot y_4 + \dots \binom{2}{0} (n-3)\, y_{n-2}}{\binom{n}{4}} \qquad 38)$$

e che :

$$Y_2 = \frac{{}_3S\, y_2 , 2\, y_3 , 3\, y_4 , \dots . (n-3)\, y_{n-2}}{\binom{n}{4}} \qquad 39)$$

e infine che :

$$Y_3 = \frac{\sum_{i=1}^{i=\binom{n-1}{n-4}} s_{3,i}}{\binom{n}{4}} =$$

$$= \frac{\binom{n-2}{n-4} y_{n-1} + \binom{n-3}{n-5} 2 \cdot y_{n-2} + \binom{n-4}{n-6} 3\, y_{n-3} + \dots \binom{2}{0} (n-3)\, y_3}{\binom{n}{4}} \qquad 40)$$

e che :

$$Y_3 = \frac{{}_3S\, y_{n-1} , 2\, y_{n-2} , \dots . (n-3)\, y_3}{\binom{n}{4}} \qquad 41)$$

Come si è già dimostrato per le sintesi terze, anche per le sintesi quarte, vale la relazione :

$$4\, M - Y_1 - Y_4 - Y_2 = Y_3 \qquad 42)$$

dove M è la media aritmetica degli n valori dati.

Quando gli n valori dati sono i primi n numeri naturali, le precedenti formule delle sintesi quarte, possono esprimersi in funzione

di n ed assumono la forma:

$$X_1 = \frac{n+1}{5} \qquad X_2 = \frac{2(n+1)}{5}$$
$$X_3 = \frac{3(n+1)}{5} \qquad X_4 = \frac{4(n+1)}{5}. \tag{43}$$

Se gli n valori sono i quadrati dei primi n numeri naturali, le sintesi quarte sono:

$$X_1^{(2)} = \frac{n^2-1}{15} \qquad X_2^{(2)} = \frac{3n^2+3n}{15}$$
$$X_{(3)}^{(2)} = \frac{6n^2+9n+3}{15} \qquad X_4^{(2)} = \frac{10n^2+18n+8}{15} \tag{44}$$

ed infine le sintesi quarte delle terze potenze dei primi n numeri naturali sono:

$$X_1^{(3)} = \frac{n^3-n^2-3n-1}{35} \; ; \; X_2^{(3)} = \frac{4n^3+3n^2-5n-4}{35}$$
$$X_3^{(3)} = \frac{10n^3+18n^2+5n-3}{35} \; ; \; X_4^{(3)} = \frac{20n^3+50n^2+38n+8}{35}. \tag{45}$$

Le 43) 44) 45) godono la proprietà d'avere costanti rispettivamente le differenze prime, seconde, terze; proprietà che facilita molto la risoluzione del sistema 12) [Parte I^a].

Formule analoghe a quelle trovate, si troverebbero per le sintesi quinte, seste, etc. Diamo, infine, qualche esempio pratico per il calcolo delle sintesi di n valori noti e dei corrispondenti valori teorici.

§ 5. — Esempio I. Determinare le sintesi quarte dei dieci valor

1765
1327
1429
1412
1163
1301
817
756
834
550

Applicando le 28), 29), 30), 31) del § 4, Parte IIª, ed indicando con s_1, s_2, s_3, le somme successive prime, seconde, terze, dei valori dati che si considerano si ha applicando la 28) :

s_1	s_2	s_3
1.765	1.765	1.765
3.092	4.857	6.622
4.521	9.378	16.000
5.933	15.311	31.311
7.096	22.407	53.718
8.397	30.804	84.522
9.214	40.018	124.540
		318.478 = ${}_4S_{1}$*)

e poichè : $Y_1 = \dfrac{{}_4S_1}{\binom{n}{4}}$, si ha : $Y_1 = 1516{,}5619$

Analogamente, applicando la 29) si ha :

s_1	s_2	s_3
550	550	500
1.384	1.934	2.484
2.140	4.074	6.558
2.957	7.031	13.589
4.258	11.289	24.878
5.421	16.710	41.588
6.833	23.543	65.131
		154.778 = ${}_4S_4$

e poichè : $Y_4 = \dfrac{{}_4S_4}{\binom{n}{4}}$, si ha : $Y_4 = 737{,}0381$

*) Per semplicità abbiamo indicato ${}_4S\, y_1, y_2, \ldots y_{n-3}$ con ${}_4S_1$.

e applicando la 30) si ottiene :

		S_1	S_2
1.327 × 1 =	1.327	1.327	1.327
1.429 × 2 =	2.858	4.185	5.512
1.412 × 3 =	4.236	8.421	13.933
1.163 × 4 =	4.652	13.073	27.006
1.301 × 5 =	6.505	19.578	46.584
817 × 6 =	4.902	24.480	71.064
756 × 7 =	5.292	29.772	100.836
			266.262 = $_3S_2$

e poichè : $Y_2 = \dfrac{{}_3S_2}{\binom{n}{4}}$, si ha : $Y_2 = 1267{,}9142$

ora, per calcolare Y_3, basta applicare la 42) P. II^a,

$$Y_3 = 4\,M - Y_1 - Y_2 - Y_4$$

dove M è la media aritmetica dei dieci valori dati ; sostituendo i numeri alle lettere si ottiene :

$Y_3 = 4541{,}6000 - 1516{,}5619 - 1267{,}9142 - 737{,}0381 = 1020{,}0858$.

§ 6. — Esempio II°. — Determinare le sintesi quarte dei dieci valori :

$$\begin{aligned}
&H_1 - 4{,}5\,H_2 + 20{,}33\,H_3 - 92{,}25\,H_4\\
&H_1 - 3{,}5\,H_2 + 12{,}33\,H_3 - 43{,}75\,H_4\\
&H_1 - 2{,}5\,H_2 + 6{,}33\,H_3 - 16{,}25\,H_4\\
&H_1 - 1{,}5\,H_2 + 2{,}33\,H_3 - 3{,}75\,H_4\\
&H_1 - 0{,}5\,H_2 + 0{,}33\,H_3 - 0{,}25\,H_4\\
&H_1 + 0{,}5\,H_2 + 0{,}33\,H_3 + 0{,}25\,H_4\\
&H_1 + 1{,}5\,H_2 + 2{,}33\,H_3 + 3{,}75\,H_4\\
&H_1 + 2{,}5\,H_2 + 6{,}33\,H_3 + 16{,}25\,H_4\\
&H_1 + 3{,}5\,H_2 + 12{,}33\,H_3 + 43{,}75\,H_4\\
&H_1 + 4{,}5\,H_2 + 20{,}33\,H_3 + 92{,}25\,H_4
\end{aligned}$$

In questo caso, per determinare le sintesi, conviene applicare le formule 34), 36), 38), 40) del § 4. P. II^a.

Applicando la 34) si ha :

$$
\begin{array}{rrrrrrrrrrrr}
84 \cdot [H_1 & - 4{,}5\, H_2 & + 20{,}33\, H_3 & - 92{,}25\, H_4] & = & 84\, H_1 & - 378\, H_2 & + 1708\, H_3 & - & 7749\, H_4 \\
56 \cdot [H_1 & - 3{,}5\, H_2 & + 12{,}33\, H_3 & - 43{,}75\, H_4] & = & 56\, H_1 & - 196\, H_2 & + 690{,}66\, H_3 & - & 2450\, H_4 \\
35 \cdot [H_1 & - 2{,}5\, H_2 & + 6{,}33\, H_3 & - 16{,}25\, H_4] & = & 35\, H_1 & - 87{,}5\, H_2 & + 221{,}66\, H_3 & - & 568{,}75\, H_4 \\
20 \cdot [H_1 & - 1{,}5\, H_2 & + 2{,}33\, H_3 & - 3{,}75\, H_4] & = & 20\, H_1 & - 30\, H_2 & + 46{,}66\, H_3 & - & 75\, H_4 \\
10 \cdot [H_1 & - 0{,}5\, H_2 & + 0{,}33\, H_3 & - 0{,}25\, H_4] & = & 10\, H_1 & - 5\, H_2 & + 3{,}33\, H_3 & - & 2{,}5\, H_4 \\
4 \cdot [H_1 & + 0{,}5\, H_2 & + 0{,}33\, H_3 & + 0{,}25\, H_4] & = & 4\, H_1 & + 2\, H_2 & + 1{,}33\, H_3 & + & 1\, H_4 \\
1 \cdot [H_1 & + 1{,}5\, H_2 & + 2{,}33\, H_3 & + 3{,}75\, H_4] & = & 1\, H_1 & + 1{,}5\, H_2 & + 2{,}33\, H_3 & + & 3{,}75\, H_4 \\
\hline
 & & & & & 210\, H_1 & - 693\, H_2 & + 2674\, H_3 & - & 10.840{,}5\, H_4 = {}_4S_1
\end{array}
$$

e poichè : $Y_1 = \dfrac{{}_4S_1}{\binom{n}{4}}$, si ha :

$$Y_1 = H_1 - 3{,}3\, H_2 + 12{,}73\, H_3 - 51{,}62\, H_4$$

analogamente, applicando la 36) si ha :

$$
\begin{array}{rrrrrrrrrrrr}
84\, [H_1 & + 4{,}5\, H_2 & + 20{,}33\, H_3 & + 92{,}25\, H_4] & = & 84\, H_1 & + 378\, H_2 & + 1708\, H_3 & + & 7749\, H_4 \\
56\, [H_1 & + 3{,}5\, H_2 & + 12{,}33\, H_3 & + 43{,}75\, H_4] & = & 56\, H_1 & + 196\, H_2 & + 690{,}66\, H_3 & + & 2450\, H_4 \\
35\, [H_1 & + 2{,}5\, H_2 & + 6{,}33\, H_3 & + 16{,}25\, H_4] & = & 35\, H_1 & + 87{,}5\, H_2 & + 221{,}66\, H_3 & + & 568{,}75\, H_4 \\
20\, [H_1 & + 1{,}5\, H_2 & + 2{,}33\, H_3 & + 3{,}75\, H_4] & = & 20\, H_1 & + 30\, H_2 & + 46{,}66\, H_3 & + & 75\, H_4 \\
10\, [H_1 & + 0{,}5\, H_2 & + 0{,}33\, H_3 & + 0{,}25\, H_4] & = & 10\, H_1 & + 5\, H_2 & + 3{,}33\, H_3 & + & 2{,}5\, H_4 \\
4\, [H_1 & - 0{,}5\, H_2 & + 0{,}33\, H_3 & - 0{,}25\, H_4] & = & 4\, H_1 & - 2\, H_2 & + 1{,}33\, H_3 & - & 1\, H_4 \\
1\, [H_1 & - 1{,}5\, H_2 & + 2{,}33\, H_3 & - 3{,}75\, H_4] & = & 1\, H_1 & - 1{,}5\, H_2 & + 2{,}33\, H_3 & - & 3{,}75\, H_4 \\
\hline
 & & & & & 210\, H_1 & + 693\, H_2 & + 2674\, H_3 & + & 10.840{,}5\, H_4 = {}_4S_4
\end{array}
$$

e poichè : $Y_4 = \dfrac{{}_4S_4}{\binom{n}{4}}$, si ha :

$$Y_4 = H_1 + 3{,}3\, H_2 + 12{,}73\, H_3 + 51{,}62\, H_4 .$$

Applicando ora la 38), si ha :

$$
\begin{array}{rrrrrrrrrrrr}
28 \cdot 1 \cdot [H_1 & - 3{,}5\, H_2 & - 12{,}33\, H_3 & - 43{,}75\, H_4] & = & 28\, H_1 & - 98\, H_2 & + 345{,}3\, H_3 & - & 1225\, H_4 \\
21 \cdot 2 \cdot [H_1 & - 2{,}5\, H_2 & - 6{,}33\, H_3 & - 16{,}25\, H_4] & = & 42\, H_1 & - 105\, H_2 & + 266\, . H_3 & - & 682{,}5\, H_4 \\
15 \cdot 3 \cdot [H_1 & - 1{,}5\, H_2 & - 2{,}33\, H_3 & - 3{,}75\, H_4] & = & 30\, H_1 & - 67{,}5\, H_2 & + 105\, . H_3 & - & 168{,}75\, H_4 \\
10 \cdot 4 \cdot [H_1 & - 0{,}5\, H_2 & - 0{,}33\, H_3 & - 0{,}25\, H_4] & = & 40\, H_1 & - 20\, H_2 & + 13{,}3\, . H_3 & - & 10\, . H_4 \\
6 \cdot 5 \cdot [H_1 & + 0{,}5\, H_2 & + 0{,}33\, H_3 & + 0{,}25\, H_4] & = & 30\, H_1 & + 15\, H_2 & + 10\, . H_3 & + & 7{,}5\, . H_4 \\
3 \cdot 6 \cdot [H_1 & + 1{,}5\, H_2 & + 2{,}33\, H_3 & + 3{,}75\, H_4] & = & 18\, H_1 & + 27\, H_2 & + 42\, . H_3 & + & 67{,}5\, . H_4 \\
1 \cdot 7 \cdot [H_1 & + 2{,}5\, H_2 & + 6{,}33\, H_3 & + 16{,}25\, H_4] & = & 7\, H_1 & + 17{,}5\, H_2 & + 44{,}3\, . H_3 & + & 113{,}75\, . H_4 \\
\hline
 & & & & & 210\, H_1 & - 231\, H_2 & + 826\, H_3 & - & 1897{,}5\, H_4 = {}_3S_2
\end{array}
$$

e poiché: $Y_2 = \dfrac{{}_3S_2}{\binom{n}{4}}$ si ha:

$$Y_2 = H_1 - 1{,}1\, H_2 + 3{,}93\, H_3 - 9{,}0357\, H_4$$

Applicando, infine, la 40), si ha:

$$\begin{aligned}
28 \cdot [H_1 + 3{,}5\, H_2 + 12{,}33\, H_3 + 43{,}75\, H_4] &= 28\, H_1 + 98\, H_2 + 345{,}3 \cdot H_3 + 1225\, H_4 \\
42 \cdot [H_1 + 2{,}5\, H_2 + 6{,}33\, H_3 + 16{,}25\, H_4] &= 42\, H_1 + 105\, H_2 + 266\, \cdot H_3 + 682{,}5\, H_4 \\
30 \cdot [H_1 + 1{,}5\, H_2 + 2{,}33\, H_3 + 3{,}75\, H_4] &= 30\, H_1 + 67{,}5\, H_2 + 105\, \cdot H_3 + 168{,}75\, H_4 \\
40 \cdot [H_1 + 0{,}5\, H_2 + 0{,}33\, H_3 + 0{,}25\, H_4] &= 40\, H_1 + 20\, H_2 + 13{,}3 \cdot H_3 + 10\, H_4 \\
30 \cdot [H_1 - 0{,}5\, H_2 + 0{,}33\, H_3 - 0{,}25\, H_4] &= 30\, H_1 - 15\, H_2 + 10\, \cdot H_3 - 7{,}5\, H_4 \\
18 \cdot [H_1 - 1{,}5\, H_2 + 2{,}33\, H_3 - 3{,}75\, H_4] &= 18\, H_1 - 27\, H_2 + 42\, \cdot H_3 - 67{,}5\, H_4 \\
7 \cdot [H_1 - 2{,}5\, H_2 + 6{,}33\, H_3 - 16{,}25\, H_4] &= 7\, H_1 - 17{,}5\, H_2 + 44{,}3 \cdot H_3 - 113{,}75\, H_4 \\
\hline
& 210\, H_1 + 231\, H_2 + 826\, H_3 + 1897{,}5\, H_4 = {}_3S_3
\end{aligned}$$

e poichè: $Y_3 = \dfrac{{}_3S_3}{\binom{n}{4}}$, si ha:

$$Y_3 = H_1 + 1{,}1\, H_2 + 3{,}93\, H_3 + 9{,}0357\, H_4 .$$

HERMAN WOLD

A Study on the Mean Difference, Concentration Curves and Concentration Ratio

1. — More than 20 years ago Mr. CORRADO GINI introduced in statistics the conception of mean difference in a set of quantities as an index of variability, and two years after he introduced the conception of concentration ratio, showing its relations with the mean difference [1]).

Contributions to the theory of these tools of statistical analysis have later on been given by C. GINI, L. GALVANI, V. CASTELLANO, and others [2]).

Thereby the conceptions mentioned have been extended to the case of continuous frequency curves, and special distributions have been investigated in detail.

There are, however, some circumstances that render the study of this subject rather inconvenient. Firstly, the definitions are different in the cases of continuous and discontinuous distributions [3]). This is a real inconveniency, for it obviously makes it necessary to treat the different cases separately. The natural tool for removing this inconveniency is of course the Stieltjes integral, which makes it possible to express sums and integrals in one formula. As regards the concentration ratio, R, the Stieltjes integral can be applied

1) CORRADO GINI: *Sulla misura della concentrazione e della variabilità dei caratteri*. Atti del R. Istituto Ven. di S. L. A., 1913-14, Vol. 73, II.

2) For full references, see C. GINI: *Intorno alle curve di concentrazione*, « Metron », Vol. IX, no. 3-4, 1932, p. 3, and V. CASTELLANO: *Sulle relazioni fra curve di frequenza e curve di concentrazione e sui rapporti di concentrazione corrispondenti a determinate distribuzioni*. « Metron », Vol. X, no. 4, 1933, p. 3.

3) Strictly speaking: The cases ξ and η in the statement at the end of this art.

without obstacles, a formula for R, (12), beeing given in art. 4, which is valid for all sorts of frequency distributions. After a slight and natural modification in the discontinuous case, of the mean difference, g, the same holds true also for g (3*).

The reasons to adopt the modification mentioned are further strengthened by calling attention to an important property of the mean difference as defined by the general Stieltjes formula, (3*). Regarding the concentration curve it is shown in art. 3 that, from a mathematical point of view, the different definitions for continuous and discontinuous distributions are analogous.

In the second place, some confusion may arise from the fact that the conceptions regarding concentration are defined only for distributions of a non-negative variable, whilst there is no such restriction as regards the mean difference. However, the formula (12) already referred to being valid also for a non-bounded variable, this lack of generality as regards R is at once supplied. Also the conception of concentration curves can be generalized, in this respect, as is shown in art. 4.

Lastly, (3*) and (12) are also well suited to simplify the inconvenient calculation of g and R in the case of a grouped frequency distribution. This is shown in art. 6 where also the Sheppard correction for these frequency constants is considered.

Art. 2 is devoted to a brief sketch of the principal conceptions and formulae known before. In view to make the following distinctions between the different cases more clear, we may in the first place remind of the following general property of the probability function of a statistical variable [1]:

$F(x)$ being a never decreasing function, it is known that $F(x)$ can be uniquely represented as the sum of three never decreasing functions

$$F(x) = \xi(x) + \eta(x) + \zeta(x),$$

such that $\xi(-\infty) = \eta(-\infty) = \zeta(-\infty) = 0$ and that:

$$A. \quad \xi(x) = \int_{-\infty}^{x} \xi'(t)\, dt = \int_{-\infty}^{x} F'(t)\, dt \quad \text{for all values of } x.$$

$B.$ $\eta(x)$, the "saltus function", is equal to the sum of the saltuses of $F(x)$ at all the points of discontinuity which are less than x.

---

1) $F(x)$ is the probability function of the statistical variable x, if $P(x < t) = F(t)$, denoting by $P(A)$ the probability of the event A.

C. $\zeta(x)$, the "singular function", is a continuous function which has "almost everywhere" a derivative equal to zero.

Referring to this statement, we shall, for the different types of distribution where one or two of the components ξ, η and ζ are identically equal to zero, use the notations ξ, η, $\xi + \eta$, etc. [1]).

I use the opportunity to express my best thanks to Mr. H. CRAMÉR, who has read this paper in manuscript, and thereby suggested several improvements in the presentation.

2. — $F(x)$ be the probability function of type ξ of the non-negative variable x. Denoting by $f(x)$ a derivative of $F(x)$, and putting $\int_0^\infty t\, f(t)\, dt = m$, the following system of equations:

$$(1) \qquad \begin{cases} p(x) = \int_0^x f(t)\, dt = \int_0^x dF(t) \\ q(x) = \frac{1}{m} \cdot \int_0^x t \cdot f(t)\, dt = \frac{1}{m} \cdot \int_0^x t \cdot dF(t), \end{cases}$$

x considered a parameter, defines the *concentration curve* $q = \varphi(p)$. The concentration curve is bound to the triangle (0,0); (0,1); (1,1), and runs from (0,0) to (1,1), q monotonically increasing with p.

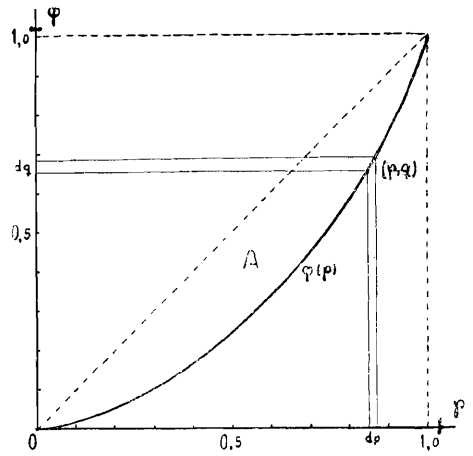

Fig. 1. — Distribution of type $\xi(f(x) = x \cdot e^{-x})$. Concentration curve, and area of concentration.

1) The statement above is, regardless of an immaterial modification, literally cited from a paper by my honoured teacher, prof. H. CRAMÉR, viz.: *On the composition of elementary errors*, «Skandinavisk Aktuarietidskrift», 1928, p. 59.

The case of a distribution in the form of a set $x_1 \leq x_2 \leq \ldots . \leq x_k$ of quantities is a special case of type η. In this case $p(x)$ and $q(x)$, as defined by (1), must be interpreted as Stieltjes integrals. But $p(x)$ and $q(x)$ now being "staircase curves", q, considered as a function of p, is strictly defined by (1) only for the following values of p, viz. 0, $p(x_1)$, $p(x)_2 \ldots . p(x_k) = 1$. In the special case $x_1 = x_2 = \ldots . = x_k$, "the case of equal distribution", the concentration curve thus is strictly defined only in the two points (0,0) and (1,1). This difficulty is over come by considering p and q as successively built up by infinitely small parts of the "quantities" x_i. This reasoning leads to the definition of the concentration curve in the interval $0 \leq q \leq 1, 0 \leq q < 1$ as that broken straight line which connects the points defined by (1); thus in the case of equal distribution the line $q = p$.

The area A between the line $q = p$ and the concentration curve is called the *area of concentration*. The minimum value of A, zero, is reached in the case of equal distribution; the upper limit of A is $1/2$. In these limit cases the distribution is said to be of resp. minimum and maximum concentration. In accordance herewith, the *concentration ratio* R is, by definition, equal to $\frac{A}{1/2} = 2A$, and we have $0 \leq R < 1$. Analytical formulae for R are, in the case ξ, supplied by the integrals

$$(2^{\xi}) \quad R = \int_{\varphi} (p\,dq - q\,dp) = 2 \cdot \int_0^1 (p - q)\,dp = 1 - 2 \cdot \int_0^1 q \,.\, dp \,.$$

The *mean difference*, g, is in the case ξ defined by [1])

$$(3^{\xi}) \quad g = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |t - u|\, f(t) f(u)\, dt\, du = 2 \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{x} f(t)\, dt \,.\, \int_{x}^{\infty} f(t)\, dt \right\} dx \,.$$

By partial integration in the first formula it is easy to deduce a formula, analogous to (2^{ξ}), showing that $g = 2Rm$. In the

---

1) The transformation is due to L. GALVANI: *Contributi alla determinazione degli indici di variabilità per alcuni tipi di distribuzione.* «Metron», Vol. IX, no. 1, 1931, p. 17.

special case of type η already mentioned is, by definition

$$(3^{\eta})\quad g = \frac{1}{k\,(k-1)} \sum_{i=1}^{k} \sum_{j=1}^{k} |x_i - x_j| = \frac{4}{k\,(k-1)} \sum_{1}^{k} i\,x_i - \frac{2\,(k+1)}{k-1}\,.\,m\ ^{1}).$$

3. — The concentration curve being strictly defined by (1) in the case ξ, we may try to find a mathematical point of view, from which there is full analogy between the cases ξ and η. Such a point of view is supplied by considering every distribution of type η, or of type $\xi + \eta$, as the limit of successive approximations of type ξ, as appears from the following lemma.

LEMMA. Let $F_1(x)$, $F_2(x)$, ... $F_n(x)$ .... be a set of probability functions of type ξ such that

$$\lim_{n \to \infty} \int_0^{\infty} d\,|F_n(t) - F(t)| = 0\,,$$

where $F(x)$ is a probability function of type $\xi + \eta$ corresponding to a variable which is never negative.

Putting

$$(4)\qquad \begin{cases} m_n = \int_0^{\infty} t\,.\,d\,F_n(t) \\ m = \int t\,.\,d\,F(t)\,, \end{cases}$$

we suppose further that

$$(5)\qquad \lim_{n \to \infty} m_n = m > 0\,.$$

Then as $n \to \infty$ the concentration curve $\varphi_n(p)$ corresponding to the distribution $F_n(x)$ tends uniformly to the concentration curve $\varphi(p)$ corresponding to the distribution $F(x)$.

Before the proof we mention that the lemma easily can be generalized. Firstly we observe that (1), interpreted as Stieltjes in-

---

1) For the transformation, given by C. GINI, see e. g. B. DE FINETTI: *Sui metodi proposti per il calcolo della differenza media.* « Metron », Vol. IX, no. 1, 1931, p. 50.

tegrals, determines $q\,(x)$ and $q\,(p)$ also in the case $\xi + \zeta$. Then it follows that the lemma still is valid if we consider a general distribution of type $\xi + \eta + \zeta$ which we approximate successively by a set $F_n\,(x)$ of type $\xi + \zeta$.

Owing to our lemma it is often sufficient to analyse only the case ξ, for the treatment of the case η can thereby be brought back to the case ξ. *E. g.*, Mr. V. CASTELLANO [1]), starting from (1), has pointed out some general properties of the concentration curves in the case ξ. Without further notice, however, his theorems are not proved for the case η or for the general case $\xi + \eta + \zeta$, where the parametric representation (1) for the concentration curve is not applicable. By the use of our lemma it is, however, easy to prove that each theorem of Mr. CASTELLANO's is true for general distributions.

The essential point of the proof is to compare the trend of the concentration curves $\varphi\,(p)$ and $\varphi_n\,(p)$ in those intervals where p, considered as a function $p\,(x) = \int_0^x d\,F\,(t)$ of the parameter x, is discontinuous. Let x_0 be such a value, and put

$$p_1 = \int_0^{x_0-0} d\,F\,(t)\,,\quad p_2 = \int_0^{x_0+0} dF\,(t)\,,\quad \text{and}\quad q_1 = \frac{1}{m}\,.\int_0^{x_0-0} t\,.\,d\,F\,(t)\,.$$

Then we have

$$q_2 = \frac{1}{m}\,.\int_0^{x_0+0} t\,.\,d\,F\,(t) = q_1 + \frac{x_0}{m}\,(p_2 - p_1)\,.$$

In the interval $(p_1\,,\,p_2)$, the length of which is equal to the "saltus" $p_2 - p_1$ of $p\,(x) = F\,(x)$ in the point x_0, the concentration curve $\varphi\,(p)$, according to art. 2, thus coincides with the straight line

$$q = q_1 + \frac{x_0}{m}\,(p - p_1)\,. \tag{6}$$

Hence, on one hand, ε being sufficiently small, when the parameter x varies from $x_0 - \varepsilon$ to $x_0 + \varepsilon$, the concentration curve $\varphi\,(p)$, regardless of amounts vanishing with ε, runs in an interval of the

---

1) Loc. cit.

length $p_2 - p_1$, $\varphi(p)$ thereby increasing linearly, by $\frac{x_0}{m}(p_2 - p_1)$ in total. On the other hand, examining the trend of the concentration curves $\varphi_n(p)$ as the parameter x varies between $x_0 - \varepsilon$ and $x_0 + \varepsilon$, we find

$$(7) \qquad p_n(x) = \int_0^x d F_n(t) = \int_0^{x_0-\varepsilon} d F_n(t) + \int_{x_0-\varepsilon}^x d F_n(t) .$$

Further is, according to (5),

$$(8) \quad q_n(x) = \frac{1}{m_n} \cdot \int_0^x t \,.\, d F_n(t) = \frac{1}{m} \cdot \int_0^{x_0-\varepsilon} t \,.\, d F_n(t) + \frac{1}{m} \int_{x_0-\varepsilon}^x t \,.\, d F_n(t) + K \,.\, \varepsilon ,$$

K being uniformly bounded. (8) gives

$$(9) \quad q_n(x) = \frac{1}{m} \cdot \int_0^{x_0-\varepsilon} t \,.\, d F_n(t) + \frac{x_0 + \varepsilon \,.\, \Theta(x)}{m} \int_{x_0-\varepsilon}^x d F_n(t) + K \,.\, \varepsilon ,$$

where $-1 \leq \Theta(x) \leq 1$. Examining (7) we find that, as x varies from $x_0 - \varepsilon$ to $x_0 + \varepsilon$, the concentration curve $\varphi_n(p)$ runs in an interval of the length $\int_{x_0-\varepsilon}^{x_0+\varepsilon} d F_n(t)$. As n increases, the length of this interval tends to $p_2 - p_1$, regardless of an amount vanishing with ε. In the same way we find from (9) that for increasing n, the increase of $q_n(x)$ in the interval mentioned tends to $\frac{x_0}{m}$, $(p_2 - p_1)$ still regardless of amounts tending to zero with ε. However, comparing (7) and (9) we see also that the trend of $q_n(p)$ tends to a straight line, the variation of $p_n(x)$ anc $q_n(x)$ being essentially proportional to $\int_{x_0-\varepsilon}^x d F_n(t)$. And it easy to see that this straight line must be the same as (6). For, examining the concentration curves $\varphi(p)$ and $\varphi_n(p)$ as the parameter x varies, we have seen that as x passes a point of discontinuity of $p(x)$, the limit concentration curve of the series $\varphi_n(p)$ varies exactly as $\varphi(p)$. And obviously the same holds true when x describes an interval without discontinuities in $p(x)$. As

further every concentration curve starts in the point (0,0) and ends in (1,1), it follows that $\varphi_n(p)$ must tend uniformly to $\varphi(p)$, which proves the lemma.

Regarding the conditions of validity, (5) is obviously also a necessary condition. *E. g.*, given a distribution $x_1, x_2, \ldots x_k$ of type η, we may not, even if $\lim_{n\to\infty} \varepsilon_n = 0$, choose for $F_n(x)$ a function, which for $x > x_k$ is equal to $1 - \varepsilon_n \cdot x^{-2-\varepsilon_n}$. On the contrary it is allowed to use for $F_n(x)$ a curve which for $x > x_k$ is equal to $1 - \varepsilon_n \cdot x^{-a-\varepsilon_n}$, a being > 2. Further, if for every n we have $F_n(x) = 1$ for $x >$ const. $C > x_k$, (5) is always satisfied.

What regards the restriction $m > 0$, it is easy to construct a set $F_n(x)$ satisfying (5), and (according to (4)) converging to the distribution (of type η) with $m = 0$, but nevertheless such that $\varphi_n(p)$ does not converge to $\varphi(p) = p$. We need only take a set of positive numbers $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n, \ldots$ where $\varepsilon_n \to 0$, and choose

$$F_n(x) = \begin{cases} \dfrac{x}{\varepsilon_n} & \text{for } 0 \leq x \leq \varepsilon_n, \\ 1 & \text{for } \varepsilon_n \leq x. \end{cases}$$

A simple calculation shows that $\varphi_n(p)$ here is independent of n, being equal to $p^2 + p$.

The restriction $m > 0$ concerns, however, only the very special case of equal distribution. It does not enter in the more general cases, which are those of practical importance. For the rest, when $m = 0$ is $q(x)$ as a function of x (and *a fortiori* of p) not strictly defined by (1). The only condition of validity of any importance for the lemma is, therefore, (5).

4. — The lemma in art. 3 shows that Mr. GINI's definition of the concentration curve is analogous in the cases ξ and η. Regarding the concentration ratio the analogy in definition is still more complete, for, as already mentioned in art. 1, there exists a formula valid for every kind of distribution.

In fact, starting from the case ξ, we have for the area of concentration

$$A = \frac{1}{2} \cdot \varphi (p\,dq - q\,dp) = \frac{1}{2m} \cdot \int_0^\infty f(t) \left\{ t \int_0^t f(u)\,du - \int_0^t u \cdot f(u)\,du \right\} dt .$$

Integration by parts gives

$$\int_0^t du \cdot \int_0^u f(v)\,dv = t \cdot \int_0^t f(u)\,du - \int_0^t u \cdot f(u)\,du ,$$

and thus we find

$$A = \frac{1}{2m} \int_0^\infty f(t)\,dt \int_0^t du \int_0^u f(v)\,dv . \tag{10}$$

Writing (10) as a Stieltjes integral, and inserting $\int_0^u f(v)\,dv = F(u)$, we find

$$A = \frac{1}{2m} \int_0^\infty d\,F(t) \cdot \int_0^t F(u)\,du . \tag{11}$$

$\int_0^x F(u)\,du$ always being a continuous function, (11) has a meaning for every distribution. That (11) gives the correct value for the area of concentration also in the case of a general frequency distribution follows from our lemma. Regardig, *e. g.*, the case η, we can choose a series of distributions of type ξ such that $\varphi_n(p)$ converges to $\varphi(p)$. Denoting by A_n and A the areas of concentration, it follows that $\lim_{n \to \infty} A_n = A$, and if we further only observe that

$$\lim_{n \to \infty} A_n = \lim_{n \to \infty} \frac{1}{2m_n} \int_0^\infty d\,F_n(t) \cdot \int_0^t F_n(u)\,du = \frac{1}{2m} \int_0^\infty d\,F(t) \cdot \int_0^t F(u)\,du ,$$

the validity of (11) is proved.

Accordingly we have for the concentration ratio the general formula

$$R = \frac{1}{m} \int_0^\infty d\,F(t) \cdot \int_0^t F(u)\,du . \tag{12}$$

Regarding the mean difference g, the formula $g = 2mR$ does not hold true in the case of a distribution $x_1, x_2, \ldots, x_k$ of type η. We have here to introduce a correction, namely

$$g = 2mR \cdot \frac{k-1}{k},$$

a fact that hitherto has been overlooked. This inconveniency can, however, be removed by attaching weight also to the vanishing differences $|x_i - x_i|$. Then (3η) becomes

$$(3\eta^*) \qquad g = \frac{1}{k^2} \sum_1^k \sum_1^k |x_i - x_j| = \frac{4}{k^2} \sum_1^k i \cdot x_i - \frac{2(k+1)}{k} \cdot m \cdot ^1)$$

Now there is full analogy between the different cases, the formulae

$$g = 2mR = 2 \int_0^\infty dF(t) \cdot \int_0^t F(u)\, du$$

being true for an arbitrary distribution. Obviously we have in the case of a variable not assuming only positive values

$$(3^*) \qquad g = 2 \cdot \int_{-\infty}^{\infty} dF(t) \cdot \int_{-\infty}^{t} F(u)\, du .$$

The slight modification proposed has another advantage, the mean difference g, as definited by (3^*), having a property which might further justify its adoption. This remark has reference to the wellknown ω^2 — method for testing goodness of fit [2]).

$F(x)$ be a probability function, and $x_1, x_2, \ldots x_k$ be a set of independent observations of a variable x. The ω^2 — method for testing whether $F(x)$ might be considered as the probability function of the variable x consists in the computation of the integral

---

1) As a matter of fact, this variety of g is mentioned also by Mr. GINI: See *Di una estensione del concetto di scostamento medio e di alcune applicazioni alla misura della variabilità dei caratteri qualitativi.* Atti del R. Ist. Veneto di Sc. Lett. ed Arti, Vol. 77, nr. 2 (1918) pag. 398.

2) The ω^2 — method seems to have been proposed for the first time by H. CRAMÉR, in his Swedish textbook on the theory of probabilities (1925). Explicit formulae are there given for the case of a normal distribution. See also his paper (already cited), where the case of the P-series of Cramér (a modification of the A-series of Charlier) is treated. See also R. v. MISES: *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und Theoretischen Physik.* Leipzig und Wien, 1931, p. 316-335.

$$J = \int_{-\infty}^{\infty} [F_h(t) - F(t)]^2 \, dt ,$$

where $F_h(t)$ is the relative number of observations $\leq t$. J is then compared with its mean value, which is

$$M[J] = \frac{1}{h} \cdot \int_{-\infty}^{\infty} F(t) \left[1 - F(t)\right] dt .$$

By partial integration we find

$$M[J] = \frac{1}{h} \cdot \int_{-\infty}^{\infty} d\,F(t) \cdot \int_{-\infty}^{t} F(u)\, du = \frac{g}{2h} ,$$

g being defined by (3*). From the point of view of the ω^2 — method for testing goodress of fit, the mean difference, as defined by (3*), thus is a very important characteristic number for a theoretical frequency distribution. The same holds true for the concentration ratio as defined in art. 2 or as directly defined by (12). In a recent paper «*Metron*» IX, l. c.) Mr. GINI has proposed a modification of R, caused by geometrical argumentation on the maximum value of the area of concentration. However, the modified R is not related to g by the simple equality (3*), not either to the ω^2 — method [1]).

5. — Only a few lines are necessary as regards the generalization of the conceptions of concentration to the case of a statistical variable assuming also negative values. In the first place we observe that these conceptions are of interest only when $m > 0$. Assuming a

1) [The modification proposed by GINI consists in a generalisation of the formule for R so that the R values may always have 0 as their lower and 1 as their upper limit. The area of concentration is thus made equal to the maximum it can actually attain for the phenomenon under consideration, a maximum equal to the area of the rectangular triangle in which the curve of concentration is inscribed when the character has no upper or lower limit, but equal to the area of a smaller triangle than this, when one at least of the aforesaid limits comes into play. The generalisation is therefore determinated not by formal considerations but by conceptual needs and so as to make comparable the R values concerning different distributions. It only implies moreover the introduction of a coefficient of correction which modifies very little the relation of R to g and to the η^2 method. (See GINI: *Sul massimo degli indici di variabilità assoluta e sulle sue applicazioni agli indici di variabilità relativa e al rapporto di concentrazione*, " Metron ", Vol. VIII, N. 3, February, 1930. And see also L. GALVANI: *Sulle curve di concentrazione relative a caratteri non limitati e limitati*, " Metron ", Vol. X, N° 3, 1932; V. CASTELLANO: *Sulle relazioni*, etc., op. cit., note 5 on pp. 16-17). *Ed. note*].

distribution with $m > 0$ everything regarding the conceptions of concentration stated above remains true if we only throughout replace zero by $-\infty$ in the lower limits for the integrals. The only modifications caused thereby are that the concentration curve no longer is bounded downwards by the p — axis, neither is $q(x)$ in all cases monotonically increasing, nor $A \leq \frac{1}{2}$, nor $R \leq 1$.

An example of a general concentration curve of type $\xi + \eta$ is given in the figure below.

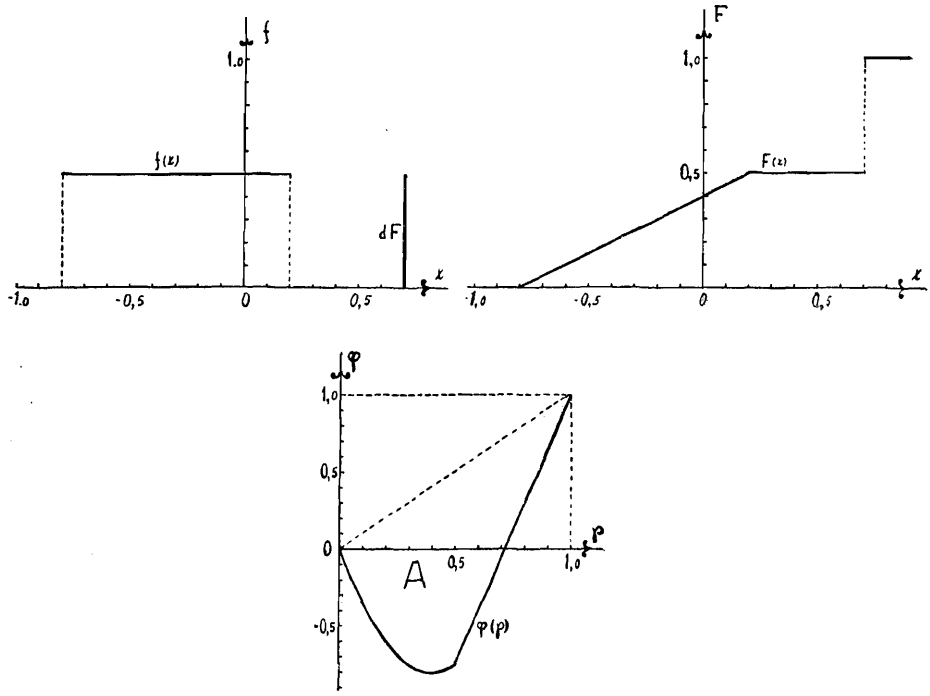

Fig. 2. — Frequency distribution of type $\xi + \eta$ [1]. Frequency function, (including one saltus) probability function, and concentration curve.

We may here call attention to the analogy between the frequency constants analysed in this paper and other frequency con-

1) The distribution describes the risk situation of an insurance company at a special form of wholelife assurance with increasing amount. For this and similar examples regarding the probability function and the frequency function, see H. CRAMÉR : *On the Mathematical Theory of Risk.* «Försäkringsaktiebolaget Skandia, 1855-1930», Stockholm, 1930, p. 22-30

stants. *E. g.* there is a certain formal analogy between the concentration ratio and the variation coefficient of CHARLIER $\left(\rho = 100 \cdot \frac{\sigma}{m}\right)$. For the sake of simplicity we give the formulae for a distribution $f(x)$ of type ξ. Denoting the two coefficients by $R[f(x)]$ and $\rho[f(x)]$, $R[f(x)]$ as well as $\rho[f(x)]$ is of dimension zero, *i. e.*

$$(13) \qquad \begin{cases} R[a \cdot f(a x)] = R[f(x)] \\ \rho[a \cdot f(a x)] = \rho[f(x)] \,. \end{cases}$$

Further we have

$$(14) \qquad \begin{cases} R[f(x - a)] = \dfrac{m}{m + a} R[f(x)] \\ \rho[f(x - a)] = \dfrac{m}{m + a} \rho[f(x)] \,, \end{cases}$$

$$(15) \qquad \begin{cases} R[f(2m - x)] = R[f(x)] \\ \rho[f(2m - x)] = \rho[f(x)] \,. \end{cases}$$

These relations, as well as corresponding properties of the mean difference, are easily proved by means of (12) and (3*), as may be given to the reader to verify.

The relation (13) shows that a change of the scale for x does not affect R. (14) gives the modification of R when the frequency distribution is translated along the x-axis. (15) shows that R is the same for two distributions, which are symmetrical round the vertical through the common arithmetical mean of the distributions [1]. Other frequency constants analogous to R, for which (13), (14) and (15) hold true, are also [2]

$$\frac{\text{Quartile deviation}}{\text{Mean of quartiles}} \quad \text{and} \quad \frac{\text{Mean deviation}}{\text{Median}} \,.$$

1) Of course, similar relations have been known long before in the special cases ξ and η; see, *e. g.*, V. CASTELLANO, l. c. p. 10-20.

2) See A. L. BOWLEY: *Elements of Statistics*, 5th ed., London, 1926, p. 116.

6. — For the numerical calculation of g and R we must distinguish different cases.

α. A distribution in the form of a set of k quantities

$$x_1 \leq x_2 \leq \ldots \leq x_k .$$

β. A grouped frequency distribution, *i. e.*

$$\begin{matrix} x_1 & x_2 & \ldots\ldots & x_n \\ f(x_1) & f(x_2) & \ldots\ldots & f(x_n) \end{matrix} .$$

α and β are special cases of type η.

γ. A frequency distribution of type ξ given in a table of equidistant division, thus

$$\begin{matrix} x_1 & & x_2 & & \ldots\ldots & & x_n \\ \int_{-\frac{h}{2}}^{\frac{h}{2}} f(x_1 + t)\, dt & & \int_{-\frac{h}{2}}^{\frac{h}{2}} f(x_2 + t)\, dt & & \ldots\ldots & & \int_{-\frac{h}{2}}^{\frac{h}{2}} f(x_n + t)\, dt \end{matrix} .$$

δ. Other cases.

Ad α. The most convenient formula is here, perhaps, the second expression in ($3^{\eta *}$), due to Mr. GINI. As an example we take the schematical distribution $x_i = i$. Accordingly we find

$$g = \frac{4}{k^2} \sum_1^k i \,.\, x_i - \frac{2(k+1)}{k} \,.\, m = \frac{4}{k^2} \sum_1^k i^2 - \frac{(k+1)^2}{k} =$$

$$= \frac{4}{k^2} \cdot \frac{k(k+1)(2k+1)}{6} - \frac{(k+1)^2}{k} = \frac{k^2 - 1}{3k} ,$$

and

$$R = \frac{g}{2m} = \frac{k-1}{3k} .$$

As a first example of the use of (12) we may calculate R as follows

(1)	(2)	(3)	(4)	(5)
x_i	$k \,.\, d\,F(x_i)$	$k \,.\, F(x_i - 0)$	$k \,.\, \int_{-\infty}^{x_i} F(t)\,dt$	$k^2 \,.\, d\,F(x_i) \,.\, \int_{-\infty}^{x_i} F \,.\, dt$
1	1	0	0	0
2	1	1	1	1
3	1	2	3	3
.	.	.	.	.
.	.	.	.	.
k	1	$k-1$	$\binom{k}{2}$	$\dfrac{\binom{k}{2}}{\binom{k+1}{3}}$

The sum of the last column represents $k^2 \,.\, \int_{-\infty}^{\infty} d\,F(t) \,.\, \int_{-\infty}^{t} F(u)\,du$, which gives

$$R = \frac{1}{m} \cdot \frac{1}{k^2} \cdot \binom{k+1}{3} = \frac{k-1}{3k},$$

i. e. the same result as before.

Ad β. In the case of equidistancy, the formulae (3*) and (12) lead to the same simple calculations as in the schematic example above. As the case γ leads to the same calculation scheme, we may here take a simple example, and choose the following classical distribution of the yearly number of deaths by equine kicks in the Prussian army [1]).

(1) Number of deaths	(2) Frequency	(3) Σ (2)	(4) Σ (3)	(5) (2) . (4)
x_i	$k \,.\, d\,F(x_i)$	$k \,.\, F(x_i - 0)$	$k \,.\, \int_{-\infty}^{x_i} F(t)\,dt$	$k^2 \,.\, d\,F(x_i) \,.\, \int_{-\infty}^{x_i} F(t)\,dt$
0	109	0	0	0
1	65	109	109	7085
2	22	174	283	6226
3	3	196	479	1437
4	1	199	678	678
	200			15,426

1) L. v. BORTKIEWICZ: *Das Gesetz der kleinen Zahlen.* Leipzig 1898, p. 25.

As we have $k = 200$ and $m = 0{,}61$, we find by (12) and (3*)

$$R = \frac{1}{k^2} \cdot \frac{1}{m} \cdot 15\,426 = \frac{15\,426}{24\,400} = 0{,}63\,,$$

$$g = \frac{15\,426}{20\,000} = 0{,}7713\,.$$

As is well known, this frequency table may be approximated by a Poisson distribution, i. e.

$$d\,F\,(x) = \frac{m^x \cdot e^{-m}}{x\,!}\,,\ (x = 0,\ 1,\ 2,\ 3,\ \ldots\ldots)\,.$$

In this case we find after some deductions

$$g = 2\,m \cdot e^{-2m} \cdot \sum_0^\infty \frac{m^{2\,i}\,(m + 1 + i)}{i\,!\,(i + 1)\,!}\,.$$

For $m = 0{,}61$ this formula gives $g = 0.7707$, in excellent accordance with the mean difference of the empirical distribution.

If the points x_i are not equidistant, the computation of col. (4) is a little more complicated. For this case we refer to the completely analogous example under δ.

An indirect and approximate method is to calculate g and R via the area of concentration. This may be done by means of a good planimeter or by numerical integration. For these methods, see Mr. GINI's paper in «Metron» IX, cited above.

Ad γ. Here the problem is to calculate $\int_{-\infty}^{\infty} f\,(x)\,d\,x \cdot \int_{-\infty}^{x} F\,(t)\,d\,t$ from a table giving $\int_{-\frac{h}{2}}^{\frac{h}{2}} f\,(x_i + t)\,d\,t$ for equidistant values of x_i. This is a special case of the classical Sheppard problem. Regarding the general problem to calculate the mean of an arbitrary function $k\,(x)$, it is known [1]) that

$$\int_{-\infty}^{\infty} k\,(x) \cdot f\,(x)\,d\,x = \sum_{-\infty}^{\infty} \varkappa\,(x_i) \cdot \int_{-\frac{h}{2}}^{\frac{h}{2}} f\,(x_i + t)\,d\,t + r\,,$$

1) H. WOLD: *Sulla correzione di Sheppard.* «Giorn. dell'Ist. Italiano degli Attuari», Vol. 5, no., 2-3, 1934, p. 308.

if

$$(16) \qquad f(x) \cdot \int_0^x \varkappa(t)\, dt \to 0 \quad \text{for} \quad x \to \pm\infty ,$$

$\varkappa(x)$ being determined by the difference equation

$$\frac{\Delta}{h}\varkappa(x) = \frac{\varkappa(x+h) - \varkappa(x)}{h} = k'\left(x + \frac{h}{2}\right),$$

and r being the remainder in the Euler-MacLaurin formula

$$\frac{1}{h} \cdot \int_{-\infty}^{\infty} \varkappa(x)\, dx \cdot \int_{-\frac{h}{2}}^{\frac{h}{2}} f(x+t)\, dt = \sum_{-\infty}^{\infty} \varkappa(x_i) \cdot \int_{-\frac{h}{2}}^{\frac{h}{2}} f(x_i + t)\, dt + r .$$

In our special case is $k(x) = \int_{-\infty}^{x} F(t)\, dt$. Thus we have for the determination of $\varkappa(x)$

$$\frac{\Delta}{h}\varkappa(x) = F\left(x + \frac{h}{2}\right).$$

This difference equation gives immediately

$$\varkappa(x) = h \cdot \sum_{-\infty}^{0} F\left(x - \frac{h}{2} + ih\right).$$

Observing that

$$F\left(x - \frac{h}{2}\right) = \sum_{-\infty}^{-1} \int_{-\frac{h}{2}}^{\frac{h}{2}} f(x - ih)\, dx ,$$

we see that the calculation of g and R in this case leads to exactly the same computation scheme as under β. Thus, according to (16) we have for distributions with finite dispersion σ, omitting the remainder $2r$,

$$(20) \qquad g = 2h \cdot \sum_{x_i=-\infty}^{\infty} \int_{-\frac{h}{2}}^{\frac{h}{2}} f(x_i + t) \cdot \sum_{j=-\infty}^{i} \sum_{k=-\infty}^{j-1} \int_{-\frac{h}{2}}^{\frac{h}{2}} f(x_k + t)\, dt .$$

As an example we take the distribution $f(x) = x \cdot e^{-x}$, $(x > 0)$. The concentration curve corresponding to this distribution is repre-

sented in fig. 1. It is easy to verify that here $F(x) = 1 - (x+1) \cdot e^{-x}$, $\int_0^x F(t)\,dt = x - 2 + (x+2) \cdot e^{-x}$, $m = 2$, and $g = 1{,}5$. On the other hand, according to (20), choosing $h = \frac{1}{2}$, the approximate calculation of g can be made from the integrals $\int_{n/2}^{\frac{n+1}{2}} f(t)\,dt$ by the following calculations which are completely analogous to the case β.

(1) x_i	(2) $\int_{-h/2}^{h/2} f(x_i+t)\,dt$	(3) $F(x_i - x/2)$	(4) $\sum_{j=-\infty}^{i} F\left(x_j - \frac{h}{2}\right)$	(5) (2) . (4)
0,25	0,090	0,000	0,000	0,0000
0,75	174	090	0,090	157
1,25	178	264	0,354	630
1,75	152	442	0,796	1210
2,25	119	594	1,390	1654
2,75	88	713	2,103	1851
3,25	63	801	2,904	1830
3,75	44	864	3,768	1658
4,25	31	908	4,676	1450
4,75	21	939	5,615	1179
5,25	14	960	6,575	921
5,75	9	974	7,549	679
6,25	6	983	8,532	512
6,75	4	989	9,521	381
7,25	3	993	10,514	315
7,75	2	996	11,510	230
8,25	1	998	12,508	125
8,75	1	999	13,507	135
9,25	0	1,000		1,4917

As $2h = 1$, we find $g = 1{,}492$, showing that the error r in this, rather unfavourable, case is less than 0,01, or 0,68 %. The approximate value of R is accordingly 0,373, whilst the exact value is 0,375.

For the indirect calculation of g and R by means of the area of concentration, see β.

Ad δ. Also in the case of non-equidistant distribution the Sheppard transformation from integrals to sums is approximatively valid. This leads obviously to the same computation scheme as un-

$$x_i - \frac{h_i}{2}$$

der γ, only the calculation of $\int_{-\infty} F(t)\,dt$ becomes a little more complicated. The simplest manner to perform the calculation is now, perhaps, to insert a column with the differences $x_{i+1} - x_i$ of the central points of the intervals. As an example, we take a distribution investigated by Mr. GINI (in the paper in « Metron » IX cited above).

The State of Victoria, 1910. *Real Estates.*

(1) Central area (acres) x_i	(2) Number of estates	(3) $x_{i+1} - x_i$	(4) Σ (2)	(5) $\Sigma \frac{(3)\cdot(4)}{1000}$
2,5	3 469	7,5	56 771	26 359
10	4 420	12,5	52 351	25 933
22,5	4 854	17,5	47 497	25 279
40	3 866	35	43 631	24 448
75	6 696	75	36 935	22 921
150	9 208	100	27 727	20 151
250	5 422	75	22 305	17 378
325	2 953	50	19 352	15 705
375	2 951	75	16 401	14 737
450	2 863	100	13 538	13 507
550	2 212	75	11 326	12 154
625	1 650	50	9 676	11 304
675	918	75	8 758	10 820
750	1 249	100	7 509	10 164
850	1 014	100	6 495	9 413
950	1 173	300	5 322	8 763
1 250	2 583	500	2 739	7 166
1 750	1 062	500	1 677	5 797
2 250	514	500	1 163	4 958
2 750	270	750	893	4 377
3 500	329	1 000	564	3 707
4 500	150	1 750	414	3 143
6 250	161	2 500	253	2 419
8 750	78	3 750	175	1 786
12 500	79	5 000	96	1 130
17 500	52	7 500	44	650
25 000	22	10 000	22	320
35 000	15	10 000	7	100
45 000	5	15 000	2	30
60 000	2			

The calculations are in this example carried upwards from the foot of the table, the direction obviously being indifferent. Col. (5) gives the successive sums of the numbers on col. (4), these numbers being weighted by the numbers in col. (3).

Multiplying the numbers in cols. (2) and (5) and adding, we find 1103 808 000. The total number of estates being 60 240, we find

$$g = \frac{2.1103\ 808\ 000\ 000}{60\ 240^2} = 608{,}3\,.$$

The exact value is 604,6 which might be found by using instead of col. (3) the differences between the mean estate areas in consecutive groups (not given here). Even in this case it is dubious whether the indirect method of calculating g by means of the area of concentration is more convenient.

The cases analysed above being those of practical importance, it is not necessary to carry the investigation further. The calculations always can be performed, either by approximating the integral (11), or by approximate evaluation of the area of concentration.

N. SMIRNOFF

Ueber die Verteilung des allgemeinen Gliedes in der Variationsreihe

Nehmen wir an, dass die Folge

$$x_1 \leq x_2 \leq x_3 \leq \ldots . \leq x_n \tag{1}$$

die beobachteten Werte einer zufälligen Variablen X vorstellt, die ihrer Grösse nach geordnet sind. Nehmen wir ferner an, dass $\Phi(x)$ das Verteilungsgesetz von X bezeichnet, so dass

$$\Phi(x) = W[X \leq x] \ (*).$$

Die Funktion $\Phi(x)$ werden wir differenzierbar voraussetzen ; dann wird $\Phi'(x) = f(x)$ die Wahrscheinlichkeitsdichte der Grösse X vorstellen.

Wir wollen die Folge (1) die Variationsreihe der Grösse X nennen, die durch n Beobachtungen erhalten worden ist und wollen vor allem das Verteilungsgesetz des k — *ten* Gliedes dieser Reihe feststellen.

Die gesuchte Wahrscheinlichkeit $W[x_k \leq x]$ erhält in unseren Voraussetzungen folgenden Ausdruck :

$$W[x_k \leq x] = \frac{\int_{-\infty}^{x} f(x_1)\,dx_1 \int_{x_1}^{x} f(x_2)\,dx_2 \ldots \int_{x_{k-1}}^{x} f(x_k)\,dx_k \int_{x_k}^{\infty} f(x_{k+1})\,dx_{k+1} \ldots \int_{x_{n-1}}^{\infty} f(x_n)\,dx_n}{\int_{-\infty}^{+\infty} f(x_1)\,dx_1 \int_{x_1}^{\infty} f(x_2)\,dx_2 \ldots . \int_{x_{n-1}}^{\infty} f(x_n)\,dx_n}. \tag{2}$$

(*) $W[E]$ wird in dieser Arbeit immer die Wahrscheinlichkeit des Ereignisses E bezeichnen.

Das n — fache Integral im Nenner (2) ist gleich $\frac{1}{n!}$ und stellt die Wahrscheinlichkeit vor, nach n Beobachtungen die nichtabnehmende Reihenfolge der Form (1) zu erhalten. Der Zähler ist leicht zum einfachen Integral

$$\frac{1}{k-1!\,n-k!}\int_{-\infty} f(x_k)\,\Phi^{k-1}(x_k)\,[1-\Phi(x_k)]^{n-k}\,dx_k$$

zurückzuführen und stellt die Wahrscheinlichkeit einer nichtabnehmenden Reihe dar, deren k erste Glieder nicht grösser als x sind.

Folglich haben wir

$$W[x_k \leq x] = V_k(x) = \frac{n!}{k-1!\,n-k!}\int_{-\infty}^{x} \Phi^{k-1}(\xi)\,[1-\Phi(\xi)]^{n-k} f(\xi)\,d\xi \quad (3).$$

Wenn die Grösse X eine gleichmässige Wahrscheinlichkeitsverteilung im Interwall (0,1) besitzt, so ist

$$f(\xi) = 1\ ,\quad \Phi(\xi) = \xi\ ,\quad \text{und folglich}$$

$$V_k(x) = \frac{n!}{k-1!\,n-k!}\int_0^x \xi^{k-1}(1-\xi)^{n-k}\,d\xi\,. \qquad (4)$$

Im allgemeinen Falle können wir den Ausdruck (3) transformieren, indem wir die Funktion $x = \Psi(z)$ $(0 < z < 1)$ heranziehen, die zur Funktion $\Phi(x) = z$ invers ist.

Dann ist:

$$W[x_k \leq \Psi(z)] = \frac{n!}{k-1!\,n-k!}\int_{-\infty}^{\Psi(z)} \Phi^{k-1}(\xi)\,[1-\Phi(\xi)]^{n-k} f(\xi)\,d\xi =$$

$$= \frac{n!}{k-1!\,n-k!}\int_0^z y^{k-1}(1-y)^{n-k}\,dy \qquad (5)$$

In Folgendem wollen wir die Grenzform des Verteilungsgesetz (3) untersuchen unter der Voraussetzung, dass n unbegrenzt zunimmt und unter verschiedenen Voraussetzungen inbetreff des Ranges des k — *ten* Gliedes der Variationsreihe.

Hier muss noch auf einen Umstand hingewiesen werden, der uns in Folgendem die Möglichkeit bieten wird unsere Untersuchung zu vereinfachen.

Das Verteilungsgesetz des Gliedes x_{n-k+1} (des k — *ten* Gliedes vom Ende der Variationsreihe gerechnet) wird nach Formel (3) folgendermassen lauten

$$W[x_{n-k+1} \leq x] = V_{n-k+1}(x) = \frac{n!}{n-k!\,k-1!} \int_{-\infty}^{x} \Phi^{n-k}(\xi)\,[1-\Phi(\xi)]^{k-1} f(\xi)\,d\xi$$

oder

$$V_{n-k+1}(x) = \frac{n!}{n-k!\,k-1!} \int_{-x}^{+\infty} \Phi^{n-k}(-\xi)\,[1-\Phi(-\xi)]^{k-1} f(-\xi)\,d\xi \quad (6).$$

Die Funktion $1-\Phi(-\xi) = \Phi_1(\xi)$ bestimmt aber das Verteilungsgesetz der Grösse $X' = -X$, da $W[X' \leq \xi] = W[X \geq -\xi]$ ist und die Funktion $f(-\xi)$ die Dichte der Grösse X' vorstellt.

Das Verteilungsgesetz von x_k', dem k — *ten* Gliede der Variationsreihe der Grösse X' wird nach derselben Formel (3) lauten

$$W[x_k' \leq x] = V_k^{(1)}(x) = \frac{n!}{n-k!\,k-1!} \int_{-\infty}^{x} [1-\Phi(-\xi)]^{k-1} \Phi^{n-k}(-\xi)\, f(-\xi)\,d\xi \quad (7).$$

Der Vergleich von (6) und (7) ergibt

$$W[x_{n-k+1} \leq x] = W[x_k' \geq -x] = W[-x_k' \leq x]$$

oder

$$V_{n-k+1}(x) = 1 - V_k^{(1)}(x)\,. \qquad (8)$$

Folglich haben die Grössen x_{n-k+1} und $-x_k'$ dieselbe Verteilung. Hieraus folgt, dass ihre Momente verschiedenen Ordnungen (wenn sie existieren) übereinstimmen:

$$E(x_{n-k+1})^s = E(-x_k')^s = (-1)^s\, E(x_k')^s \quad (*). \qquad (9)$$
$$(s = 0, 1\,2 \ldots.)$$

Insbesondere ist

$$E(x_{n-k+1})^s = -E(x_k')$$
$$E(x_{n-k+1})^2 = \quad E(x_k')^2\,. \qquad (10)$$

(*) Mit $E(x)$ wird die mathematische Erwartung der Grösse x bezeichnet.

Aus der Formel (10) folgt, dass die Streuung von x_{n-k+1} derjenigen von x_k' gleich ist.

II.

Untersuchen wir zuerst den Fall, wo beide Grössen k und $n-k$ unbegrenzt zunehmen bei der Zunahme von n:

Es sei

$$\lambda_k = \frac{k-1}{n-1} \quad \nu_k = \frac{n-k}{n-1} \quad (\lambda_k + \nu_k = 1) \tag{11}$$

$$\tau_k = \sqrt{\frac{\lambda_k \nu_k}{n-1}}. \tag{12}$$

Wir wollen jetzt mit Hilfe der Formel (5) die Wahrscheinlichkeit $W_n[z_1, z_2]$ der Ungleichungen

$$\Psi(\lambda_k + z_1 \tau_k) \leq x_k \leq \Psi(\lambda_k + z_2 \tau_k) \; ; \; (z_1 \leq z_2) \tag{13}$$

feststellen.

Wir haben

$$W_n[z_1, z_2] = \frac{n!}{k-1!\, n-k!} \int_{\lambda_k + z_1 \tau_k}^{\lambda_k + z_2 \tau_k} y^{k-1} (1-y)^{n-k} \, dy$$

oder

$$W_n[z_1, z_2] = \frac{n!}{k-1!\, n-k!} \tau_k \int_{z_1}^{z_2} (\lambda_k + v\tau_k)^{k-1} (\nu_k - v\tau_k)^{n-k} \, dv \,. \tag{14}$$

$$(y = \lambda_k + v\tau_k)$$

Indem wir den Ausdruck

$$\frac{n!}{k-1!\, n-k!} = \frac{n \,.\, n-1!}{k-1!\, n-k!}$$

nach der Stirlingschen Formel abschätzen, erhalten wir

$$\frac{n!}{k-1!\, n-k!} = \frac{1}{\sqrt{2\pi}} \lambda_k^{-(k-1)} \nu_k^{-(n-k)} \tau_k^{-1} (1 + \delta_n) \tag{15}$$

wo $\delta_n \to 0$ für $n \to \infty$.

Die Formel (14) wird deshalb folgende Gestalt annehmen

$$W_n[z_1, z_2] = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} \left(1 + v\frac{\tau_k}{\lambda_k}\right)^{k-1} \left(1 - v\frac{\tau_k}{\nu_k}\right)^{n-k} dv\,(1+\delta_n). \quad (16)$$

Setzen wir

$$U(v) = \left(1 + v\frac{\tau_k}{\lambda_k}\right)^{k-1} \left(1 - v\frac{\tau_k}{\nu_k}\right)^{n-k},$$

so erhalten wir (indem wir die Logarithmen als Integrale darstellen)

$$\lg U(v) = (k-1)\lg\left(1 + v\frac{\tau_k}{\lambda_k}\right) + (n-k)\lg\left(1 - v\frac{\tau_k}{\nu_k}\right) =$$

$$= \frac{(k-1)\tau_k}{\lambda_k}\int_0^v \frac{dx}{1 + x\frac{\tau_k}{\lambda_k}} - \frac{(n-k)\tau_k}{\nu_k}\int_0^v \frac{dx}{1 - x\frac{\tau_k}{\nu_k}}$$

oder

$$\lg U(v) = -\int_0^v \frac{x\,dx}{\left(1 + x\frac{\tau_k}{\lambda_k}\right)\left(1 - x\frac{\tau_k}{\nu_k}\right)}$$

und schliesslich, indem wir im letzten Integral die Variablentransformation $x = vt$ durchführen, erhalten wir

$$\lg U(v) = -\frac{1}{2}v^2\varphi(v)$$

wo

$$\varphi(v) = \int_0^1 \frac{2t\,dt}{\left(1 + v\frac{t\tau_k}{\lambda_k}\right)\left(1 - v\frac{t\tau_k}{\nu_k}\right)} \quad (17)$$

gesetzt ist.

Folglich ist

$$U(v) = e^{-\frac{1}{2}v^2\varphi(v)} \quad (18)$$

und

$$W_n[z_1, z_2] = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{1}{2}v^2} \varphi(v)\, dv\, (1 + \delta_n). \tag{19}$$

Da die Verhältnisse $\frac{\tau_k}{\lambda_k} = \sqrt{\frac{\nu_k}{k-1}}$ und $\frac{\tau_k}{\nu_k} = \sqrt{\frac{\lambda_k}{n-k}}$ unter unseren Voraussetzungen gegen Null streben (bei $n \to \infty$) so konvergiert $\varphi(v)$ gegen

$$\int_0^1 2\, t\, dt = 1 \tag{20}$$

(was die Gleichung (17) mit Leichtigkeit beweist) und zwar gleichmässig inbezug auf v in jedem endlichen Interwall.

Hieraus ist auf Grund bekannter Eigenschaften der Verteilungsgesetze leicht zu folgern, dass

$$\lim_{n \to \infty} W_n[z_1, z_2] = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{1}{2}v^2} dv \tag{21}$$

gleichmässig inbezug auf z_1, und z_2.

Nehmen wir jetzt an, dass der Rang k bei $n \to \infty$ einen festen Wert bewahrt.

Wollen wir mit Hilfe der Formel (5) in diesem Falle die Wahrscheinlichkeit $W_n[a_1, a_2]$ der Ungleichungen

$$\Psi\left(\frac{k}{n} + \frac{a_1}{n}\right) \leq x_k \leq \Psi\left(\frac{k}{n} + \frac{a_2}{n}\right) \tag{22}$$

$$(-k < a_1 \leq a_2 < n - k)$$

ausdrücken.

Wir haben

$$W_n[a_1, a_2] = \frac{n!}{k-1!\, n-k!} \int_{\frac{k}{n} + \frac{a_1}{n}}^{\frac{k}{n} + \frac{a_2}{n}} y^{k-1} (1-y)^{n-k}\, dy$$

oder

$$W_n[a_1, a_2] = \frac{n!}{k-1!\, n-k!\, n^k} \int_{k+a_1}^{k+a_2} v^{k-1} \left(1 - \frac{v}{n}\right)^{n-k} dv. \tag{23}$$

Wenn wir nun die Stirlingsche Formel anwenden, erhalten wir

$$\frac{n\,!}{n-k\,!\,n^k} = 1 + \delta_n \tag{24}$$

wo $\delta_n \to 0$, bei $n \to \infty$.

Wenn wir ausserdem noch bemerken, dass

$$\left(1 - \frac{v}{n}\right)^{n-k} = e^{-v}\,(1 + \rho_n\,(v)) \tag{25}$$

ist, wo $\rho_n\,(v) \to 0$ wenn v absolut beschränkt ist, erhalten wir folgende Grenzleichung

$$\lim_{n \to \infty} W_n\,[a_1\,,\,a_2] = \frac{1}{k-1\,!}\int_{k+a_1}^{k+a_2} e^{-v}\,v^{k-1}\,dv = \frac{e^{-k}}{\Gamma_{(K)}}\int_{a_1}^{a_2} e^{-x}\,(x+k)^{k-1}\,dx \tag{26}$$

(und zwar gleichmässig inbezug auf die möglichen Werte von a_1 und $a_2 - k < a_1 < a_2$) .

III.

Nehmen wir jetzt an, dass die Funktion $f\,(x) = \Phi'\,(x)$ überall stetig und positiv ist mit Ausnahme vielleicht der Enden des Verteilungsinterwalles (wenn es beschränkt ist), wo sie verschwinden darf.

Es sei ausserdem

$$0 < \underline{\lambda} \leq \lambda_k \leq \bar{\lambda} < 1\,, \tag{27}$$

so dass bei der Vergrösserung von k und n , $\lambda_k = \dfrac{k-1}{n-1}$ beständig in einem Interwall $(\underline{\lambda}\,,\,\bar{\lambda})$ bleibt, dass sich im Inneren des Interwalles (0,1) befindet.

Es sei $$\bar{x}_k = \Psi\,(\lambda_k) \tag{28}$$

so dass

$$\lambda_k = \Phi\,(\bar{x}_k)\;;\; \nu_k = 1 - \Phi\,(\bar{x}_k)\,. \tag{29}$$

Es sei ausserdem

$$\sigma_k = \frac{\tau_k}{f\,(\bar{x}_k)}\,. \tag{30}$$

Wenn wir

$$z = \frac{\Phi(\bar{x}_k + t\sigma_k) - \Phi(\bar{x}_k)}{\tau_k} = A(t) \tag{31}$$

setzen, erhalten wir

$$\Psi(\lambda_k + z\tau_k) = \Psi\left[\Phi(\bar{x}_k + t\sigma_k)\right] = \bar{x}_k + t\sigma_k \tag{32}$$

Bezeichnen wir mit $W_n[t_1, t_2]$ die Wahrscheinlichkeit der Ungleichungen

$$\bar{x}_k + t_1\sigma_k \leq x_k \leq \bar{x}_k + t_2\sigma_k \ldots (t_1 < t_2). \tag{33}$$

Es ist leicht zu ersehen, mit Hilfe von (32), dass die Ungleichungen (33) den Ungleichungen (13) äquivalent sind und dass folglich

$$W_n[t_1, t_2] = W_n[z_1, z_2] \tag{34}$$

wenn

$$z_1 = A(t_1)$$
$$z_2 = A(t_2). \tag{34'}$$

Aus den Bedingungen (27) kann man mit Leichtigkeit schliessen, dass die Zahlen k und $n - k$ unbegrenzt zunehmen.

Mit Berücksichtigung von (21) und (34) können wir die Wahrscheinlichkeit $W_n[t_1, t_2]$ folgendermassen ausdrücken

$$W_n[t_1, t_2] = \frac{1}{\sqrt{2\pi}} \int_{A(t_1)}^{A(t_2)} e^{-\frac{v^2}{2}} dv\,(1 + \delta_n) \tag{35}$$

wo δ_n gleichmässig inbezug auf t_1 und t_2 gegen Null strebt.

Auf Grund von (31) folgern wir aber leicht, dass

$$A(t) = t\frac{\sigma_k}{\tau_k} f(\bar{x}_k + \theta t\sigma_k) = t\frac{f(\bar{x}_k + \theta t\sigma_k)}{f(\bar{x}_k)} \quad (0 \leq \Theta \leq 1). \tag{36}$$

Unter unseren Voraussetzungen über die Funktion $f(x)$ ist $f(\bar{x}_k) > q > 0$ für alle n (wo q eine gewisse positive Konstante vorstellt), da die Grösse $\bar{x}_k$ auf Grund von (27) und (29) sich nicht unbegrenzt den Enden des Verteilungsinterwalles nähern kann, wenn dasselbe begrenzt ist, (bzw. unbeschränkt dem Absolutwert nach zuneh-

men, wenn das letztere unbegrenzt ist) ; daher folgt aus der Formel (36), dass

$$A(t) = t + \rho_n(t) \tag{37}$$

wo $\rho_n(t) \to 0$ bei Zunahme von n $\left(\text{zugleich mit } \sigma_k < \frac{\tau_k}{q}\right)$, wenn t in seiner Veränderung beschränkt bleibt.

Die Gleichungen (37) und (35) ergeben die Schlussfolgerung

$$\lim_{n=\infty} W_n(t_1, t_2] = \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-\frac{v^2}{2}} dv. \tag{38}$$

Wenn also *die Wahrscheinlichkeitsdichte* $f(x)$ *der Grösse* X *stetig ist und nirgends verschwindet, mit Ausnahme vielleicht der Enden des Verteilungsinterwalles (wenn es begrenzt ist) und wenn das Verhältnis* $\frac{k}{n}$, *bei der Zunahme von* n *sich nicht beliebig nahe an* 0 *oder* 1 *herankommt, so nähert sich die Verteilung des* k*-ten Gliedes der Variationsreihe zur normalen mit dem Mittelwert* $\bar{x}_n = \Psi(\lambda_n)$ *und die Streuung* $\sigma_k = \frac{\tau_k}{f(\bar{x}_k)}$.

IV.

Wir wollen jetzt die möglichen Verallgemeinerungen des letzten Satzes betrachten auf die Fälle, wenn die Grösse λ_k sich der Null nähert bei unbegrenzter Zunahme von k und n (wobei $k = 0(n)$). Vor allem muss bemerkt werden, dass in den Fällen, wenn das Verteilungsinterwall von links begrenzt ist und $\Phi(x) \equiv 0$, wenn $x \leq l$ und die Funktion $f(x)$ in der Umgebung von $x = l$ stetig und ist, wenn dabei $f(l) > 0$ ist, der Satz offenbar ohne jede Veränderungen auch auf diesen Fall übertragen werden kann.

Untersuchen wir jetzt den Fall, wenn die Funktion $f(x)$ in der Nähe des Punktes $x = l$ in der Form

$$f(x) = (x - l)^{\alpha} u(x) \tag{39}$$

dargestellt werden kann, wo $\alpha > -1$ ist und $u(x)$ eine stetige Funktion darstellt, wobei $u(l) > 0$ ist.

Vorausgesetzt

$$z = \frac{\Phi(\bar{x}_k + t\,\sigma_k') - \Phi(\bar{x}_k)}{\tau_k} = \beta(t) \tag{40}$$

wo

$$\sigma_k' = \frac{\tau_k}{(\bar{x}_k - l)^\alpha\, u(l)} \tag{41}$$

erhalten wir

$$\Psi(\lambda_k + z\,\tau_k) = \Psi\left[\Phi(\bar{x}_k + t\,\sigma_k')\right] = \bar{x}_k + t\,\sigma_k'. \tag{42}$$

Die Wahrscheinlichkeit $W_n[z_1, z_2]$ der Ungleichungen (13) gleicht auf Grund von (42) der Wahrscheinlichkeit $W_n[t_1, t_2]$ der Ungleichungen

$$\bar{x}_k + t_1\,\sigma_k' \leq x_k \leq \bar{x}_k + t_2\,\sigma_k' \quad (t_1 < t_2)$$

$$W_n[z_1, z_2] = W_n[t_1, t_2] \tag{43}$$

wenn $z_1 = \beta(t_1)$ und $z_2 = \beta(t_2)$.

Folglich werden wir nach (21) erhalten

$$W_n[t_1, t_2] = \frac{1}{\sqrt{2\pi}} \int_{\beta(t_1)}^{\beta(t_2)} e^{-\frac{v^2}{2}}\, dv\,(1 + \delta_n) \tag{44}$$

wo $\delta_n \to 0$ gleichmässig inbezug auf t_1, und t_2.

Aus (40) und (41) erhalten wir aber

$$\beta(t) = t\,\frac{f(\bar{x}_k + \theta\,t\,\sigma_k')}{(\bar{x}_k - l)^\alpha\, u(l)} = t\left(\frac{\bar{x}_k - l + \theta\,t\,\sigma_k'}{\bar{x}_k - l}\right)^\alpha \frac{u(\bar{x}_k + \theta\,t\,\sigma_k')}{u(l)}, \quad (0 < \theta < 1)$$

oder

$$\beta(t) = t\left[1 + \theta\,t\,\frac{\tau_k}{(\bar{x}_k - l)^{1+\alpha}\, u(l)}\right]^\alpha \frac{u(\bar{x}_k + \theta\,t\,\sigma_k')}{u(l)}. \tag{45}$$

Andererseits haben wir laut Formel (29)

$$\lambda_k = \Phi(\bar{x}_k) = \int_l^{\bar{x}_k} (x - l)^\alpha\, u(x)\, dx = [u(l) + \gamma_n]\,\frac{(\bar{x}_k - l)^{1+\alpha}}{1 + \alpha} \tag{46}$$

wo $\gamma_n \to 0$ bei $n \to \infty$ und $\bar{x}_k \to l$.

Nach (46) ist

$$(\bar{x}_k - l)^{1+\alpha} = \frac{\lambda_k (1+\alpha)}{u(l) + \gamma_n}. \tag{47}$$

Da jedoch $\frac{\tau_k}{\lambda_k} = \sqrt{\frac{\nu_k}{k-1}}$ gegen Null strebt bei der Zunahme von k (zugleich mit der Vergrösserung von n), so wird auch die Grösse

$$\frac{\tau_k}{(\bar{x}_k - l)^{1+\alpha}} = \frac{\tau_k}{\lambda_k (1+\alpha)} [u(l) + \gamma_n]$$

gegen Null konvergieren.

Zugleich wird auch die Grösse

$$\sigma_k' = \frac{\tau_k}{(\bar{x} - l)^{\alpha} u(l)}$$

unendlich klein sein, weshalb

$$\frac{u(\bar{x}_k + \theta t \sigma_k')}{u(\bar{x}_k)} = 1 + \eta_n$$

sein wird, wo $\eta_n \to 0$, bei $n \to \infty$.

Der Ausdruck $\beta(t)$ (45) wird aus diesem Grunde die Form

$$\beta(t) = t + \rho_n(t) \tag{48}$$

erhalten, wo $\rho_n(t)$ gleichmässig gegen Null strebt, wenn t absolut beschränkt bleibt.

Hieraus ziehen wir, wie früher, die Folgerung, dass

$$\lim_{n \to \infty} W_n[t_1, t_2] = \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-\frac{v^2}{2}} dv. \tag{49}$$

Wir bemerken, dass in dem untersuchten Falle der Mittelwert und die Streuung sich, annähernd folgendermassen ausdrücken

$$\bar{x}_k \approx l + \left\{ \frac{\lambda_k (1+\alpha)}{u(l)} \right\}^{\frac{1}{1+\alpha}} \tag{50}$$

$$\sigma_k' \approx \frac{\tau_k}{(\bar{x}_k - l)^{\alpha} u(l)} = \frac{\tau_k}{[(1+\alpha)\lambda_k]^{\frac{\alpha}{1+\alpha}} [u(l)]^{\frac{\alpha}{1+\alpha}}} \tag{50'}$$

Auf analoge Weise kann die normale Grenzverteilung auch in dem Falle festgestellt werden, wenn $\lambda_k \to 1$, bei Zunahme von n (bei $\nu_k \to 0$) und wenn die Funktion $f(x)$ im rechten Ende des Verteilungsinterwalles verschwindet oder von der Ordnung $\alpha > -1$ unendlich wird.

Ebenso einfach ist die Untersuchung des Falles, wenn das Verteilungsinterwall unbegrenzt ist, wenn man bestimmte Voraussetzungen über die Ordnung der Kleinheit der Funktion $f(x)$ bei unbegrenzter Zunahme von x macht.

Es sei, wie früher, $\lambda_k \to 0$, bei $n \to \infty$ $(k \to \infty\ ,\ k = 0\,(n))$; die Funktion $f(x)$ soll aber für absolut grosse negative Werte von x durch den Ausdruck

$$f(x) = \frac{u(x)}{|x|^{1+\beta}} \tag{51}$$

dargestellt sein, wo $\beta > 0$, die Funktion $u(x) > 0$ und stetig, und dabei $\lim_{x \to -\infty} u(x) = c > 0$ ist.

Wenn man annimmt, dass in diesem Falle

$$\sigma_k'' = \frac{\tau_k}{c} |\bar{x}_k|^{1+\beta} \tag{52}$$

ist, so kann man mit Hilfe der Formel (21) Wahrscheinlichkeit $W_n[t_1, t_2]$ der Ungleichungen

$$\bar{x}_k + t_1 \sigma_k'' \leq x_k \leq \bar{x}_k + t_2 \sigma_k''$$

folgendermassen ausdrücken :

$$W_n[t_1, t_2] = \frac{1}{\sqrt{2\pi}} \int_{D(t_1)}^{D(t_2)} e^{-\frac{v^2}{2}} dv\ (1 + \delta_n) \tag{53}$$

wo

$$D(t) = \frac{\Phi(\bar{x}_k + t\,\sigma_k'') - \Phi(\bar{x}_k)}{\tau_k} = z \tag{54}$$

$t_1 , < t_2$ und $\delta_n \to 0$ gleichmässig bezüglich t_1, und t_2 $(n \to \infty)$

Auf dem Wege derselben Erwägungen, wie im vorhergehenden Falle, finden wir

$$\lambda_k = \int_{-\infty}^{\bar{x}_k} \frac{u(x)\,dx}{|x|^{1+\beta}} = \frac{c + \varepsilon_n}{\beta\, x_k^{\beta}}$$

wo

$$\varepsilon_n \to 0$$

und hieraus

$$|x_k|^\beta = \frac{c + \varepsilon_n}{\beta \lambda_k} \tag{55}$$

Wenn wir die Gleichungen (54) und (52) berücksichtigen, so erhalten wir

$$D(t) = t \frac{\sigma_k''}{\tau_k} f(\bar{x}_k + \theta t \sigma_k'') = \frac{t}{c} |\bar{x}_k|^{1+\beta} f(\bar{x}_k + \theta t \sigma_k'') \,; (0 < \theta < 1)$$

oder wegen (51)

$$D(t) = t \frac{u(\bar{x}_k + \theta t \sigma_k'')}{c} = \frac{1}{\left(1 + \theta \frac{t \sigma_k''}{|\bar{x}_k|}\right)^{1+\beta}} \tag{56}$$

Indem wir bemerken, dass $\bar{x}_k \to -\infty$ und dass nach (52) und (55)

$$\frac{\sigma_k''}{|\bar{x}_k|} = \frac{\tau_k}{c} |\bar{x}_k|^\beta = \frac{\tau_k}{\lambda_k} \frac{c + \varepsilon_n}{c\,\beta}$$

$\left(\text{zugleich mit } \sqrt{\frac{\nu_k}{k-1}} = \frac{\tau_k}{\lambda_k}\right)$ ist, finden wir (56)

$$D(t) = t + \rho_n(t)\,, \tag{57}$$

wo $\rho_n(t) \to 0$, gleichmässig inbezug auf t.

Auf diese Weise gelangen wir auf Grund von (54) und (57) zur Schlussfolgerung

$$\lim_{n \to \infty} W_n[t_1, t_2] = \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-\frac{v^2}{2}} dv \ldots .$$

Zu bemerken ist, dass die Grösse der Streuung

$$\sigma_k'' = \frac{\tau_k |\bar{x}_k|^{1+\beta}}{c} = \frac{\tau_k}{\frac{1+\beta}{\beta}} Q$$
$$\lambda_k$$

wo $Q = \frac{1}{c}\left(\frac{c+\varepsilon_n}{\beta}\right)^{\frac{1+\beta}{\beta}}$ ist, nicht gegen Null zu konvergieren braucht, bei der Zunahme von n und k.

Es sei, zum Beispiel $\beta = 1$.

In diesem Fall ist

$$\sigma_k'' = \frac{\tau_k}{\lambda_k^2} Q = \sqrt{\frac{\nu_k (n-1)^2}{(k-1)^3}}\, Q$$

Ist $\frac{k}{n^{\frac{2}{3}}} \to 0$, so ist $\sigma_k'' \to \infty$.

Auf ebendieselbe Weise kann man die normale Grenzverteilung auch in anderen Fällen feststellen, wenn die Funktion $f(x)$, bei unbegrenztem Anwachsen von x, wie eine Exponentialfunktion e^{-ax} oder $e^{-ax} x^{\beta} \ldots$ u. s. w. abnimmt. Ohne diese Frage in ihrem ganzen Umfange zu untersuchen, wollen wir uns nur mit dem Falle befassen, wenn die Anfangsverteilung der Grösse x selbst normal ist und

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (E(x) = 0\,,\ \sigma = 1)\,.$$

Es sei, wie früher, $\lambda_k \to 0$ ($k = 0(n)$ und unendlich gross). Man setze

$$\sigma_k = \frac{\tau_k}{f(\bar{x}_k)}\,, \tag{58}$$

wo $\bar{x}_k$ durch die Gleichung

$$\Phi(\bar{x}_k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\bar{x}_k} e^{-\frac{x^2}{2}}\, dx = \lambda_k \quad \text{bestimmt wird.}$$

Offenbar ist $\bar{x}_k \to -\infty$ mit $\lambda_k \to 0$.

Es ist zu beweisen, dass die Funktion

$$E(t) = t\,\frac{\Phi(\bar{x}_k + t\,\sigma_k) - \Phi(\bar{x}_k)}{\tau_k} \tag{59}$$

bei ihrer Vergrösserung n gleichmässig gegen t strebt.

Aus (59) und (58) folgt, dass

$$E(t) = t\,\frac{f(\bar{x}_k + \theta\, t\,\sigma_k)}{f(\bar{x}_k)} = t\, e^{-\theta\, t\,\sigma_k\, x_k - \frac{(\theta\, t\,\sigma_k)^2}{2}} \tag{60}$$

ist. Die Grössen σ_k und $\sigma_k \left|\bar{x}_k\right|$ konvergieren aber gegen Null; denn

$$\sigma_k = \frac{\tau_k}{f(\bar{x}_k)} = \frac{\tau_k}{\lambda_k} \frac{\Phi(\bar{x}_k)}{f(\bar{x}_k)} = \sqrt{\frac{\nu_k}{k-1}} \frac{\Phi(\bar{x}_k)}{f(\bar{x}_k)} \tag{61}$$

Andererseits ist bekanntlich

$$\Phi(\bar{x}_k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\bar{x}_k} e^{-\frac{x^2}{2}} dx < \frac{e^{-\frac{\bar{x}_k^2}{2}}}{\sqrt{2\pi}\left|\bar{x}_k\right|} = \frac{f(\bar{x}_k)}{\left|\bar{x}_k\right|} .$$

Darum folgt auf Grund von (61), dass

$$\sigma_k \left|\bar{x}_k\right| < \sqrt{\frac{\nu_k}{k-1}} \tag{62}$$

udn dass

$$\sigma_k \left|\bar{x}_k\right| \to 0$$

$$\text{und} \qquad \sigma_k \to 0 . \tag{63}$$

Nach (63) und auf Grund von (60) folgt, dass

$$E(t) = t + \rho_n(t) , \tag{64}$$

wo $\rho_n(t) \to 0$ für $n \to 0$ gleichmässig inbezug auf t in jedem endlichen Interwall.

Hieraus erhalten wir für die Wahrscheinlichkeit $W_n(t_1, t_2]$ der Ungleichungen

$$\bar{x}_k + t_1 \sigma_k \leq x_k \leq \bar{x}_k + t_2 \sigma_k ,$$

welche den Ungleichungen (13) äquivalent sind, bei $Z_1 = E(t_1)$ und $Z_2 = E(t_2)$,

$$\lim_{n \to \infty} W_n[t_1, t_2] = \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-\frac{v^2}{2}} dv .$$

V.

Wir wollen jetzt den Fall untersuchen, wenn der Rang des Gliedes der Variationsreihe (1) (vom Anfang oder vom Ende gezählt) unverändert bleibt bei der Zunahme von n. In diesem Falle haben wir die Grenzgleichung (26).

Man setze

$$\bar{x}_k = \Psi\left(\frac{k}{n}\right), \tag{64}$$

$$\text{so dass } \Phi(\bar{x}_k) = \frac{k}{n}$$

und $a = \Phi(\bar{x}_k + t\nu) - \Phi(\bar{x}_k) = D(t)$ (wo ν eine positive Zahl ist, die wir später wählen werden),

$$a_1 = D(t_1), \; a_2 = D(t_2); \tag{65}$$

wir wollen die Wahrscheinlichkeit $W_n[t_1, t_2]$ der Ungleichungen

$$\bar{x}_k + t_1\nu \leq x_k < \bar{x}_k + t_2\nu \qquad (t_1 < t_2) \tag{66}$$

bestimmen.

Es ist leicht ersichtlich, dass

$$\Psi\left(\frac{k}{n} + \frac{a_1}{n}\right) = \Psi[\Phi(\bar{x}_k + t\nu)] = \bar{x}_k + t\nu$$

ist; aus diesem Grunde sind die Ungleichungen (66) den Ungleichungen (22) äquivalent, und deshalb ist

$$W_n[t_1, t_2] = W_n[a_1, a_2], \quad \text{wo} \quad W_n[a_1, a_2]$$

die Wahrscheinlichkeit der Ungleichungen (22) bedeutet.

Indem wir den Rang von k als konstant voraussetzen, können wir mit Hilfe der Formel (26) die Gleichung aufsetzen

$$W_n[t_1, t_2] = \frac{e^{-k}}{\Gamma(k)} \int_{D(t_1)}^{D(t_2)} e^{-x}(x+k)^{k+1}\, dx\,(1+\delta_n) \tag{67}$$

wo $\delta_n \to 0$, gleichmässig inbezug auf t_1 und t_2. Da jedoch $n = \frac{\Phi(\bar{x}_k)}{k}$ ist, werden wir dem Ausdruck $D(t)$ (64) folgende Form geben:

$$D(t) = k\left\{\frac{\Phi(\bar{x}_k + t\nu)}{\Phi(\bar{x}_k)} - 1\right\}. \tag{68}$$

Wenn wir für wachsendes n (bei konstantem k und bei entsprechend gewähltem ν) finden, dass

$$\frac{\Phi(\bar{x}_k + t\nu)}{\Phi(\bar{x}_k)} \to q(t), \quad \text{(gleichmässig inbezug auf } t\text{)}, \text{ wo } q(t)$$

irgend eine bestimmte Funktion von t vorstellt, so werden wir auf Grund von (68) erhalten, dass

$$\lim_{n \to \infty} D(t) = k(q(t) - 1)$$

ist, und auf Grund von (67)

$$\lim_{n \to \infty} W_n[t_1, t_2] = \frac{e^{-k}}{\Gamma(k)} \int_{k[q(t_1)-1]}^{k[q(t_2)-1]} e^{-x}(x+k)^{k-1} dx. \tag{69}$$

Nehmen wir zuerst an, dass das Verteilungsinterwall von links beschränkt ist, so dass $\Phi(x) \equiv 0$ für $x \leq l$, und dass die Wahrscheinlichkeitsdichte in der Form

$f(x) = (x - l)^{\alpha} u(x)$ dargestellt werden kann, wobei

$$\alpha > -1 \quad \text{und} \quad u(l) > 0 \quad \text{ist.}$$

Offensichtlich ist $\frac{k}{n} \to 0$ und $\bar{x}_k \to l$.

Wir setzen $\nu = \bar{x}_k - l$ (70) und wollen den Limes des Ausdrucks

$$\frac{\Phi[\bar{x}_k + t(\bar{x}_k - l)]}{\Phi(\bar{x}_k)} \quad (t > -1)$$

für $\bar{x}_k \to l$ suchen.

Nach der l'Hospitalschen Regel erhalten wir

$$\lim_{\bar{x}_k \to l} \frac{\Phi[\bar{x}_k + t(\bar{x}_k - l)]}{\Phi(\bar{x}_k)} = \lim_{\bar{x}_k \to l} \frac{f[\bar{x}_k + t(\bar{x}_k - l)]}{f(\bar{x}_k)}(1+t) =$$

$$= \lim_{\bar{x}_k \to l} \frac{U[\bar{x}_k + t(\bar{x}_k - l)]}{u(\bar{x}_k)}(1+t)^{1+\alpha} = (1+t)^{1+\alpha} = q(t). \tag{70}$$

Folglich ist in diesem Falle auf Grund von (68)

$$\lim_{n \to \infty} W_n [t_1 , t_2] = \frac{e^{-k}}{\Gamma (k)} \int_{k\,[(1+t_1)^{1+\alpha}-1]}^{k\,[(1+t_2)^{1+\alpha}-1]} e^{-x} (x + k)^{k-1} \, dx =$$

$$= \frac{k^k (1 + \alpha)}{\Gamma (k)} \int_{t_1}^{t_2} e^{-k(1+t)^{1+\alpha}} (1 + t)^{(k-1)(1+\alpha)+\alpha} \, dt \qquad (71)$$

hierbei ist

$$W_n [t_1 , t_2] = W \left[\bar{x}_k + t_1 (\bar{x}_k - l) \leq x_k \leq \bar{x}_k + t_2 (\bar{x}_k - l)\right]$$

und $t > -1$.

Nun wollen wir annähernd die Werte der mathematischen Erwartung und der Streuung der Grösse $\bar{x}_k$ in unserem Falle bestimmen. Aus dem, was früher bewiesen war, folgt, dass

$$I_1 = E \left\{ \frac{x_k - \bar{x}_k}{\bar{x}_k - l} \right\} \approx \frac{k^k (1 + \alpha)}{\Gamma (k)} \int_{-1}^{\infty} e^{-k(1+t)^{1+\alpha}} (1 + t)^{(k-1)(1+\alpha)+\alpha} t \, dt =$$

$$= \frac{\Gamma\left(k + \frac{1}{1+\alpha}\right)}{k^{\frac{1}{1+\alpha}} \Gamma (k)} - 1 \qquad (72)$$

$$I_2 = E \left\{ \frac{x_k - \bar{x}_k}{\bar{x}_k - l} \right\}^2 \approx \frac{k^k (1 + \alpha)}{\Gamma (k)} \int_{-1}^{+\infty} e^{-k(1+t)^{1+\alpha}} (1 + t)^{(k-1)(1+\alpha)+\alpha} t^2 \, dt =$$

$$= \frac{\Gamma\left(k + \frac{2}{1+\alpha}\right)}{k^{\frac{2}{1+\alpha}} \Gamma (k)} - 2 \frac{\Gamma\left(k + \frac{1}{1+\alpha}\right)}{\Gamma (k) \, k^{\frac{1}{1+\alpha}}} + 1 \qquad (72')$$

Ausserdem erhält man leicht auf Grund analoger Erwägungen wie in Abschnitt IV, dass

$$\bar{x}_k = l + \left[\frac{k}{n} \frac{(1 + \alpha)}{u (l)}\right]^{\frac{1}{1+\alpha}} . \qquad (73)$$

Aus (72) und (73) folgt

$$E (x_k) = x_k^0 = \bar{x}_k + (\bar{x}_k - l) I_1 \approx l + \left[\frac{1 + \alpha}{n \, . \, u (l)}\right]^{\frac{1}{1+\alpha}} \frac{\Gamma\left(k + \frac{1}{1+\alpha}\right)}{\Gamma (k)} \qquad (74)$$

Ebenso folgt aus (73), dass

$$E\,(x_k - x_0)^2 = (x_k - l)^2\,[I_2 - I_1^2] \approx \left[\frac{1+\alpha}{n\,u\,(l)}\right]^{\frac{2}{1+\alpha}}$$

$$\left[\frac{\Gamma\left(k+\frac{2}{1+\alpha}\right)}{\Gamma\,(k)} - \frac{\Gamma^2\left(k+\frac{1}{1+\alpha}\right)}{\Gamma^2\,(k)}\right]. \tag{75}$$

Im Spezialfalle, wenn $\alpha = 0$ ist, wird die Funktion $f\,(x)$, für $x = l$, von Null verschieden und $f\,(x) = U\,(x)$ sein. Die Formeln (73), (74) und (75) ergeben

$$E\,(\bar{x}_k) = x_k^0 = \bar{x}_k = l + \frac{k}{n\,f\,(l)}$$

$$E\,(x_k - \bar{x}_k)^2 = \approx \frac{k}{n^2\,f^2\,(l)}.$$

Dabei haben wir auf Grund von (71) in diesem Falle

$$\lim_{n\to\infty} W_n\,[t_1, t_2] = \frac{k\,e^{-k}}{\Gamma\,(k)} \int_{t_1}^{t_2} e^{-k\,t}\,(1+t)^{k-1}\,\ddot{a}\,t \tag{76}$$

Folglich hat in diesem Falle die Grenzverteilung die Form einer Pearson'schen Kurve vom Typus III mit dem Mittelwert M, der Mode M_0, der Streuung σ und der Assymmetrie γ, welche nach folgenden Formeln bestimmt werden

$$\left.\begin{aligned} M &= l + \frac{k}{n\,f\,(l)} & \sigma &= \frac{\sqrt{k}}{n\,f\,(l)} \\ M_0 &= l + \frac{k-1}{n\,f\,(l)} & \gamma &= \frac{1}{\sqrt{k}} \end{aligned}\right\} \tag{77}$$

Untersuchen wir, z. B., die Verteilung des absolut kleinsten Fehlers ε_1 in n Beobachtungen. Unter der Voraussetzung, dass die Fehler ε ein normales Verteilungsgesetz haben, so dass

$$W\,[|\varepsilon| < x] = \frac{2\,h}{\sqrt{\pi}} \int_0^x e^{-h^2 x^2}\,d\,x$$

$$f(x) = \frac{2h}{\sqrt{\pi}} e^{-h^2 x^2}; f(0) = \frac{2h}{\sqrt{\pi}} \quad (l = 0, k = 1)$$

ist, werden wir nach den Formeln (77) feststellen, dass

$$M_{\varepsilon_1} = \frac{\sqrt{\pi}}{2hn}$$

$$\sigma_{\varepsilon_1} = \frac{\sqrt{\pi}}{2hn}$$

$$W[\varepsilon_1 - M_{\varepsilon_1} < t\,\sigma_{\varepsilon_1}] = 1 - e^{-(t+1)}.$$

Nun wollen wir die Verteilung des Gliedes x_{n-k+1} (des k-ten vom Ende) der Reihe (I) betrachten, indem wir k bei zunehmenden n als konstant annehmen, und das Verteilungsinterwall als von rechts begrenzt, so dass $\Phi(x) \equiv 1$ für $x < l_1$. Nehmen wir ausserdem an, dass die Wahrscheinlichkeitsdichte $f(x)$ folgende Forme besitzt

$$f(x) = (l_1 - x)^\alpha u(x)$$

wo $u(x)$ eine stetige Funktion vorstellt und $u(l_1) > 0$ ist.

Auf Grund der Bemerkung, die wir am Ende des Abschnittes I gemacht haben, genügt es zur Feststellung der Grenzverteilung von x_{n-k+1}, die Verteilung des Gliedes x_k' zu untersuchen (wobei k konstant ist und $n \to \infty$), in der Variationsreihe zu der Grösse $-x$, die in unserem Falle die Wahrscheinlichkeitsdichte

$$f(-\xi) = (l_1 + \xi)^\alpha u(-\xi) \quad (\alpha > -1)$$

besitzt.

Bestimmen wir die Grössen $\bar{x}_k'$ und $\bar{x}_{n-k+1}$ aus den Gleichungen

$$\frac{k}{n} = 1 - \Phi(-\bar{x}_k') \qquad 1 - \frac{k}{n} = \Phi(\bar{x}_{n-k+1}). \tag{78}$$

Es ist offenbar

$$\bar{x}_{n-k+1} = -\bar{x}_k'.$$

Aus der Gleichung (8) folgt

$$V_{n-k+1}[\bar{x}_{n-k+1} + t(l_1 - \bar{x}_{n-k+1})] = 1 - V_k'[\bar{x}_k' - t(l_1 + \bar{x}_k')] \tag{79}$$
$$(t < 1).$$

Laut (76) haben wir

$$\lim_{n\to\infty} V_k'[\bar{x}_k' - t(l_1 + \bar{x}_k')] = \frac{k^k(1+\alpha)}{\Gamma(k)} \int_{-1}^{-t} e^{-k(1+t)^{1+\alpha}} (1+t)^{(k-1)(1+\alpha)+\alpha} dt$$

und folglich auf Grund (79)

$$\lim_{n\to\infty} W_n'[x_{n-k+1} < \bar{x}_{n-k+1} + t(l_1 - \bar{x}_{n-k+1})] =$$

$$= \lim_{n\to\infty} V_{n-k+1}[\bar{x}_{n-k+1} + t(l_1 - \bar{x}_{n-k+1})] =$$

$$= \frac{k^k(1+\alpha)}{\Gamma(k)} \int_{-t}^{\infty} e^{-k(1+t)^{1+\alpha}} (1+t)^{(k-1)(1+\alpha)+\alpha} dt =$$

$$= \frac{k^k(1+\alpha)}{\Gamma(k)} \int_{-\infty}^{t} e^{-k(1-t)^{1+\alpha}} (1-t)^{(k-1)(1+\alpha)+\alpha} dt . \qquad (80)$$

Die Mathematische Erwartung $E(x_{n-k+1})$ und die Streuung finden wir mit Hilfe der Formel (10).

Auf dieselbe Weise können wir die Grenzverteilung für die Glieder x_k und x_{n-k+1} bei Konstanten k auch in andere Fällen bestimmen, indem wir das Verteilungsinterwall unbegrenzt voraussetzen und eine gewisse Ordnung für das Abnehmen der Funktion $f(x)$ bei $|x| \to \infty$ postulieren.

Die Verteilungskurven der Grenzverteilungen erhält man durch eine gewisse Transformation der Kurve des III Typus von Pearson.

Wir wollen hier genauer nur den Fall betrachten, wenn die Ausgangsverteilung selbst normal ist

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (E(x) = 0 , \sigma = 1)$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}} dx .$$

In diesem Falle wird $\bar{x}_k$ durch die Gleichung

$$\frac{k}{n} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\bar{x}_k} e^{-\frac{v^2}{2}} dv$$

bestimmt, und offenbar ist $\bar{x}_k \to -\infty$ für $n \to \infty$.

Indem wir ebenso wie früher vorgehen, finden wir den Limes für den Ausdruck.

$$\frac{\Phi\left(\bar{x}_k + \frac{t}{|\bar{x}_k|}\right)}{\Phi(\bar{x}_k)}$$

bei der Annahme, dass $\nu = \frac{1}{|\bar{x}_k|}$ ist, und dass $\bar{x}_k \to -\infty$.

Mittels Anwendung der l'Hospitalschen Regel erhalten wir mit Leichtigkeit

$$\lim_{\bar{x}_k \to -\infty} \frac{\Phi\left(\bar{x}_k + \frac{t}{|\bar{x}_k|}\right)}{\Phi(\bar{x}_k)} = \lim_{x_k \to -\infty} \frac{f\left(\bar{x}_k + \frac{t}{|\bar{x}_k|}\right)}{f(x_k)}\left(1 + \frac{t}{|\bar{x}_k|}\right) =$$

$$= \lim_{\bar{x}_k \to -\infty} e^{t - \frac{t^2}{(\bar{x}_k)^2}}\left(1 + \frac{t}{|\bar{x}_k|}\right) = e^t = q(t) \tag{81}$$

Folglich haben wir auf Grund von (69)

$$\lim_{n \to \infty} W_n\left[\bar{x}_k + \frac{t_1}{|\bar{x}_k|} \leq x_k \leq \bar{x}_k + \frac{t_2}{|\bar{x}_k|}\right] = \frac{e^{-k}}{\Gamma(k)} \int_{k(e^{t_1}-1)}^{k(e^{t_2}-1)} e^{-x}(x+k)^{k-1}\,dx =$$

$$= \frac{k_k}{\Gamma(k)} \int_{t_1}^{t_2} e^{-k e^t + kt}\,dt \tag{82}$$

Die mathematische Erwartung und die Streuung (wie auch die anderen Momente) von x_k können annähernd bestimmt werden.

Auf Grund von (82) ist

$$I_1 = E\left[|x_k|(x_k - \bar{x}_k)\right] \approx \frac{k^k}{\Gamma(k)} \int_{-\infty}^{+\infty} e^{-k e^t + kt}\, t\,dt =$$

$$= \frac{1}{\Gamma(k)} \int_0^{\infty} e^{-\gamma}\gamma^{k-1} \lg\frac{\gamma}{k}\,d\gamma = \frac{\Gamma'(k)}{\Gamma(k)} - \lg k$$

$$I_2 = E\left[\bar{x}_k^2 (x_k - \bar{x}_k)^2\right] \approx \frac{k^k}{\Gamma(k)} \int_{-\infty}^{+\infty} e^{-k e^t + kt}\, t^2\,dt =$$

$$= \frac{k^k}{\Gamma(k)} \int_0^{+\infty} e^{-kz} z^{k-1} \lg^2 z\,dz = \frac{\Gamma''(k)}{\Gamma(k)} - 2\lg k \frac{\Gamma'(k)}{\Gamma(k)} + \lg^2 k\,.$$

Hieraus folgt

$$x_k^0 = E(x_k) = \bar{x}_k + \frac{I_1}{|\bar{x}_k|} = \bar{x}_k + \left(\frac{\Gamma'(k)}{\Gamma(k)} - \lg k\right)\frac{1}{|\bar{x}_k|} \tag{83}$$

$$E(x_k - x_k^0)^2 = \frac{1}{\bar{x}_k^2}[I_2 - I_1^2] = \frac{1}{\bar{x}_k^2}\left[\frac{\Gamma''(k)}{\Gamma(k)} - \frac{\Gamma'^2(k)}{\Gamma^2(k)}\right]. \tag{84}$$

Im Sonderfalle, wenn $k = 1$ ist, erhalten wir für die Verteilung des kleinsten Gliedes x_1, der Reihe (1)

$$x_1^0 = E(x_1) = \bar{x}_1 - \frac{c}{|x_1|} \tag{85}$$

(wo $c = 0{,}577215\ldots$ die Eulersche Konstante bedeutet).

$$\sigma_1^2 = E(x_1 - x_1^0)^2 = \frac{1}{|x_1|^2}\left(\frac{\Gamma''(1)}{\Gamma(1)} - c^2\right) = \frac{\pi_1^2}{6\bar{x}_1^2} \tag{86}$$

wo $\bar{x}_1$, aus der Gleichung

$$\Phi(\bar{x}_1) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\bar{x}_1} e^{-\frac{x^2}{2}}\,dx = \frac{1}{n} \tag{87}$$

bestimmt wird. Ausserdem haben wir auf Grund von (32) in diesem Falle

$$\lim_{n\to\infty} W_n\left[\bar{x}_1 + \frac{t_1}{|\bar{x}_1|} \leq x_1 \leq \bar{x}_1 + \frac{t_2}{|\bar{x}_1|}\right] = \int_{t_1}^{t_2} e^{-e^t + t}\,dt = e^{-e^{t_1}} - e^{-e^{t_2}} \tag{88}$$

Demnach

$$\lim_{n\to\infty} V_1\left(\bar{x}_1 + \frac{t}{|x_1|}\right) = 1 - e^{-e^t}. \tag{88'}$$

Auf Grund (8) ist die Grenzform der Verteilung des maximalen Gliedes x_n leicht zu ermitteln *.

$$\lim V_n\left[\bar{x}_n + \frac{t}{|\bar{x}_n|}\right] = 1 - \lim V_1\left(\bar{x}_1 - \frac{t}{|\bar{x}_1|}\right) = e^{-e^{-t}} \tag{89}$$

(*) Vergl. BRUNO DE FINETTI. *Sulla legge di probabilità degli estremi.* « Metron » V. IX, No. 1, 1931, p. 127-138.

J. O. IRWIN

Tests of Significance for differences between percentages based on small numbers *

If we wish to compare the significance of the difference between the percentages of marked individuals in two samples it is usual to compute the standard error of the difference from the formula

$$\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

where the samples contain respectively n_1 , n_2 individuals and p is the proportion of marked individuals obtained by combining the two given samples. This method gives the same result as applying the χ^2 test to the corresponding fourfold table, and may be used without serious error when the expected numbers in any cell of the fourfold table are greater than 5.

But it is a matter of some little difficulty to know how to treat pairs of samples in which this condition is not fulfilled ; three possible methods suggest themselves. For the sake of definiteness let us suppose our data when arranged in fourfold table form are as follows :

(*) This paper was concluded in May 1933, but its publication has been unavoidably delayed. Meanwhile a paper dealing with the same subject, in some respects more completely, has been published by F. YATES (" J. Roy Stat. Soc. " Suppl. I, 2 p. 217). YATES' correction to the usual method of calculating χ^2, consists in deducting one half from the deviations between observed and expected frequencies. The effects of applying it is shown in Table X of this paper, it will be seen as a rule to give good agreement with method (3) at the critical levels.

	Marked	*Not Marked*	
Sample I . . .	a	b	$a + b$
Sample II . . .	c	d	$c + d$
	$a + c$	$b + d$	$a + b + c + d$

and let us suppose $\frac{b}{a+b} > \frac{d}{c+d}$

or $bc > ad$

This is really no restriction as it could always be effected by suitably arranging the table.

(1) We could suppose our sample sizes $a + b$ and $c + d$ to be fixed but allow our other marginal totals to vary from sample to sample, we could then enumerate the $(a + b + 1)(c + d + 1)$ possible tables subject to this condition and work out from the exact binomial expressions the probabilities of each table arising by chance from a universe in which there was a proportion p of masked individuals. We could then take $\frac{a+c}{a+b+c+d}$ as our estimate of p, and calculate the probability of getting a table as improbable as, or more improbable than that observed. This procedure, however, would be impracticable owing to the large number of tables to be enumerated.

(2) We might take $\frac{a+c}{a+b+c+d}$ as our estimate of p and work out the probability that Sample I should contain b or more unmarked individuals and sample II, d or less, i. e.

$$\left(\sum_{b}^{a+b} {}^{a+b}C_r\, p^{a+b-r}\, q^r\right)\left(\sum_{c}^{c+d} {}^{c+d}C_r\, p^r\, q^{c+d-r}\right)$$

The disadvantage of this method is that it does not account for all the tables which are less probable than that observed, for example if we consider the following table (considered in detail below)

TABLE I.

	Marked	*Not Marked*	
Sample I . . .	26	2	28
Sample II . . .	61	2	63
	87	4	91

the percentages of unmarked individuals are 7.14 per cent and 3.17 per cent. respectively, a difference of 3.97 per cent. Method (2) would give us a probability of $(1 - 28\,p^{27}q - p^{28}) \times (p^{63} + 63\,p^{62}q + 1953\,p^{61}q^2)$ with $p = \frac{87}{91}$ but such a table as

TABLE II.

	Marked	*Not Marked*	
Sample I . . .	3	25	28
Sample II . . .	51	12	63
	54	37	91

which makes the difference between the two percentages $89 - 19 = 70$ per cent, is less probable than that observed, but would not be taken into account in the calculation, method (2) therefore gives odds which are rather too low and significance would tend to be overestimated.

(3) We can suppose all the marginal totals to be fixed, it is then easy to enumerate all the possible tables which can arise work out the probability of each and thence to deduce the probability of a table arising by chance as probable or less probable than that observed. The process can best be illustrated by numerical examples.

Example (I). — Suppose we have the following fourfold table.

TABLE III.

	Marked	*Not Marked*	
Sample I . . .	2	4	6
Sample II . . .	6	0	6
	8	4	12

We may estimate the proportion marked by $p = \frac{8}{12} = \frac{2}{3}$

Suppose we now take samples of fixed size (6 each) from a universe in which p is the probability of a marked individual, the chance of getting 8 marked and 4 unmarked individuals is ${}^{12}C_4\, p^8 q^4$ and we may easily enumerate the fourfold tables which fulfil this condition by supposing r individuals in sample I to be marked, when the constitution of the table is easily seen to be

r	$6-r$	6
$8-r$	$r-2$	6
8	4	12

All possible tables are then obtained by giving r integral value from 2 to 6.

The chance of obtaining the above table is evidently

$$ {}^6C_r\, p^r q^{6-r} \times {}^6C_{8-r}\, p^{8-r} q^{r-2} = {}^6C_r \times {}^6C_{8-r}\, p^8 q^4 $$

We thus have the following probabilities for the possible tables.

2	${}^6C_2 \times {}^6C_6\, p^8 q^4 = 15\, p^8 q^4$
3	${}^6C_3 \times {}^6C_5$ » » $= 120$ » »
4	${}^6C_4 \times {}^6C_4$ » » $= 225$ » »
5	${}^6C_5 \times {}^6C_3$ » » $= 120$ » »
6	${}^6C_6 \times {}^6C_2$ » » $= 15$ » »
Total	${}^{12}C_4\, p^8 q^4 = 495\, p^8 q^4$

Thus if $r = 2$, the chance of a table arising with as small or a smaller number of marked individuals in sample I is $\frac{15}{495} = \frac{1}{33}$ and the chance of an equally probable or less probable table arising is $\frac{30}{495} = \frac{2}{33}$. These two probabilities are respectively .030 and .061.

From the fourfold table we find

$$ \chi^2 = 6 \qquad \chi = 2.4495 $$

and the corresponding probabilities are .007 and .014. Thus the ordinary test would considerably overestimate the significance.

If $r = 3$ the corresponding probabilities are $\frac{27}{99}$ and $\frac{54}{99}$ or .273 and

.545 while the χ^2 test gives

$$\chi^2 = 1.5 \qquad \chi = 1.225$$

and the two probabilities are .110 and .220. Here both tests would agree in giving a non-significant result, and the difference between the two tests is relatively smaller than when $r = 2$. For $r = 2$, method (2) gives a probability of .009.

Example (II).

As another example we may take the table

TABLE IV.

	Marked	*Not Marked*	
	—	—	
Sample I . . .	10	2	12
Sample II . . .	4	8	12
	14	10	24

Here the expected numbers are 5 or greater and we might ordinarily use the χ^2 test. The possible tables are obtained from the following schema by giving r the values from 2 to 12.

Marked	*Not Marked*	
—	—	
r	$12 - r$	12
$14 - r$	$r - 2$	12
14	10	12

The probabilities are ${}^{12}C_r \times {}^{12}C_{14-r}\, p^{14} q^{10}$

$$r = 2, 3, \ldots\ldots 12.$$

and we obtain the following values.

TABLE IV-*a*.

r	$^{12}C_r \times {}^{12}C_{14-r}$
2	66
3	2.640
4	32.670
5	174.240
6	457.380
7	627.264
8	457.380
9	174.240
10	32.670
11	2.640
12	66
Total.	$1.961.256 = {}^{24}C_{10}$

The chance of a table arising with a number of marked individuals in sample I as improbable or less probable than 10 is therefore $\frac{35.376}{1.961.256} = .018$ and the chance of any table as improbable or less probable than the observed one is .036.

The χ^2 test gives $\chi^2 = 6.1714$
$\chi^2 = 2.484$

and the two probabilities are .007 and .014. Thus even here the χ^2 test exaggerates the significance but both tests would agree in giving a significant difference if we took the five per cent level of significance. Method (2) gives a probability of .005.

It is to be noted that in this process we have only used samples giving the same estimate of p as that observed ; on the other hand the only assumption made in using the χ^2 test, in addition to the normality of the distribution of cell frequencies is that in each sample the expected frequencies are estimated from that sample.

Not to have kept constant the marginal numbers of marked and unmarked individuals would approximately have squared the number of possible tables to be considered and rendered the process impracticable ; further this procedure would not be comparable with the χ^2 test either, since we have two degrees of freedom instead of one.

The process can at once be extended to the case when the two samples are not of the same size. In this case we have the following schema :

$a + b - s$	s	$a + b$
$c - b + s$	$b + d - s$	$c + d$
$a + c$	$b + d$	$a + b + c + d$

The probability of the above table arising will be

$$\frac{{}^{a+b}C_s \times {}^{c+d}C_{b+d-s}}{{}^{a+b+c+d}C_{a+c}}$$

if we suppose the marginal totals fixed.

We may then give s all values consistent with the frequencies being positive, and discuss any case we wish. s must clearly be smaller than or at most equal to the smaller of $a + b$ and $c + d$ and can be zero if $c - b$ is positive or zero, but cannot be less than $b - c$ if $b - c$ is positive.

As illustrations we may take the following three tables, giving the number of cases of measles prevented and not prevented by the use of convalescent serum in each of two different schools.

Example (III).

TABLE V.

	Prevented	*Not Prevented*	
School I . . .	26	2	28
School II . . .	61	2	63
	87	4	91

The possible tables may be enumerated by giving s the values 0, 1, 2, 3, 4 in the following schema :

$28 - s$	s	28
$59 + s$	$4 - s$	63
87	4	91

The chance of a table with any particular value of s will then be given by dividing the corresponding entry in column 2 of Table Va by the total of that column.

TABLE V *a*.

s	$^{28}C_s \times {}^{63}C_{4-s}$
0	595.665
1	1.111.908
2	738.234
3	206.388
4	20.475
	$2.672.670 = {}^{91}C_4$

The chance of a table as improbable or less probable than that observed is therefore :

$$\frac{1.560.762}{2.672.670} = .584$$

We have $\chi^2 = .7264 \qquad \chi = .8523$

and the corresponding probability is .394. As already noted above, method (2) would give us a probability of

$$(1 - 28\, p^{27} q - p^{28})\ (p^{63} + 63\, p^{62} q + 1953\, p^{61} q^2)$$

with $p = .9560$. The numerical result is .166.

Example (IV).

TABLE VI.

	Prevented	*Not Prevented*	
School I . . .	16	1	17
School II . . .	13	0	13
	29	1	30

Here we only have two alternatives ; our schema is

$17 - s$	s	17
$12 + s$	$1 - s$	13
29	1	30

and s can either be 0 or 1.

TABLE VI *a*.

s	$^{17}C_s \times {}^{13}C_{1-s}$
0	13
1	17
	30

and the observed table is the most likely that can occur ; the chance of a table as unlikely or less likely is therefore 1.00.

We have $\chi^2 = .7911$ $\chi = .8894$

and the corresponding probability is .374.

Method (2) would give $p^{13}(1 - p^{17}) = .282$.

Example (V).

TABLE VII.

	Prevented	*Not Prevented*	
School I . . .	79	3	82
School II . . .	56	7	63
	135	10	145

Here s can take the values 0, 1, 2, . . . 10 in the schema.

$82 - s$	s	82
$53 + s$	$10 - s$	63
135	10	145

and using Tables of Log. $\Gamma(x)$, we find the following values for $\frac{{}^{82}C_s \times {}^{63}C_{10-s}}{{}^{145}C_{10}}$.

TABLE VII *a*.

s	$({}^{82}C_s \times {}^{63}C_{10-s}) / {}^{145}C_{10}$
0	.0002
1	.0024
2	.0156
3	.0594
4	.1442
5	.2327
6	.2530
7	.1831
8	0844
9	.0224
10	.0026
Total . . .	1.0000

and the chance of a table as improbable or less probable than that osebrved is .103.

We have $\chi^2 = 3.0818$ $\chi = 1.756$

and the corresponding probability is .080. Method (2) would give us

$$\left(\sum_{r=0}^{r=3} p^{82-r} q^r\right)\left(1 - \sum_{r=0}^{r=6} p^{63-r} q^r\right)$$

with $p = .9310$; the numerical result is .025.

Example (VI).

TABLE VIII.

	Prevented	*Not Prevented*	
I . . .	67	2	69
II . . .	76	12	88
	143	14	157

Here all the expected values are over 5 and the χ^2 test could be used ; the example is, however, included because the χ^2 test gives a siginificant difference and the comparison between the three methods is therefore of some interest.

Using method (3) s can take value from 0 to 14 in the schema

$69 - s$	s	69
$74 + s$	$14 - s$	88
143	14	157

and we find the following values for $\dfrac{{}^{69}C_s \times {}^{88}C_{14-s}}{{}^{157}C_{14}}$.

TABLE VIII *a*.

s	$({}^{69}C_s \times {}^{88}C_{14-s})/{}^{157}C_{14}$	s	$({}^{69}C_s \times {}^{88}C_{14-s})/{}^{157}C_{14}$
0 . . .	.0002	8 . . .	.1298
1 . . .	.0024	9 . . .	.0636
2 . . .	.0138	10 . . .	.0227
3 . . .	.0480	11 . . .	.0067
4 . . .	.1118	12 . . .	.0010
5 . . .	.1839	13 . . .	.0001
6 . . .	.2207	14 . . .	.0004
7 . . .	.1962	—	
		Total . . .	1.0000

The chance of a table as improbable as or less probable than that observed is .031.

$$\chi^2 = 5.490 \qquad \chi = 2.343$$

and the corresponding probability .019. Method (2) gives

$$\left(\sum_{r=0}^{2} {}^{69}C_r\, p^{69-r}\, q^r\right)\left(1 - \left\{\sum_{r=0}^{11} {}^{88}C_r\, p^{88-r}\, q^r\right\}\right)$$

with $p = .9108$, the numerical value of this expression is .004.

Example (VII).

Finally we may consider the following example in which not all the expected frequencies are greater than 5 but in which the χ^2 test gives a significant result.

TABLE IX.

	Marked	*Not Marked*	
Sample I . . .	18	3	21
Sample II . . .	129	4	133
	147	7	154

Using method (3) s can take the values from 0 to 7 in the schema

$21 - s$	s	21
$126 + s$	$7 - s$	133
147	7	154

and we find the following values of

$$({}^{21}C_s \times {}^{133}C_{7-s}) \,/\, {}^{154}C_7 .$$

s	$({}^{21}C_s \times {}^{133}C_{7-s})/{}^{154}C_7$
0 . . .	.3505
1 . . .	.4057
2 . . .	.1902
3 . . .	.0467
4 . . .	.0065
5 . . .	.0005
6 . . .	.00002
7 . . .	.0000003
	1.0000

The chance of a table as improbable as or less probable than that observed is therefore .053.

We find $\chi^2 = 5.3169$ $\chi = 2.3058$

and the corresponding probability is .021. Method (2) gives:

$$\left(1 - \sum_{r=0}^{2} p^{21-r} q^r\right)\left(\sum_{r=0}^{4} p^{133-r} q^r\right)$$

with $p = .9545$, the numerical value is 018.

We may summarise the results as follows:

TABLE X.

Probability of as improbable or a less probable result arising by chance.

	Usual Method *	*Method* (2)	*Method* (3)
Table III . . .	.014 (.066)	.009	.061
Table IV . . .	.014 (.038)	.005	.036
Table V . . .	.394 (.766)	.166	.584
Table VI . . .	.374 (—)	.282	1.000
Table VII . . .	.080 (.154)	.025	.103
Table VIII . . .	.019 (.039)	.004	.031
Table IX . . .	.021 (.082)	.018	.054

Method (2) gives results which are undoubtedly too low; in all these examples, except the first and last, method (3) would give the same result as the usual method, if we used the 5 per cent level of significance.

When, as in Table III the numbers in all cells are small the non-approximative method should be used, but if samples are of reasonable size (say giving 30 individuals or more when clubbed together) and there are small frequencies (i. e. yielding expected frequencies < 5) in one or two cells owing to the percentage of marked individuals being 90 per cent or more, we shall seldom be misled by applying the usual test and taking the 1 per cent level of significance, that is to say taking a difference between the two percentages exceeding 2.6 times the standard error as significant.

* The figures in brackets show the results of applying of YATES' correction.

CARLOS E. DIEULEFAIT

Généralisation des courbes de K. Pearson

Dans la méthode de TSCHEBYCHEW, pour interpoler les données $(x_i\ ;\ y_i)$ $(i = 0, 1, 2, \ldots . n.)$ au lieu d'utiliser les développements courants du type :

$$f(x) = a_0 + a_1 x + a_2 x^2 + \ldots . + a_k x^k \tag{1}$$

on les substitue par une série de polynomes :

$$f(x) = \alpha_0 P_0(x) + \alpha_1 P_1(x) + \ldots . + \alpha_k P_k(x) \tag{2}$$

qui ont la propriété d'être orthogonaux dans le champ où x_i est considéré et que nous désignerons par ω.

L'orthogonalité des $P_0(x)$, $P_1(x)$, ... revient à imposer la condition :

$$\sum_{\omega} P_s(x)\, P_j(x) \begin{cases} = 0 \text{ pour } s \neq j \\ \neq 0 \quad \text{» } \quad s = j . \end{cases} \tag{3}$$

Il y a interpolation, c'est à dire on a : $f(x_i) = y_i$ si $k = n$ pour $i = 0 , 1 , 2 , \ldots . n$. Dans les cas courants d'ajustement l'on a : $k < n$ et la détermination des paramètres α_s de la (2) se fait par la méthode des moindres carrés, c'est à dire que :

$$S_n = \sum_{\omega} [y_i - f(x_i)]^2 \qquad \text{soit minimum.} \tag{4}$$

Pour cela l'on doit avoir :

$$\frac{\partial S_n}{\partial \alpha_s} = 0 \qquad\qquad s = 0, 1, 2, \ldots . k .$$

d'où :

$$(5) \qquad \sum_{\omega} [y_i - f(x_i)] P_s(x_i) = 0$$

mais à cause de la (3) on arrive facilement à la

$$\alpha_s = \frac{\sum_{\omega} y_i P_s(x_i)}{\sum_{\omega} P_s^2(x_i)} .$$

qui met en évidence l'indépendance de chaque coefficient par rapport aux autres, de façon de pouvoir augmenter le degré de l'approximation donnée par (2) en conservant tous les calculs faits auparavant, ce qui n'était pas possible avec le développement (1).

L'avantage de la méthode de TSCHEBYSCHEW est une question de simplification, mais au point de vue spéculatif la détermination d'un développement (2) reste toujours d'une valeur empirique.

Dans beaucoup de questions quand on dispose d'un ensemble de mesures $(x_i ; y_i)$ relatives à un phénomène, il est possible de connaître une fonction ou loi approchée $\varphi(x)$ telle que

$$y_i = \varphi(x_i) + \varepsilon_i \quad ; \quad \varepsilon_i \to 0$$

avec l'exactitude de la loi $\varphi(x)$ et avec amélioration des techniques d'observation. Tel est un des cas de l'ajustement de diverses distributions des fréquences dans la méthode du Prof. K. PEARSON.

Mais souvent, malgré la haute discrimination de cette méthode, les valeurs ε_i empèchent de se contenter de la fonction $\varphi(x_i)$. On a besoin de pousser l'approximation plus loin. La généralisation des courbes de PEARSON répond à ces faits et sa résolution devient une question facile quand on se place en face du problème théorique qui implique une généralisation de la méthode de TSCHEBYSCHEW.

Au lieu du développement (2) pour ajuster les données $(x_i ; y_i)$ posons :

$$(6) \qquad f(x) = \varphi(x) [1 + \beta_1 F_1(x) + \beta_2 F_2(x) + \ldots + \beta_k F_k(x)]$$

$\varphi(x)$ étant la loi du phénomène.

Si cette loi était suffisante pour interpoler les paires $(x_i ; y_i)$ alors

on aurait $\beta_s = 0$ pour $s = 1, 2, 3, \ldots$ ce qui laisse voir le rôle correcteur des coefficients.

On peut dans la (6) maintenir la même simplification apportée par TSCHEBYSCHEW dans la détermination des coefficients. Pour cela on imposera aux $F_s(x)$ la condition d'être orthogonales avec la fonction $\varphi(x)$ prise comme noyau, c'est à dire :

$$(7) \qquad \sum_{\omega} \varphi(x) F_s(x) F_j(x) \begin{cases} = 0 & \text{pour } s \neq j \\ \neq 0 & \text{» } s = j . \end{cases}$$

Il suit de la (7) que les calculs des β_s de la (6) qui assurent la condition d'un minimum pour

$$S_n = \sum_{\omega} \left[\frac{y_i}{\sqrt{(\varphi x_i)}} - \sqrt{\varphi(x_i)} \sum_s \beta_s F_s(x_i) \right]^2$$

s'obtiennent par la

$$(8) \qquad \beta_s = \frac{\sum_{\omega} y_i F_s(x_i)}{\sum_{\omega} \varphi(x_i) F_s^2(x_i)} .$$

Il reste à voir comment on peut construire effectivement ces fonctions $F_s(x)$.

On peut arriver à cette génération des $F_s(x)$ en généralisant le procédé trouvé par V. ROMANOVSKY et que nous allons reprendre en le développant par une autre voie (*).

Soient les polynomes $A_0(x)$, $A_1(x)$, $A_2(x)$, ...
étant

$$(9) \qquad A_s(x) = \delta_{s.0} + \delta_{s.1} x + \delta_{s.2} x^2 + \ldots + \delta_{s.s} x^s .$$

Nous posons :

$$(10) \qquad F_s(x) = A_s(x) - A_{s-1}(x) \quad \text{avec} \quad F_0(x) = A_0(x) .$$

(*) V. ROMANOVSKY. « Comptes Rendus. Académie des Sciences ». Tome 181, pag. 595. V. aussi « Biometrika ». Vol. XIX pag. 93.

V. aussi C. DIEULEFAIT. « Anales. Socied. Científica Argentina », Tome CXV, pag. 23.

La condition (7) appliquée de proche en proche conduit, pour déterminer $F_s(x)$ au système

$$(11) \qquad \sum_{\omega} \varphi(x) F_l(x) . A_s(x) = \sum_{\omega} \varphi(x) F_l(x) A_{s-1}(x)$$

pour $l = 0, 1, 2, \ldots . (s-1)$ et avec les $s+1$ inconnues : $\delta_{s.j}$ pour $j = 0, 1, 2, 3, \ldots . s$.

On peut fixer une de ces $\delta_{s.j}$ après quoi on calcule les autres. On a ainsi une des solutions pour le système (11) et alors $A_s(x)$ est déterminé et par suite d'accord à la (10) on a $F_s(x)$.

Parmi les valeurs que l'on peut donner à une des $(s+1)$ inconnues on peut prendre partout $\delta_{k.k} = 1$ pour $k = 0, 1, 2, 3, \ldots . s$. et alors on obtient la formule :

$$(12) \qquad F_s(x) = -\sum_{l=0}^{s-1} \frac{\sum_{\omega} \varphi(x) . x^s . F_l(x)}{\sum_{\omega} \varphi(x) F_l^2(x)} . F_l(x) + x^s$$

qui est précisément la généralisation de celle qui a été donnée par V. ROMANOVSKY.

Il est évident que l'on pourra substituer aux $\sum_{\omega}$ des $\int$ prises dans l'intervalle de variation des x ; étant aussi toujours possible de substituer les bases x^s par des fonctions convenables $\theta_s(x)$ ce qui arrivera si au lieu du polynome intermediaire (10) on part de :

$$A_s(x) = \delta_{s.0} \theta_0(x) + \delta_{s.1} \theta_1(x) + \delta_{s.2} \theta_2(x) + \ldots . + \delta_{s.s} \theta_s(x) .$$

Il suffit maintenant d'appliquer ces raisonnements aux courbes de PEARSON pour obtenir leur généralisation. Nous considérons les types V et VI que nous croyons n'ont pas été étudiés.

On sait que la courbe VII ou de LAPLACE a été développée par CHARLIER et que c'est M. ROMANOVSKY qui a résolu le problème pour les courbes I, II et III en employant des méthodes différentes à celles que nous exposons ici (*).

Soit $\varphi(x)$ la courve V de PEARSON (**).

(*) V. ROMANOVSKY. « Biometrika ». Vol. XVI, pag. 106.

(**) W. P. ELDERTON. *Frequency Curves*, etc. pág. 94.

$$(13) \qquad \varphi(x) = y_0 \,.\, x^{-p} \,.\, e^{-\frac{\gamma}{x}} \qquad \text{avec} \quad x \left\{ \begin{matrix} 0 \\ \infty \end{matrix} \right\}.$$

avec $y_0 = \dfrac{\gamma^{p-1}}{\Gamma(p-1)}$. En appliquant la (12) on a :

$$F_1(x) = -\frac{\gamma}{p-2} + x$$

$$F_2(x) = \frac{\gamma^2}{(p-3)\,p-4)} - \frac{2\gamma}{(p-4)}\,x + x^2$$

$$F_3(x) = -\frac{\gamma^3}{(p-4)\,(p-5)\,(p-6)} + \frac{3\gamma^2}{(p-5)\,(p-6)}\,x - \frac{3\gamma}{(p-6)}\,x^2 + x^3$$

et en général :

$$(14) \qquad F_n(x) = \sum_{s=0}^{n} (-1)^{n+s} \,.\, \frac{\binom{n}{s}\,\gamma^{n-s}}{\prod_{j=n+s+1}^{2n} (p-j)}\,x^s \,.$$

Pour déterminer les α_s du développement

$$(15) \qquad f(x = \varphi(x)\,[1 + \alpha_1 F_1(x) + \ldots + \alpha_s F_s(x) + \ldots]$$

on tient :

$$(16) \qquad \alpha_n = \frac{\int_0^{\infty} f(x) \,.\, F_n(x)\,dx}{\int_0^{\infty} \varphi(x) \,.\, F_n^2(x)\,dx}$$

et si nous indiquons les moments expérimentaux par :

$$\mu_j \,.\, = \int_0^{\infty} x^j \,.\, f(x)\,dx$$

en tenant compte de la (14), le numérateur de la (16) s'écrit :

$$\sum_{s=0}^{n} (-1)^{n+s} \frac{\binom{n}{s}\,\gamma^{n-s}}{\prod_{j=n+s+1}^{2n} (p-j)} \cdot \mu_s$$

et pour le dénominateur de la (16) l'on a :

$$\int_0^\infty \varphi(x)\, F_n^2(x)\, dx = H_n \qquad \text{étant}$$

$$H_1 = \frac{\gamma^2}{(p-2)^2\,(p-3)}$$

$$H_2 = \frac{2\,\gamma^4}{(p-2)\,(p-3)^2\,(p-4)^2\,(p-s)}$$

$$H_3 = \frac{6\,\gamma^6}{(p-2)\,(p-3)\,(p-4)^2\,(p-5)^2\,(p-6)^2\,(p-7)}$$

et en général :

$$H_n = \frac{n!\,\gamma^{2n}}{\prod_{s=2}^{2n+1}(p-s)\ \prod_{l=n+1}^{2n}(p-l)} \qquad \text{d'où l'on a :}$$

$$(17) \qquad \alpha_n = \frac{1}{H_n} \sum_{s=0}^{n} (-1)^{n+s} \cdot \frac{\binom{n}{s}\gamma^{n-s}}{\prod_{j=n+s+1}^{2n}(p-j)} \cdot \mu_s$$

indiquant maintenant par (λ_s) les moments d'ordre s du noyau, c'est à dire :

$$(18) \qquad \lambda_s = \int_0^\infty x^s \cdot \varphi(x) \cdot dx .$$

Dans la méthode de PEARSON les paramètres de (13) ont été déterminés de façon à avoir :

$$(19) \qquad \lambda_s = \mu_s \qquad \text{pour} \qquad s = 1 \,.\, 2 \,.\, 3 \text{ et } 4 .$$

Mais en calculant sur la (18) on trouve :

$$\lambda_1 = \frac{\gamma}{(p-2)}$$

$$\lambda_2 = \frac{\gamma^2}{(p-2)\,(p-3)}$$

$$\lambda_3 = \frac{\gamma^3}{(p-2)\,p-3)\,(p-4)}$$

et en général :

$$\lambda_s = \frac{\gamma^s}{\prod_{l=2}^{s+1} (p - l)}$$

Mais en vertu de la (19) la (17) pourra s'écrire :

$$\alpha_n = \frac{\gamma^n}{H_n} \sum_{s=0}^{n} (-1)^{n+s} \cdot \frac{\binom{n}{s}}{\prod_{j=n+s+1}^{2n} (p - j) \prod_{r=2}^{s+1} (p - r)}$$

pour : $n = 1 \,.\, 2 \,.\, 3$. et 4 .

Mais comme :

$$\sum_{s=0}^{n} (-1)^{n+s} \cdot \frac{\binom{n}{s}}{\prod_{j=n+s+1}^{2n} (p - j) \sum_{r=2}^{s+1} (p - r)} = 0$$

ce que l'on voit immédiatement par induction, il résulte que :

(20) $$\alpha_n = 0 \quad \text{pour} \quad n = 1 \,.\, 2 \,.\, 3 \quad \text{et} \quad 4$$

et en particulier pour toute valeur de (t) par laquelle on aura :

$$\mu_t = \lambda_t$$

A cause du résultat (20) le développement (15) se réduit

$$f(x) = \varphi(x) [1 + \alpha_5 F_5(x) + \alpha_6 F_6(x) + \ldots] .$$

Nous allons généraliser maintenant avec le même procédé la courbe (VI) de PEARSON qui est :

$$\psi(x) = J_0 \,.\, (x - a)^{q_2} \,.\, x^{-q_1} \qquad \text{avec } x \text{ variable} \begin{cases} a \\ \infty \end{cases} \text{ (*)}$$

et :

$$J_0 = \frac{1}{a^{q_2 - q_1 + 1} B(q_2 + 1 \;;\; q_1 - q_2 - 1)} \qquad \text{d'où :}$$

(*) W. P. ELDERTON. Loc. cit. pag. 74. Je dois le développement pour cet type VI à mon ancienne élève Mlle. Clotilde Bula.

$$\psi(x) = \frac{\Gamma(q_1)}{a^{q_2 - q_1 + 1}\,\Gamma(q_2 + 1)\,\Gamma(q_1 - q_2 - 1)}\,(x - a)^{q_2}\,.\,x^{-q_1}$$

les fonctions orthogonales avec ce noyau calculé selon la (12) sont :

$$P_0(x) = 1$$

$$P_1(x) = -\frac{a(q_1 - 1)}{(q_1 - q_2 - 2)} + x$$

$$P_2(x) = \frac{a^2(q_1 - 1)(q_1 - 2)}{(q_1 - q_2 - 3)(q_1 - q_2 - 4)} - \frac{2a(q_1 - 2)}{(q_1 - q_2 - 4)}\,x + x_2$$

et en général :

$$P_n(x) = \sum_{k=0}^{n} (-1)^{n+k}\,.\,\frac{\binom{n}{k}\prod_{j=k+1}^{n}(q_1 - j)\,a^{n-k}}{\prod_{l=n+k+1}^{2n}(q_1 - q_2 - l)}\,.\,x^k$$

et pour les coefficients du développement on aura :

$$\alpha_n = \frac{1}{H_n}\sum_{k=0}^{n} (-1)^{n+k}\,.\,\frac{\binom{n}{k}\prod_{j=k+1}^{n}(q_1 - j)\,a^{n-k}\,\mu_k}{\prod_{l=n+k+1}^{2n}(q_1 - q_2 - l)}$$

étant :

$$\mu_k = \int_a^{\infty} f(x)\,x^k\,dx$$

et $$H_n = \int_a^{\infty} \Psi(x)\,P_n^2(x)\,dx$$ avec

$$H_1 = \frac{a^2(q_1 - 1)}{(q_1 - q_2 - 2)^2(q_1 - q_2 - 3)}\,.\,(q_2 + 1)$$

$$H_2 = \frac{2a^4(q_1 - 1)(q_1 - 2)}{(q_1 - q_2 - 2)(q_1 - q_2 - 3)^2(q_1 - q_2 - 4)^2(q_1 - q_2 - s)}\,.\,(q_2 + 1)(q_2 + 2)$$

et en général :

$$H_n = \frac{n\,!\;a^{2n}\prod_{l=1}^{n}(q_1 - l)}{\prod_{s=2}^{2n+1}(q_1 - q_2 - s)\prod_{l=n+1}^{2n}(q_1 - q_2 - l)}\prod_{j=1}^{n}(q_2 + j)\,.$$

Mais pour la détermination des paramètres de la $\psi(x)$ on a dû faire :

$$\mu_h = \lambda_h \qquad \text{pour} \qquad h = 1\,.\,2\,.\,3 \text{ et } 4$$

avec :

$$\lambda_h = \int_a^{\infty} \Psi(x)\, x^h\, d\,x$$

qui donne :

$$\lambda_1 = \frac{a\,(q_1 - 1)}{(q_1 - q_2 - 2)}$$

$$\lambda_2 = \frac{a^2\,(q_1 - 1)\,(q_1 - 2)}{(q_1 - q_2 - 2)\,(q_1 - q_2 - 3)}$$

et en général :

$$\lambda_n = \frac{a^n\prod_{j=1}^{n}(q_1 - j)}{\prod_{k=2}^{n+1}(q_1 - q_2 - k)}$$

et alors :

$$\text{pour } n = 1\,.\,2\,.\,3\,.\text{ et } 4$$

on aura :

$$\alpha_n = \frac{a^n}{H_n}\sum_{k=0}^{n}(-1)^{n+k}\,.\,\frac{\binom{n}{k}\prod_{j=k+1}^{n}(q_1 - j)\prod_{t=1}^{k}(q_1 - t)}{\prod_{l=n+k+1}^{2n}(q_1 - q_2 - l)\prod_{r=2}^{k+1}(q_1 - q_2 - r)}$$

mais étant nulle la summatoire il se suit que :

$$\alpha_n = 0 \quad \text{pour} \quad n = 1 \,.\, 2 \,.\, 3 \text{ et } 4$$

d'où :

$$f(x) = \Psi(x) \left[1 + \alpha_5 P_5(x) + \alpha_6 P_6(x) + \ldots\right] . \quad (*)$$

(*) Après avoir présenté cet article à la Rédaction de « Metron » j'ai pris connaissance que M. ROMANOVSKY avait donné aussi entre autres les généralisations ici traitées. Sa méthode différente à celle que nous exposons, fait partie d'un procédé que nous avons appelé méthode dérivative et qui est moins générale que le procédé direct que nous avons suivi. Le lecteur pourra consulter à cet égard notre memoire *Contribution à l'étude de la théorie de la corrélation* « Biometrika ». Vol. XXVI, pag. 379, et pour le mentionné article de M. ROMANOVSKY, voir « Atti del Congresso dei Matematici di Bologna », 1928.

HANS KOEPPLER

Das Fehlergesetz des Korrelationskoeffizienten und andere Wahrscheinlichkeitsgesetze der Korrelationstheorie

Zu den seltenen Darstellungen der Korrelationstheorie gehören zweifellos jene des wahrscheinlichen und des mittleren Fehlers, mit dem der Korrelationskoeffizient behaftet ist. Die meisten Lehrbücher begnügen sich mit der Angabe, dass der wahrscheinliche Fehler im Korrelationskoeffizienten

$$0.6745\, M = 0.6745\, \frac{1 - r^2}{\sqrt{n}} \quad \text{(I) *)}$$

betrage. Dabei bedeutet M den mittleren Fehler, der nach der Formel

$$M^2 = \frac{H}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-H^2 x^2} x^2 \, dx = \frac{1}{2 H^2}$$

*) Neben diesem Wert für den mittleren Fehler des Korrelationskoeffizienten, welcher die weiteste Verbreitung gefunden hat, wurde von Pearson noch der Wert

$$M = \frac{1 - r^2}{\sqrt{n - 1}}$$

aufgestellt. Wie sich dieser ergibt, kann z. B. aus dem Werk von COOLIDGE (s. a. a. O., pag. 134 u. f., § 1 Das Fehlergesetz) ersehen werden. BERNSTEIN (s. a. a. O.) bemerkt, dass FISHER den Wert

$$M = \frac{1 - r^2}{\sqrt{n - 2}}$$

aufgestellt hat.

berechnet wird,

$$r = \frac{S(x \,.\, y)}{\sqrt{S(x^2)\; S(y^2)}}$$

den Korrelationskoeffizienten und n die Zahl der Beobachtungspaare. Der rechts vom Gleichheitszeichen stehende Ausdruck der Gleichung (I) wird z. B. in einer Fussnote auf pag. 182 des sehr unterrichtenden Referats von Dr. KARL E. RANKE 1) gegeben, aus dem auch schon hervorgeht, dass PEARSON 2) diesen Ausdruck aufgefunden hat. Auch in der kleinen Einführung, welche W. BETZ 3) geschrieben hat, findet man den obigen Ausdruck mitgeteilt und kann aus dem der Schrift beigegebenen, sehr reichhaltigen Litteraturverzeichnis wiederum ersehen, dass der angegebene Ausdruck auf PEARSON zurückzuführen ist. Ferner findet man den mittleren, bezw. den wahrscheinlichen Fehler des Korrelationskoeffizienten wohl auch mit Rücksicht auf den beschränkten Raum ohne Herleitung angegeben bei W. JOHANNSEN 4), FELIX M. EXNER 5), P. RIEBESELL 6), C. H. FORSYTH 7), G. UDNY YULE 8), W. PALIN ELDERTON 9), ROBERT WILBUR BURGESS 10), FELIX BERNSTEIN 11), R. A. FISHER 12), C. V. L. CHARLIER 13), RIETZ-BAUR 14), der in einer Fussnote auch auf TSCHUPROW 15) hinweist. Letzterer leitet aber den mittleren Fehler für normale Korrelation auf andere Weise her, als es hier geplant ist. Ueber das Verfahren von TSCHUPROW geben auch die Bücher von BAUR 16), G. DARMOIS 17), R. RISSER et C.-E. TRAYNARD 18) Aufschluss, die diese Berechnungsweise teils kurz, teils ausführlich behandeln. Es gibt noch eine stattliche Zahl neuerer in Deutschland, Frankreich und Italien erschienen Lehrbücher, die wohl im Hinblick auf den theoretischen Charakter die Fehlerberechnung für den Korrelationskoeffizienten unberücksichtigt lassen. Dagegen fällt in dem beachtenswerten Werk von WILHELM WIRTH 19) die ausführliche Berechnung der verschiedenen Fehlerarten der Korrelationstheorie auf, die durch Bearbeitung der Darstellungen PEARSONS entstanden ist. Ohne die Vorbildlichkeit der PEARSON'schen Berechnungen oder ihrer WIRTH' schen Bearbeitung irgendwie kritisieren zu wollen, glaubte der Verfasser dieses Aufsatzes keine müssige Arbeit zu vollrichten, wenn er sich die Aufgabe stellte, das Wahrscheinlichkeitsgesetz des Fehlers im Korrelationskoeffizienten und einige verwandte Fehlergesetze nach den sonst in der Wahrscheinlichkeitsrechnung und in der Fehlertheorie üblichen Methoden

zu berechnen. Dabei ergaben sich aber, wie die nachstehenden Ausführungen zeigen, verschiedene Abweichungen von den gebräuchlichen Darstellungen.

Unseren Betrachtungen schicken wir den folgenden Satz voraus, der eine Verallgemeinerung einer in der Wahrscheinlichkeitsrechnung häufig gezeigten Aufgabe 20) ist.

Besteht die Fehlerfunktion

$$u = c_1 x_1 + c_2 x_2 + c_3 x_3 ,$$

in welcher c_1, c_2, c_3 drei bekannte Koeffizienten und x_1, x_2, x_3 drei von einander unabhängige Fehler sind, und unterliegen diese drei Fehler dem gemeinsamen Gesetz

$$P(x_1 , x_2 , x_3)\, d x_1\, dx_2\, dx_3 = \frac{\sqrt{D}}{\pi^{3/2}}\, e^{-f(x_1 , x_2 , x_3)}\, d x_1\, d x_2\, d x_3 ,$$

in welchem

$$f(x_1, x_2, x_3) = a_{11} x_1^2 + a_{22} x_2^2 + a_{33} x_3^2 + 2 a_{12} x_1 x_2 + 2 a_{13} x_1 x_3 + 2 a_{23} x_2 x_3$$

und

$$D = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{vmatrix} = a_{11} a_{22} a_{33} - a_{11} a_{23}^2 - a_{22} a_{13}^2 - a_{33} a_{12}^2 + 2 a_{12} a_{13} a_{23}$$

ist, so hat das Gesetz der Fehlerfunktion u die bekannte Exponentialform

$$P(u)\, d u = \frac{H}{\sqrt{\pi}}\, e^{-H^2 u^2}\, d u .$$

Die Präzision H hat dabei den Wert

$$H = \sqrt{\frac{D}{E}} ,$$

in welchem E den Ausdruck

$$E = (a_{22} a_{33} - a_{23}^2)\, c_1^2 + (a_{11} a_{33} - a_{13}^2)\, c_2^2 + (a_{11} a_{22} - a_{12}^2)\, c_3^2 +$$
$$+ 2 (a_{13} a_{23} - a_{12} a_{33})\, c_1 c_2 + 2 (a_{12} a_{23} - a_{13} a_{22})\, c_1 c_3 +$$
$$+ 2 (a_{12} a_{13} - a_{23} a_{11})\, c_2 c_3$$

bezeichnet. Mit der Darstellung eines einfacheren Falles dieser Fehlerfunktion hat sich der Verfasser 21) schon früher einmal beschäftigt. Man gelangt zu diesem Fehlergesetz, indem man z. B. die Variable x_1 mittels der aus der linearen Fehlerfunktion u abgeleiteten Gleichung

$$x_1 = \frac{1}{c_1}(u - c_2 x_2 - c_3 x_3)$$

aus der Exponentfunktion $f(x_1, x_2, x_3)$ eliminiert und so die neue Exponentfunktion

$$f(u, x_2, x_3) = \frac{a_{11}}{c_1^2} u^2 + \frac{1}{c_1^2}(a_{11} c_2^2 + a_{22} c_1^2 - 2 a_{12} c_1 c_2) x_2^2 +$$

$$+ \frac{1}{c_1^2}(a_{11} c_3^2 + a_{33} c_1^2 - 2 a_{13} c_1 c_3) x_3^2 + 2 \frac{1}{c_1^2}(a_{12} c_1 - a_{12} c_2) u x_2 +$$

$$+ 2 \frac{1}{c_1^2}(a_{13} c_1 - a_{11} c_3) u x_3 + 2 \frac{1}{c_1^2}(a_{11} c_2 c_3 - a_{12} c_1 c_3 - a_{13} c_1 c_2 + a_{23} c_1^2) x_2 x_3 =$$

$$= A_{11} u^2 + A_{22} x_2^2 + A_{33} x_3^2 + 2 A_{12} u x_2 + 2 A_{13} u x_3 + 2 A_{23} x_2 x_3$$

herleitet. Die Exponentialfunktion selbst hat man mit $\frac{1}{c_1}$ zu multiplizieren, weil man nunmehr nicht nach x_1, sondern nach u zu integrieren hat, und $d x_1 = \frac{1}{c_1} du$ zu setzen ist. Die Integration einer Exponentialfunktion mit dem vorstehenden Exponenten erfolgt in der Weise, dass man die Funktion $f(u, x_2, x_3)$ zunächst auf die Form

$$B_0 u^2 + (B_1 u + B_2 x_2)^2 + (B_3 u + B_4 x_2 + B_5 x_3)^2$$

bringt und die Koeffizienten B_0, B_1, B_2, B_3, B_4, B_5 nach der Methode der unbestimmten Koeffizienten ermittelt. Nach einigen Berechnungen findet man sodann, dass B_0 den Wert annimmt:

$$B_0 = \frac{1}{A_{22} A_{33} - A_{23}^2}(A_{11} A_{22} A_{33} - A_{12}^2 A_{33} - A_{13}^2 A_{22} - A_{23}^2 A_{11} +$$

$$+ 2 A_{12} A_{13} A_{23}) = \frac{D'}{A_{22} A_{33} - A_{23}^2},$$

in welchem D' die Determinante der Funktion $f(u, x_2, x_3)$ bedeutet. Setzt man die Werte der Koeffizienten A_{ik} ein, so findet man nach einigen Umformungen

$$D' = \frac{D}{c_1^2} \text{ und folglich } B_0 = \frac{D}{c_1 (A_{22} A_{33} - A_{23}^2)} .$$

Nach x_2 und x_3 integriert man, indem man

$$B_1 u + B_2 x_2 = y \quad , \text{ mithin } d x_2 = \frac{1}{B_2} d y$$

$$B_3 u + B_4 x_2 + B_5 x_3 = z \, , \text{ mithin } d x_3 = \frac{1}{B_5} d z$$

setzt. Mittels der hier übergegangenen Werte für B_2 und B_5 findet man

$$B_2 B_5 = \sqrt{A_{22} A_{33} - A_{23}^2} .$$

Der Koeffizient C der Exponentialfunktion

$$C e^{-B_0 u^2}$$

nimmt nach der beschriebenen Berechnung den Wert an:

$$C = \frac{\sqrt{D}}{\pi^{3/2}} \frac{1}{c_1} \frac{\pi}{B_2 B_5} = \frac{1}{c_1 \sqrt{\pi}} \sqrt{\frac{D}{A_{22} A_{33} - A_{23}^2}} .$$

Durch eine einfache Berechnung findet man, dass

$$A_{22} A_{33} - A_{23}^2 = \frac{1}{c_1^2} E ,$$

sodass sich, wie vordem angegeben,

$$C = \frac{1}{\sqrt{\pi}} \sqrt{\frac{D}{E}} \quad \text{und} \quad B_0 = \frac{D}{E}$$

ergibt. Man kann auch die Methode des Diskontinuitätsfaktors zur Darstellung dieser Fehlerfunktion verwenden, wenn man das dreifache Integral

$$\frac{\sqrt{D}}{\pi^{3/2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-f(x_1, x_2, x_3)} d x_1 d x_2 d x_3 = 1 ,$$

welches die Summe aller Wahrscheinlichkeiten darstellt, mit dem bekannten Besselschen Diskontinuitätsfaktor

$$\frac{d\,u}{2\,\pi}\int_{-\infty}^{\infty} e^{-(c_1 x_1 + c_2 x_2 + c_3 x_3 - u)\,z\,i}\,d\,z\,,$$

dessen Eigenschaften hier als bekannt vorausgesetzt werden sollen, multipliziert (*). Darauf forme man zunächst den Exponenten in der Weise um, dass man

$$P(u)\,d\,u = \frac{\sqrt{D}\,d\,u}{\pi^{3/2} A_1 A_4 A_6}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\varphi_1^2 - \varphi_2^2 - \varphi_3^2}\,d\,\varphi_1\,d\,\varphi_2\,d\,\varphi_3\,.$$

$$\frac{2\,\pi}{1}\int_{-\infty}^{\infty} e^{-\Omega + u\,z\,i}\,d\,z$$

erhält. Dabei ist

$$\varphi_1^2 = (A_1 x_1 + A_2 x_2 + A_3 x_3 + B_1 i)^2$$
$$\varphi_2^2 = (A_4 x_2 + A_5 x_3 + B_2 i)^2$$
$$\varphi_3^2 = (A_6 x_3 + B_3 i)^2$$
$$\Omega = B_1^2 + B_2^2 + B_3^2\,.$$

Die Koeffizienten A_1, A_2, A_3, A_4, A_5, A_6 und B_1, B_2, B_3 werden nach Auflösung der Quadrate φ_1^2, φ_2^2, φ_3^2 und Zusammenfassung der Glieder mit denselben Unbekannten durch Gegenüberstellung mit der ursprünglichen Funktion

$$f(x_1, x_2, x_3) + c_1 z x_1 i + c_2 z x_2 i + c_3 z x_3 i$$

ermittelt. Diese Transformation wird übrigens schon von BIENAYMÉ 22) in allgemeinerer Weise gezeigt. Im Verlaufe der weiteren Berechnungen erhält man den bekannten Laplaceschen Integral-Ausdruck

$$P(u)\,d\,u = \frac{d\,u}{2\,\pi}\int_{-\infty}^{\infty} e^{-\frac{E}{4D}z^2 + u\,z\,i}\,d\,z\,,$$

(*) Einen analogen Fall für zwei Variable behandelte der Verfasser in seinem Aufsatz *Die Berechnung des jährlichen Risikos schwierigerer Versicherungsarten*, welcher in den «Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker», 11. Heft, Bern 1916, erschienen ist.

welcher auf den eingangs angegebenen Ausdruck für das Differential der Fehlerfunktion $P(u)\,du$ führt. Unter der Voraussetzung, dass das Fehlergesetz $P(u)$ bekannt ist, gestalten sich die folgenden Berechnungen recht einfach. Seine Darstellung hier vorauszuschicken, dürfte im Hinblick auf den Umstand, dass in der bekannteren Litteratur, wie schon angedeutet wurde, nur das Gesetz der Fehlerfunktion

$$u = c_1 z_1 + c_2 z_2 + c_3 z_3 +$$

aus dem Produkt der Integrale

$$\frac{h_1 h_2 h_3 \ldots}{\sqrt{\pi \cdot \pi \cdot \pi \ldots}} \int_{-\infty}^{\infty} e^{-h_1^2 z_1^2}\, dz_1 \int_{-\infty}^{\infty} e^{-h_2^2 z_2^2}\, dz_2 \int_{-\infty}^{\infty} e^{-h_3^2 z_3^2}\, dz_3 \ldots = 1$$

abgeleitet wird, wohl berechtigt erscheinen. In diesem Fall wird die Präzision der Fehlerfunktion $P(u)$ bekanntlich nach der einfachen Formel

$$H^2 = \frac{1}{\Sigma \frac{c_i^2}{h_i^2}}$$

berechnet. Eine erweiterte Anwendung dieses Satzes hat der Verfasser in dem Aufsatz "Zur begründenden Darstellung des ferneren Risikos verwickelterer Versicherungsformen" 23) gegeben. Ueber die zuerst angewendete Methode bei der Integration der Funktion $e^{-\Sigma a_{ik} x_i x_k}$ geben z. B. Prof. K. HATTENDORFF 24), Prof. Dr. DIENGEN (wenn man den Kosinus = 1 setzt) 25), Prof. CZUBER 26) und Prof. LINDELÖF 27) Aufschluss. Eine elegante Transformation der Funktion $f(x_1, x_2, x_3)$ zwecks Integration nach x_2 und x_3 zeigt auch Prof. WITTSTEIN 28). Ein neueres Werk, das diese Integration bespricht, ist jenes von Prof. I. L. COOLIDGE 29). Ferner kann man sich über diese Integrationsmethode auch mit Hilfe des Aufsatzes von WILLERS, "Korrelation zwischen drei Veränderlichen" 30) unterrichten. Dieselbe Zerlegung der Funktion $f(x_1, x_2, x_3)$ findet man übrigens bei G. DARMOIS ((*s. a. a. O*) auf pag. 252 und bei R. RISSER et C.-E. TRAYNARD (*s. a. a. O.*) auf pag. 228; doch ist sie dort dem Zweck entsprechend nicht vollends durchgeführt.

Die Wahrscheinlichkeits- oder Frequenzfunktion für ein Paar zusammengehöriger Abweichungen x und y von ihren Mittelwerten, welche aus einer ausreichend grossen Anzahl n von Paaren zu einander

gehörender Abweichungen herausgegriffen ist, wird, wie aus den Lehrbüchern der Korrelationstheorie zu ersehen ist, bei Normalverteilung oder sogenannter normalen Korrelation in der Form

$$P(x, y)\, dx\, dy = \frac{1}{2\pi \mathfrak{G}_1 \mathfrak{G}_2 \sqrt{1-r^2}}\, e^{-\frac{1}{2(1-r^2)}\left\{\frac{x^2}{\mathfrak{G}_1^2} - 2r\frac{x}{\mathfrak{G}_1}\frac{y}{\mathfrak{G}_2} + \frac{y^2}{\mathfrak{G}_2^2}\right\}} dx\, dy$$

dargestellt, in welcher

$$\mathfrak{G}_1^2 = \frac{S(x^2)}{n} \quad \text{und} \quad \mathfrak{G}_2^2 = \frac{S(y^2)}{n}$$

die arithmetischen Mittel der Quadrate aller Abweichungen und

$$r = \frac{\mathfrak{G}_{1,2}}{\mathfrak{G}_1 \cdot \mathfrak{G}_2}$$

den schon in anderer Bezeichnung erwähnten Korrelationskoeffizienten darstellt. Abweichend von der englischen Bezeichnungsweise bedienen wir uns für den Mittelwert der Produkte x, y der sonst nicht üblichen Bezeichnung

$$\frac{S(x, y)}{n} = \mathfrak{G}_{1,2}.$$

Unsere nächste Aufgabe soll es sein, das Fehlergesetz der drei Fehler Δ_1, Δ_2, $\Delta_{1,2}$ darzustellen, mit denen die Mittelwerte $\mathfrak{G}_1$, $\mathfrak{G}_2$, $\mathfrak{G}_{1,2}$ behaftet sein können. Wir wollen uns dabei der bekannten Methode der wahrscheinlichsten Hypothese bedienen und das Produkt der Wahrscheinlichkeiten sämtlicher Beobachtungspaare

$$\prod_{i=1}^{i=n} P(x_i, y_i) = \frac{1}{(2\pi \mathfrak{G}_1 \mathfrak{G}_2)^n (1-r^2)^{\frac{n}{2}}}\, e^{-\frac{1}{2(1-r^2)}\left\{\frac{S(x^2)}{\mathfrak{G}_1^2} - 2r\frac{S(x.y)}{\mathfrak{G}_1 \cdot \mathfrak{G}_2} + \frac{S(y^2)}{\mathfrak{G}_2^2}\right\}}$$

unter Fortfassung der Differentiale $\prod_{i=1}^{i=n} dx_i\, dy_i$ bilden. Wir können dieses Produkt in der Form

$$\frac{1}{(2\pi)^n}\, e^{-\Omega(\mathfrak{G}_1, \mathfrak{G}_2, \mathfrak{G}_{1,2})}$$

schreiben, wenn wir setzen :

$$\Omega\,(6_1\,,\,6_2\,,\,6_{1,2}) = n\,l\,n\,(6_1\,6_2) + \frac{n}{2}\,l\,n\,(1-r^2) +$$

$$+\frac{1}{2\,(1-r^2)}\left\{\frac{S\,(x^2)}{6_1^2} + \frac{S\,(y^2)}{6_2^2} - \frac{2\,r}{6_1\,6_2}\,S\,(x\,.\,y)\right\}.$$

Bei Berücksichtigung der Werte von $6_1\,,\,6_2\,,\,6_{1,2}$ liesse sich der dritte Ausdruck der rechten Seite noch zu "n" vereinfachen. Warum dieses aber nicht geschehen darf, werden wir sogleich dartun. Sind nämlich $6_1' = 6_1 + \Delta_1\,,\; 6_2' = 6_2 + \Delta_2\,,\; 6_{1,2}' = 6_{1,2} + \Delta_{1,2}$ die wahren Werte, so wird der genaue Exponent von e lauten :

$$\Omega\,(6_1'\,,\,6_2'\,,\,6_{1,2}') = n\,l\,n\,6_1' + n\,l\,n\,6_2' + \frac{n}{2}\,l\,n\,(1-r'^2) +$$

$$+\frac{1}{2\,(1-r'^2)}\left\{\frac{S\,(x^2)}{6_1'^2} + \frac{S\,(y^2)}{6_2'^2} - \frac{2\,r'}{6_1'\,6_2'}\,S\,(x\,.\,y)\right\}.$$

Nach der Theorie der wahrscheinlichsten Hypothese ist aber die Wahrscheinlichkeit, dass unter allen möglichen Werten von Δ_1, Δ_2, $D_{1,2}$ grade diese Fehler auftreten, näherungsweise

$$P\,(\Delta_1\,,\,\Delta_2\,,\,\Delta_{1,2}) = \frac{e^{-\Psi\,(\Delta_1\,,\,\Delta_2\,,\,\Delta_{1,2})}}{\int_{-\infty}^{\infty}\!\!\int\!\!\int e^{-\Psi\,(\Delta_1\,,\,\Delta_2\,,\,\Delta_{1,2})}\,d\,\Delta_1\,,\,d\,\Delta_2\,,\,d\,\Delta_{1,2}}\,.$$

Hat man $\Psi\,(\Delta_1\,,\,\Delta_2\,,\,\Delta_{1,2})$ hinreichend genau bestimmt, so liefert das Integral des Nenners die Vorzahl der Exponentialfunktion der Fehler $\Delta_1\,,\,\Delta_2\,,\,\Delta_{1,2}$. Den Exponenten der Fehlerfunktion finden wir aber, wenn wir die Funktion $\Psi\,(6_1+\Delta_1\,,\,6_2+\Delta_2\,,\,6_{1,2}+\Delta_{1,2})$ nach dem Taylorschen Satz in die nach dem zweiten Grade abgebrochene Reihe

$$\Omega\,(6_1\,,\,6_2\,,\,6_{1,2}) + \Psi\,(\Delta_1\,,\,\Delta_2\,,\,\Delta_{1,2}) = \Omega\,(6_1\,,\,6_2\,,\,6_{1,2}) +$$

$$+\frac{\partial\,\Omega}{\partial\,6_1}\Delta_1 + \frac{\partial\,\Omega}{\partial\,6_2}\Delta_2 + \frac{\partial\,\Omega}{\partial\,6_{1,2}}\Delta_{1,2} +$$

$$+\frac{1}{2}\left[\frac{\partial^2\,\Omega}{\partial\,6_1^2}\,\Delta_1^2 + \frac{\partial^2\,\Omega}{\partial\,6_2^2}\,\Delta_2^2 + \frac{\partial^2\,\Omega}{\partial\,6_{1,2}^2}\,\Delta_{1,2}^2 + 2\,\frac{\partial^2\,\Omega}{\partial\,6_1\,\partial\,6_2}\,\Delta_1\,\Delta_2 +\right.$$

$$\left. + 2\,\frac{\partial^2\,\Omega}{\partial\,6_1\,\partial\,6_{1,2}}\,\Delta_1\,\Delta_{1,2} + 2\,\frac{\partial^2\,\Omega}{\partial\,6_2\,\partial\,6_{1,2}}\,\Delta_2\,\Delta_{1,2}\right]$$

entwickeln, so dass, wie die Berechnung zeigen wird, $\Psi(\Delta_1, \Delta_2, \Delta_{1,2})$ den zweiten Differentialquotienten der Reihenentwicklung darstellt. Die Berechnung der Differentialquotienten von Ω erfolgt mit Hilfe des an erster Stelle angegebenen Originalausdrucks $\Omega(6_1, 6_2, 6_{1,2})$. Zum Schluss möge dann $S(x^2) = n\,6_1^2$, $S(y^2) = n\,6_2^2$, $S(x.y) = n\,6_{1,2}$ gesetzt werden. Diese Berechnungsweise führt zu den gleichen Ergebnissen, wie wenn man die Ableitungen von $\Omega(6_1', 6_2', 6_{1,2}')$ bilden würde, sodann $6_1'$ in 6_1, $6_2'$ in 6_2, $6_{1,2}'$, in $6_{1,2}$ übergehen liesse und damit auch gleichzeitig den Grössen $6_1, 6_2, 6_{1,2}$ ihre bekannten Werte beilegen würde Man erhält auf diese Weise

$$\frac{\partial\,\Omega}{\partial\,6_1} = \frac{n}{6_1} + \frac{n\,r^2}{6_1(1-r^2)} - \frac{2\,n\,r^2}{6_1(1-r^2)} - \frac{n(1-2\,r^2)}{6_1(1-r^2)} = 0$$

$$\frac{\partial\,\Omega}{\partial\,6_2} = \frac{n}{6_2} + \frac{n\,r^2}{6_2(1-r^2)} - \frac{2\,n\,r^2}{6_2(1-r^2)} - \frac{n(1-2\,r^2)}{6_2(1-r^2)} = 0$$

$$\frac{\partial\,\Omega}{\partial\,6_{1,2}} = -\frac{n\,r^2}{6_{1,2}(1-r^2)} + \frac{2\,n\,r^2}{6_{1,2}(1-r^2)} - \frac{n\,r^2}{6_{1,2}(1-r^2)} = 0$$

$$\frac{\partial^2\,\Omega}{\partial\,6_1^2} = -\frac{n}{6_1^2} - \frac{n\,r^2(3-r^2)}{6_1^2(1-r^2)^2} + \frac{n(3+r^2)}{6_1^2(1-r^2)^2} = \frac{2\,n}{6_1^2(1-r^2)^2}$$

$$\frac{\partial^2\,\Omega}{\partial\,6_2^2} = -\frac{n}{6_2^2} - \frac{n\,r^2(3-r^2)}{6_2^2(1-r^2)^2} + \frac{n(3+r^2)}{6_2^2(1-r^2)^2} = \frac{2\,n}{6_2^2(1-r^2)^2}$$

$$\frac{\partial^2\,\Omega}{\partial\,6_{1,2}^2} = -\frac{n\,r^2(1+r^2)}{6_{1,2}^2(1-r^2)^2} + \frac{2\,n\,r^2(1+r^2)}{6_{1,2}^2(1-r^2)^2} = \frac{n\,r^2(1+r^2)}{6_{1,2}^2(1-r^2)^2}$$

$$\frac{\partial^2\,\Omega}{\partial\,6_1\,\partial\,6_2} = -\frac{2\,n\,r^2}{6_1\,6_2(1-r^2)^2} + \frac{4\,n\,r^2}{6_1\,6_2(1-r^2)^2} = \frac{2\,n\,r^2}{6_1\,6_2(1-r^2)^2}$$

$$\frac{\partial^2\,\Omega}{\partial\,6_1\,\partial\,6_{1,2}} = \frac{2\,n\,r^2}{6_1\,6_{1,2}(1-r^2)^2} - \frac{4\,n\,r^2}{6_1\,6_{1,2}(1-r^2)^2} = -\frac{2\,n\,r^2}{6_1\,6_{1,2}(1-r^2)^2}$$

$$\frac{\partial^2\,\Omega}{\partial\,6_2\,\partial\,6_{1,2}} = \frac{2\,n\,r^2}{6_2\,6_{1,2}(1-r^2)^2} - \frac{4\,n\,r^2}{6_2\,6_{1,2}(1-r^2)^2} = -\frac{2\,n\,r^2}{6_2\,6_{1,2}(1-r^2)^2}$$

und folglich

$$\Psi(\Delta_1, \Delta_2, \Delta_{1,2}) = \frac{n}{(1-r^2)^2}\left\{\frac{\Delta_1^2}{6_1^2} + \frac{\Delta_2^2}{6_2^2} + \frac{r^2(1+r^2)}{2}\,\frac{\Delta_{1,2}^2}{6_{1,2}^2} + \right.$$

$$\left. + 2\,r^2\,\frac{\Delta_1}{6_1}\,\frac{\Delta_2}{6_2} - 2\,r^2\,\frac{\Delta_1}{6_1}\,\frac{\Delta_{1,2}}{6_{1,2}} - 2\,r^2\,\frac{\Delta_2}{6_2}\,\frac{\Delta_{1,2}}{6_{1,2}}\right\}.$$

Die Determinate dieser Funktion liefert den Wert

$$D = \frac{n^3 r^2 (1 - r^2)^3}{(1 - r^2)^6 \cdot 2\, 6_1^2 \cdot 6_2^2\, 6_{1,2}^2} = \frac{n^3 r^2}{2 (1 - r^2)^3\, 6_1^2\, 6_2^2\, 6_{1,2}^2},$$

sodass das Nennerintegral den Wert

$$\sqrt{\frac{\pi^3}{D}} = \frac{\sqrt{2}\, \pi^{3/2} (1 - r^2)^{3/2} 6_1 \cdot 6_2 \cdot 6_{1,2}}{n^{3/2} r}$$

und mithin die Vorzahl der Exponentialfunktion den Wert

$$K = \frac{n^{3/2} \cdot r}{\sqrt{2}\, \pi^{3/2} 6_1 \cdot 6_2 \cdot 6_{1,2} (1 - r^2)^{3/2}}$$

annimmt. Wir kennen nun das bisher wohl nur selten dargestellte gemeinsame Gesetz der Fehler Δ_1, Δ_2, $\Delta_{1,2}$.

Wie beiläufig bemerkt werde, kann dieses Gesetz dazu dienen, die Gesetze der Fehler Δ_1, Δ_2, $\Delta_{1,2}$ unabhängig von einander darzustellen. Hierzu hat man, wie eingangs gezeigt wurde, nur das Quadrat der Präzision

$$H^2 = \frac{1}{D_k} \{ A_{11} A_{22} A_{33} - A_{12}^2 A_{33} - A_{13}^2 A_{22} - A_{23}^2 A_{11} + 2 A_{12} A_{13} A_{23} \}$$

zu berechnen, indem man für die Unterdeterminanten D_k die Werte setzt, wie sie für die Berechnung der Fehlerwahrscheinlichkeitsfunktionen von Δ_1, Δ_2, $\Delta_{1,2}$ zu wählen sind. Auf diesem Wege findet man für die Quadrate der Präzisionen der Fehlerwahrscheinlichkeitsfunktionen von

$$\begin{matrix} \Delta_1 & \Delta_2 & \Delta_{1,2} \\ D_1 = A_{12} A_{33} - A_{23}^2 & D_2 = A_{11} A_{33} - A_{13}^2 & D_3 = A_{11} A_{22} - A_{12}^2 \end{matrix} \quad (D_3 = D_{1,2})$$

und folglich

$$h_1^2 = \frac{n}{6_1^2} \qquad h_2^2 = \frac{n}{6_2^2} \qquad h_{1,2}^2 = \frac{n r^2}{2\, 6_{1,2}^2 (1 + r^2)}.$$

Die letztere ist bisher wohl nur ganz selten dargestellt worden.

Um die Fehlerfunktion des Korrelationskoeffizienten zu berechnen, bilden wir nach dem Taylorschen Satz die Näherungsgleichung

$$r' = r + \frac{\partial r}{\partial 6_1} \Delta_1 + \frac{\partial r}{\partial 6_2} \Delta_2 + \frac{\partial r}{\partial 6_{1,2}} \Delta_{1,2},$$

bei welcher alle den ersten Grad übersteigenden Fehlerpotenzen vernachlässigt worden sind. r' stellt den angenäherten wahren Wert des Korrelationskoeffizienten r dar, sofern die verbesserten, den wahren Werten näher kommenden Werte

$$6'_1 = 6_1 + \Delta_1 \qquad 6'_2 = 6_2 + \Delta_2 \qquad 6'_{1,2} = 6_{1,2} + \Delta_{1,2}$$

bestehen würden. Für den Fehler im Korrelationskoeffizienten ergibt sich aus der vorstehenden Näherungsgleichung die lineare Fehlerfunktion

$$\Delta_r = r' - r = -\frac{r}{6_1}\Delta_1 - \frac{r}{6_2}\Delta_2 + \frac{r}{6_{1,2}}\Delta_{1,2}.$$

Jetzt sind alle Grössen bekannt, um die Präzision des Gesetzes zu bestimmen, welchem der Fehler Δ_r des Korrelationskoeffizienten unterworfen ist. Für die eingangs näher beschriebene Determinantengrösse E finden wir

$$E = \frac{n^2}{(1-r^2)^4}\,\frac{r^2\,(1-r^2)^3}{6_2^2\,6_1^2\,6_{1,2}^2} = \frac{n^2\,r^2}{(1-r^2)\,6_1^2\,6_2^2\,6_{1,2}^2}$$

und erhalten nun für das Quadrat der Präzision des Fehlergesetzes von Δ_r:

$$H^2 = \frac{D}{E} = \frac{n}{2\,(1-r^2)^2}.$$

Anstelle der Funktion $\Omega\,(6_1, 6_2, 6_{1,2})$ kann man auch die mit ihr gleichbedeutende Funktion $\Omega\,(6_1, 6_2, r)$ betrachten und sogleich das Exponentialgesetz der Fehler $\Delta_1, \Delta_2, \Delta_r$ herleiten. Dabei darf man dann aber den Korrelationskoeffizienten r nur als Funktion der Grössen $6_1, 6_2$ ansehen. Aus den vorhergehenden Berechnungen erkennt man auch, dass die Fehler $\Delta_1, \Delta_2, \Delta_r$ voneinander unabhängig sind, weil der den funktionalen Zusammenhang bewirkende Fehler $\Delta_{1,2}$ bei dieser Entwicklung nicht in Erscheinung tritt, was zur Folge hat, dass man Δ_r zunächst als konstant ansehen kann. Die Entwicklung nach dem Taylorschen Satz gestaltet sich unter diesen Voraussetzungen wesentlich einfacher. Unter Berücksichtigung der Grenzübergänge erhält man leicht

$$\frac{\partial\,\Omega}{\partial\,r} = -\frac{n\,r}{1-r^2} + \frac{2\,n\,r}{1-r^2} - \frac{n\,r}{1-r^2} = 0$$

$$\frac{\partial \Omega}{\partial 6_1} = \frac{n}{6_1} - \frac{n}{6_1} = 0$$

$$\frac{\partial \Omega}{\partial 6_2} = \frac{n}{6_2} - \frac{n}{6_2} = 0$$

$$\frac{\partial^2 \Omega}{\partial r^2} = -\frac{n(1+r^2)}{(1-r^2)^2} + \frac{2n(1+3r^2)}{(1-r^2)^2} - \frac{4nr^2}{(1-r^2)^2} = \frac{n(1+r^2)}{(1-r^2)^2}$$

$$\frac{\partial^2 \Omega}{\partial 6_1^2} = -\frac{n}{6_1^2} + \frac{n(3-2r^2)}{6_1^2(1-r^2)} = \frac{n(2-r^2)}{6_2^2(1-r^2)}$$

$$\frac{\partial^2 \Omega}{\partial 6_2^2} = -\frac{n}{6_2^2} + \frac{n(3-2r^2)}{6_2^2(1-r^2)} = \frac{n(2-r^2)}{6_2^2(1-r^2)}$$

$$\frac{\partial^2 \Omega}{d\,r\,d\,6_1} = -\frac{2nr^2}{6_1(1-r^2)} + \frac{nr}{6_1(1-r^2)} = -\frac{nr}{6_1(1-r^2)}$$

$$\frac{\partial^2 \Omega}{\partial\, r\, d\, 6_2} = -\frac{2nr^2}{6_2(1-r^2)} + \frac{nr}{6_2(1-r^2)} = -\frac{nr}{6_2(1-r^2)}$$

$$\frac{\partial^2 \Omega}{\partial 6_1 \partial 6_2} = \frac{1}{2(1-r^2)}\left(-\frac{2nr^2}{6_1 6_2}\right) = -\frac{nr^2}{6_1 6_2 (1-r^2)}.$$

Für die der Funktion $\Psi(\Delta_1, \Delta_2, \Delta_{1,2})$ entsprechende Funktion $\Psi(\Delta_r, \Delta_1, \Delta_2)$ erhält man mithin den Näherungsausdruck

$$\Psi(\Delta_r, \Delta_1, \Delta_2) = \frac{n}{2(1-r^2)}\left[\frac{1+r^2}{1-r^2}\Delta_r^2 + \frac{2-r^2}{6_1^2}\Delta_1^2 + \right.$$

$$\left. + \frac{2-r^2}{6_2^2}\Delta_2^2 - \frac{2r}{6_1}\Delta_r\Delta_1 - \frac{2r}{6_2}\Delta_r\Delta_2 - \frac{2r^2}{6_1 6_2}\Delta_1\Delta_2\right],$$

der sich auch ergeben hätte, wenn $\Delta_{1,2}$ aus $\Psi(\Delta_1, \Delta_2, \Delta_{1,2})$ mittels der Fehlerfunktion Δ_r eliminiert worden wäre.

Diesen Ausdruck haben aber auch Pearson und Filon (*s. a. a. o.*) und Wirth (*s. a. a. o.*) auf andere Weise ermittelt. Setzt man in der Funktion $\Psi(\Delta_r, \Delta_1, \Delta_2)$ für die Koeffizienten der Reihe nach A_{11}, A_{22}, A_{33}, A_{12}, A_{13}, A_{23}, so weiss man, dass das Quadrat der Präzision des Fehlers Δ_r nach der Formel

$$H^2 = \frac{1}{A_{22}A_{33} - A_{23}^2}\left\{A_{11}A_{22}A_{33} - A_{12}^2 A_{33} - A_{13}^2 A_{22} - A_{23}^2 A_{11} + \right.$$
$$\left. + 2\,A_{12}A_{13}A_{23}\right\}$$

zu berechnen ist, und gelangt zu dem Wert

$$H^2 = \frac{n}{2\,(1 - r^2)^2}.$$

Nach der bekannten Formel

$$M^2 \,.\, H^2 = \frac{1}{2}$$

findet man für den mittleren Fehler im Korrelationskoeffizienten den schon eingangs angeführten Wert

$$M = \frac{1 - r^2}{\sqrt{n}}.$$

Wird das eingangs beschriebene Gesetz der linearen Fehlerfunktion als allgemein bekannt vorausgesetzt, was es allerdings nicht ist, und berücksichtigt man dass unsere Ausführungen sich mit zwei Darstellungen des Gesetzes des Fehlers im Korrelationskoeffizienten und einigen Abschweifungen beschäftigen, so dürften diesen Untersuchungen die Eigenschaften kurz und einfach zu sein, wohl zuerkannt werden. Zweifelsohne haben sie aber den Vorzug, einen guten und unterrichtenden Einblick in die besprochenen Fehlergesetze zu gewähren.

VERZEICHNIS DES ANGEFÜHRTEN SCHRIFTEN UND AUFSÄTZE

1) *Die Theorie der Korrelation.* «Archiv für Anthropologie». Neue Folge. Band IV, Heft 2-3, Braunschweig, 1906.

2) *Mathematical Contributions to the Theory of Evolution.* IV. *On the probable errors of frequency constants and on the influence of random selection on variation and correlation.* (By KARL PEARSON and L. N. G. FILON). «Philosophical transactions of the Royal Society of London». Series A, Vol. 191 (1898). London.

3) *Ueber Korrelation.* Leipzig, 1911, Verlag von JOHANN AMBROSIUS BARTH, (pag. 21).

4) *Elemente der exakten Erblichkeitslehre.* Jena, 1909, pag. 509.

5) *Ueber die Korrelationsmethode.* Jena, 1913, pag. 13.
6) *Variationsstatistik.* Berlin-Wien, 1925, pag. 807.
7) *An Introduction to the mathematical Analysis of Statistics.* New York, 1924, pag. 218.
8) *An Introduction to the Theory of Statistics.* London, 1927, pag. 352.
9) *Frequency Curves and Correlation.* London, 1927, pag. 183.
10) *Introduction to the Mathematics of Statistics.* «The Riverside Press Cambridge» (ohne Jahr), pag. 262.
11) *Variations- und Erblichkeitsstatistik.* Berlin, 1929, pag. 27.
12) *Statistical Methods for Research Workers.* Edinburgh-London, 1930, pag. 158.
13) *Vorlesungen über die Grundzüge der mathematischen Statistik.* Lund, (ohne Jahr), pag. 91.
14) *Handbuch der mathematischen Statistik.* Leipzig und Berlin, 1930, pag. 165.
15) *Grundbegriffe und Grundprobleme der Korrelationstheorie.* Leipzig-Berlin, 1925, pag. 90.
16) *Korrelationsrechnung.* Leipzig und Berlin, 1928, pag. 48.
17) *Statistique mathématique.* Paris, 1928, pag. 224.
18) *Les principes de la statistique mathématique.* Paris, 1933, pag. 202.
19) *Spezielle psychophysische Massmethoden.* Berlin-Wien, 1920, pag. 126-145.
20) LAPLACE. *Théorie analytique des probabilités.* Paris 1814, pag. 382-383.

S. D. POISSON. *Lehrbuch der Wahrscheinlichkeitsrechnung und deren wichtigsten Anwendungen.* Deutsch von Dr. C. H. Schnuse, Braunschweig 1841, pag. 184-186

Prof. Dr. I. DIENGER, *Ausgleichung der Beobachtungsfehler.* Braunschweig, 1857, pag. 36-38.

GEORGE BIDDELL AIRY, *On the Algebraical and numerical theory of errors of observations and the combination of observations.* Cambridge - London, 1861, pag. 30-33.

G. ZACHARIAE. *De mindste Kvadraters Methode.* Nyborg, 1871, pag. 102-105.

Prof. Dr. MEYER, *Vorlesungen über Wahrscheinlichkeitsrechnung.* (Deutsch bearbeitet von E. Czuber). Leipzig, 1879, pag. 278-279.

AUGUSTO OTTAVIO FORTI. *La teoria degli errori e il metodo dei minimi quadrati.* Milano, 1880, pag. 36-39.

Prof. HARALD WESTERGAARD. *Die Lehre von der Mortalität und Morbidität.* Jena, 1901, pag. 190-191.

Dr. G. I. D. MOUNIER. *Hets over de grondslagen van de Methode der kleinste kwadraten.* «Archief voor de Verzekerings-Wetenschap en aanverwante Vakken's». — Gravenhage, 1902, pag. 263-265.

Prof. Dr. E. BLASCHKE. *Vorlesungen über mathematische Statistik* Leipzig und Berlin, 1906, pag. 187.

N. SABUDSKI. *Die Wahrscheinlichkeitsrechnung, ihre Anwendung auf das Schiessen und auf die Theorie des Einschiessens.* (Uebersetzt von Ritter von Eberhard). Stuttgart, 1906, pag. 156-157.

ALFONS CAPPELLERI, *Einführung in die Ausgleichungsrechnung.* Leipzig und Wien 1907, pag. 21-23.

S. BURILEANO. *Probabilité du tir.* Paris, 1911, pag. 102-103.

H. POINCARÉ. *Calcul des probabilités.* Paris, 1912, pag. 110-113.

Prof. E. CZUBER. *Wahrscheinlichkeitsrechnung und ihre Anwendung auf Fehlerausgleichung.* « Statistik und Lebensversicherung ». Band I. Leipzig und Berlin, 1914, pag. 303-305.

GUIDO CASTELNUOVO. *Calcolo delle probabilità.* Volume I. Bologna, 1925, pag. 146-147.

WILLIAM BURNSIDE. *Theory of probability.* Cambridge, 1928, pag. 88-91, pag. 94.

Prof. R. v. MISES. *Vorlesungen aus dem Gebiete der angewandten Mathematik,* 1. Band. Leipzig und Berlin, 1931, pag. 216-217, pag. 402-403.

21) *Allgemeine Herleitung eines Satzes* von LAPLACE. « Oesterreichische Revue », XXXIX. Jahrgang, Wien, 1914.

22) *Mémoire sur la probabilité des erreurs d'après la méthode des moindres carrés* (vergl. auch Meyer (20)).

23) ABSCHNITT C. *Die Herleitung der Wahrscheinlichkeit eines Verlustes oder Gewinnes auf kombinatorischer Grundlage.* « Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker », 14. Heft, Bern, 1919.

24) *Ueber die Berechnung der Reserven und des Risico bei der Lebensversicherung.* « Rundschau der Versicherungen » von Dr. E. A. MASIUS, Leipzig, 1868.

25) *Die Laplaoesche Methode der Ausgleichung von Beobachtungsfehlern bei zahlreichen Beobachtungen,* Wien, 1875.

26) *Theorie der Beobachtungsfehler.* Leipzig, 1891.

27) *Ueber die Ermittelung der Genauigkeit der Beobachtungen bei der Analyse periodischer Erscheinungen und in der Methode der kleinsten Quadrate.* Helsingfors, 1901.

28) *Methode der kleinsten Quadrate, Anhang zur Wittstein'schen Uebersetzung des Lehrbuchs der Differential- und Integralrechnung von Navier.* Hannover, 1875.

29) *Einführung in die Wahrscheinlichkeitsrechnung* « Deutsche Ausgabe » von Dr. URBAN, Leipzig, 1927.

30) *Skandinavisk Aktuarietidskrift.* Uppsala, 1931.

LOUIS I. DUBLIN,
ALFRED J. LOTKA
AND MORTIMER SPIEGELMAN

The Construction of Life Tables by Correlation

For a number of years past the *Statistical Bureau* of the *Metropolitan Life Insurance Company* has prepared, as a matter of routine, a series of life tables representing the mortality experience among its Industrial policyholders. It was understood from the beginning that this experience, taken from a group numbering many millions of persons, would afford an index not only of the mortality in the entire industrial population, but also, indirectly, of the mortality in the general population of the United States.

How accurate this index is, was not, however, fully realized until recently, when, for the purposes of a more extended study which will appear in book form, an investigation was made of the correlation between the mortality at specified ages among the Industrial policyholders on the hand, and the corresponding mortality in the general population on the other. This investigation revealed so close a relation that it has been found actually possible, from a life table constructed on the basis of the mortality among our Industrial policyholders, to prepare a corresponding table for the general population, which has distinct value as a preliminary table until the official figures of the Census Bureau become available.

The application of the new method has thus many advantages. In the first place, Federal mortality and population statistics do not become available until two or more years after the period to which they relate. Statistics of deaths and exposure among the policyholders of the life insurance company are available currently. The method illustrated, therefore, enables us to obtain advance information regarding the trend of longevity in the general population.

Another advantage of the method is as follons. During postcensal periods, in computing age-specific mortality in the general population, we are forced to rely upon *estimates* of the population classified by

color, sex and age. In past decades estimates of this kind proved fairly satisfactory. But the task has proved particularly difficult in the present decade, owing to the rapid decline that has taken place in the birthrate. An independent method of computing the life table characteristics is therefore of special value.

The method is best described by means of an example. In table I, following, are shown the values of q_x for age $x = 22$, for a succession of years, 1920 to 1930, according to

A) the experience among white males in the Death Registration States of 1920; and

B) the experience among the white male policyholders of the Metropolitan Life Insurance Company, Industrial Department.

From the data in columns *A* and *B*, the correlation coefficient and the regression equation between the mortality q_{22} in the general United States population and among the Industrial policyholders of the Metropolitan Life Insurance Company was computed. The correlation was exceedingly close, the value of the coefficient being $.982 \pm .011$.

The regression equation by which the mortality q'_{22} in the United States is determined from that among the Industrial policyholders q_{22} is

$$q'_{22} = .5462\ q_{22} + 1.552 \qquad (1)$$

TABLE I.

Chances per 1,000 *of Dying within One Year after Attaining Age* 22 *among White Males in the United States Death Registration States of* 1920, *and the Corresponding Chances among White Males in the Industrial Department of the Metropolitan Life Insurance Company,* 1920-1930.

YEAR	United State Death Registration States of 1920 (*A*)	Metropolitan Life Insurance Company Industrial Department (*B*)	YEAR	United States Death Registration States of 1920 (*A*)	Metropolitan Life Insurance Company Industrial Department (*B*)
1930	3.41	3.30	1924 . . .	3.64	3.89
1929	3.63	3.81	1923 . . .	3.79	3.94
1928	3.60	3.71	1922 . . .	3.59	3.97
1927	3.40	3.59	1921 . . .	3.75	4.15
1926	3.63	3.56	1920 . . .	4.94 *	6.16
1925	3.65	3.81			

* U. S. Life Table (Foudray) 1919-1920.

Thus, for example, the mortality at age twenty-two among the white male Industrial policyholders being 3.18 per 1,000, in 1931, we find for the corresponding mortality in the general population

$$q'_{22} = .5462 \times 3.18 + 1.552$$
$$= 3.29 \qquad (2)$$

In this way, forming a separate regression equation for every age required, the values of q'_x for the United States population were computed for quinquennial ages $7, 12 \ldots 72$, and also for ages 1, 2, 3, 4.

For age zero, the experience of the Metropolitan Life Insurance Company for obvious reasons does not furnish adequate data. In their place infant mortality rates as observed in New York State (excepting New York City) were employed * as the basis of the correlation.

At ages seventy-five and over, also, adequate data are not obtainable from the experience of the Industrial Department of the Metropolitan Life Insurance Company, because Industrial policies become " paid up " at age seventy-five and the data readily available are those relating to policies on which premiums are currently being paid. At ages above 75, however, approximate data will suffice, since we are dealing with only about one third of the life table cohort. Actually the values of q_x from the life table of the previous calendar year, 1930, were simply accepted for 1931. This is the more permissible since at these higher ages there is usually relatively little variation from year to year. The use of data for 1930 is particularly advantageous, inasmuch as it was a census year, for which data are known with greater accuracy than for other years.

The values of q_x having been obtained as indicated above for every fifth year of life and individually for the first five years of life, the construction of the complete life table can now be effected by any one of the established methods, such as that of King or of Jenkins.

We present herewith in Table II a set of life tables for the United States in the years 1930, 1931, 1932 and 1933, computed by the method described above. The table for 1930 was prepared by the

---

* These rates for New York State become available in provisional form shortly after the close of the current year, and long before the Federal data.

way of a test, to ascertain how nearly it agreed with a table prepared directly from the mortality statistics and the population census in the United States (Registration States of 1920). A glance at the figures shows that the correspondence between the figures of the " direct " life table and that computed by correlation for the year 1930 is very close indeed so far as the expectation of life e_x^o is concerned, and is also very good as regards the values of q_x. For example, at age one the expectation of life as computed directly from Census data is 62.22 years ; as computed by correlation method, 62.31 years. At age 12, the corresponding values are 53.22 years from Census data and 53.42 years by correlation. At age 62, the figures are 35.86 years from Census data, 36.07 by correlation. At age 62, the figures are 13.50 from Census data and 13.63 by correlation. Making a similar comparison regarding the value of q_x, we note at age one the figure of 9.00 from Census data as against 9.10 by correlation ; at age 12, 1.51 from Census data, 1.56 by correlation ; at age 32, 4.29 from Census data and 4.54 by correlation ; at age 62, 31.30 from Census data and 29.71 by correlation. All these agreements are most satisfactory. Only at age 0 is there a more considerable divergence in the values of q_x, namely 61.05 from Census data and 63.26 by correlation. It will be remembered that at this point data from our Industrial experience are inadequate and New York State data had to be used as a basis for the correlation.

In the column of survivors, l_x, the agreement may seem at first sight rather less close, as for example, at age one, 93,895 from Census data and 93,674 by correlation. But it must be remembered that, following the usual custom, the entries here are given to five significant figures, and that fair agreement in the first three figures is all that can reasonably be expected, and is quite sufficient to give a very good idea of the course of the l_x curve. In graphic representation on any ordinary scale, the two l_x curves would be practically indistinguishable.

THE CORRELATION METHOD APPLIED DIRECTLY TO THE EXPECTATION OF LIFE. — In the description above the usual procedure has been followed of building up the life table on the basis of the q_x column. This involves two steps, first the computation of the values of q_x by correlation, and second the complete process of preparing a life table from the values of q_x as a basis.

If the main object in view is to determine the expectation of

life e_x^0 at successive ages, or at any particular age, such as age zero, then the work can be very greatly reduced by applying the method of correlation directly to the already computed values of e_x^0 of the Metropolitan experience. Or, better still, the method of correlation can be applied, not directly to the values of e_x^0, but to the deviations of these values, in successive calendar years, from a trend line passed through the series. Thus, in Table III, there are exhibited the series of values of e_x^0 for the years 1921 to 1930 for the United States (line designated as A in the table) and for the Industrial policyholders of the Metropolitan Life Insurance Company (designated as B). A straight line trend* was put through each series of such values at every tenth year of life. The deviations from these trend lines were found and are exhibited in Table IV, which also gives the trend line equations. These deviations were then made the basis of a computation of the correlation and regression equation. Thus, if we denote by δ' the deviation from the trend line in values of e_x^0 for the United States, and by δ the corresponding value among the Industrial policyholders, we have a regression equation, at age 20 for example,

$$\delta' = .7746\,\delta\,. \tag{3}$$

If, now, the value of e_x^0 for the United States white males at age 20 is desired for the year 1933, for example, we may first find the point B' in the trend line corresponding to an observed value of $e_{20}^0 = 45.24$ in the Industrial insurance experience. Thus

$$\begin{aligned} B' &= 43.5756 - .0427 \times (1933\text{-}1925) \\ &= 43.23 \end{aligned} \tag{4}$$

We therefore find that

$$\delta = 45.24 - 43.23 = 2.01\,. \tag{5}$$

Similarly, for the corresponding point A' on the trend line for the United States data we have

$$A' = 45.9722 - .1362 \times (1933\text{-}1925) = 44.88\,. \tag{6}$$

Now by equation (3)

$$\delta' = .7746 \times 2.01 = 1.56\,. \tag{7}$$

* The years 1921 to 1929 were included in computing the trend.

To the trend value, A', we add δ', so our estimated value of e°_{20} is $44.88 + 1.56 = 46.44$.

It is evident that the three equations (3), (4) and (6), can be combined into a single equation. For example,

$$A - (45.9722 - .1362\,t) = .7746\,\{B - (43.5756 - .0427\,t)\} \qquad (8)$$

where t is the time in years dated from a zero at 1925.
Or

$$A = 12.218 - .1031\,t + .7746\,B\,. \qquad (9)$$

In Table III, the single equation thus obtained is shown for every tenth year of life.

The results of the computation of e°_{x} in this way are exceedingly good, as seen in columns (12) to (14) of Table III. For example, at age zero the expectation of life computed directly from the observed mortality in the United States in 1930 was 59.39, as against 59.60 computed by correlation through the Metropolitan Life Insurance Company data. At age ten, the two corresponding figures are 55.07 as against 55.06. At age 20, 46.08 as against 46.03 ; at age thirty 37.57 as against 37.46, etc. throughout the table. It will be seen that the agreement is almost perfect, except at age zero, and even there the discrepancy is small.

It is worth repeating that these results have been obtained without actually computing a new life table, this method differing in respect from the one first described.

If values of e°_{x} at every age of life are computed, it is also possible by the aid of well known relations between the several life table functions, to compute the other columns of the life table. The method described, however, probably has its chief utility in the computation of abridged tables setting forth only values of e°_{x} at selected ages.

Conclusion.

From examples given above, it is quite clear that a very serviceable life table, closely representative of the general population, can be obtained on the basis of mortality statistics compiled for a sufficiently large group forming part of that population. In this case, the sample has included about 16 million white lives. It is not necessary that the part of the population so employed should be an unbiased

sample of the total, although in this example the group is very widely distributed geographically and industrially. The bias, if any, of the sample, is fully compensated, according to the method described, since this method is based on the observed correlation and regression equation between the mortality rates in the sample as compared with that in the general population. The trend in the smaller group appears to reflect with accuracy the changes in the whole population.

Aside from its obvious practical interest, the method, therefore, also presents this somewhat unique feature, that an admittedly biased sample is used as a gauge of the population from which it is drawn. As to practical applications, it has already been pointed out that these are at least two-fold. On the one hand, the earlier availability of insurance statistics as compared with official government statistics of the general population, enables us, profiting from the method described, to prepare life tables for the general population before the government statistics are published. In the examples shown it is seen that a life table so prepared in advance has been found to check very closely with the corresponding life table computed directly from Census data of the general population when these come to hand.

The second advantage is that in intercensal years the information regarding the exposure among policyholders of a life insurance company is much more precise than in the general population. This advantage should be particularly felt at the present moment, because the rapidly declining birthrate has made it very difficult, if not impossible, to obtain reasonably close estimates of the population and its age distribution in the years since the last census. Just at the present time, therefore, the life tables prepared by the correlation method may be of particular value.

TABLE II. — *Life Tables by Correlation of Mortality Rates; Mortality Rate per* 1000, *Expectation of Life, and Survivors from Birth, for White Males and White Females in the United States Death Registration States of* 1920 *for Each Year from* 1930 *to* 1933.

AGE	Mortality Rate (1000 q_x)					Expectation of Life (e^0_x)					Survivors from Birth (l_x)				
	from Census data	by correlation method				from Census data	by correlation method				from Census data	by correlation method			
	1930	1930	1931	1932	1933	1930	1930	1931	1932	1933	1930	1930	1931	1932	1933
						White Males.									
0*	61.05	63.26	63.10	59.44	58.82	59.39	59.33	59.46	60.21	60.28	100 000	100 000	100 000	100 000	100 000
1	9.00	9.10	8.70	6.40	6.41	62.22	62.31	62.43	62.99	63.02	93 895	93 674	93 690	94 056	94 118
2	5.00	5.55	5.65	5.10	5.07	61.78	61.87	61.98	62.39	62.42	93 050	92 821	92 875	93 454	93 515
7	1.97	2.00	2.02	1.82	1.77	57.78	57.96	58.05	58.37	58.39	91 520	91 165	91 233	91 955	92 028
12	1.51	1.56	1.53	1.44	1.39	53.22	53.42	53.51	53.80	53.80	90 786	90 418	90 486	91 264	91 361
17	2.53	2.50	2.33	2.18	2.12	48.70	48.90	48.97	49.23	49.22	89 941	89 566	89 683	90 496	90 623
22	3.41	3.35	3.29	3.15	3.14	44.37	44.58	44.61	44.84	44.82	88 642	88 285	88 469	89 333	89 477
27	3.69	3.64	3.69	3.50	3.50	40.12	40.31	40.34	40.54	40.52	87 088	86 769	86 950	87 872	88 012
32	4.29	4.54	4.54	4.31	4.26	35.86	36.07	36.11	36.27	36.25	85 402	85 062	85 220	86 214	86 363
37	5.50	5.59	5.66	5.28	5.18	31.66	31.90	31.95	32.07	32.03	83 405	82 987	83 129	84 225	84 405
42	7.63	7.45	7.58	7.26	7.24	27.59	27.84	27.90	27.97	27.92	80 811	80 417	80 512	81 726	81 928
47	10.47	10.00	10.01	9.80	10.00	23.70	23.92	24.00	24.03	23.99	77 366	77 110	77 167	78 431	78 600
52	14.92	14.43	14.29	13.73	14.16	20.03	20.20	20.28	20.27	20.26	72 795	72 739	72 806	74 128	74 179
57	21.30	21.21	20.67	20.03	20.51	16.61	16.78	16.83	16.76	16.79	66 728	66 767	66 941	68 379	68 265
62	31.30	29.71	29.45	29.69	30.07	13.50	13.63	13.65	13.55	13.62	58 796	59 047	59 339	60 710	60 459
67	45.43	44.90	44.78	45.78	44.96	10.73	10.81	10.82	10.73	10.82	48 830	49 336	49 627	50 631	50 438
72	67.60	65.82	65.52	67.36	65.51	8.32	8.36	8.37	8.33	8.37	37 085	37 682	37 943	38 417	38 545
77	100.87	100.87	100.87	100.87	100.87	6.33	6.33	6.33	6.33	6.33	24 449	24 991	25 190	25 349	25 591
82	147.82	147.82	147.82	147.82	147.82	4.79	4.79	4.79	4.79	4.79	12 988	13 270	13 375	13 465	13 588
87	203.81	203.81	203.81	203.81	203.81	3.58	3.58	3.58	3.58	3.58	5 146	5 258	5 299	5 335	5 383
92	281.57	281.57	281.57	281.57	281.57	2.52	2.52	2.52	2.52	2.52	1 371	1 401	1 412	1 422	1 435

* Computations on this line are based on a correlation of q_0 in the United States Death Registration States of 1920 with deaths under 1 year per 1000 live births in New York State, exclusive of New York City.

AGE	Mortality Rate (1000 q_x)					Expectation of Life (e^0_x)					Survivors from Birth (l_x)				
	from Census data	by correlation method				from Census data	by correlation method				from Census data	by correlation method			
	1930	1930	1931	1932	1933	1930	1930	1931	1932	1933	1930	1930	1931	1932	1933
White Females.															
0*	48.38	49.71	49.58	46.58	46.07	63.02	62.78	62.96	63.52	63.89	100 000	100 000	100 000	100 000	100 000
1	7.86	8.01	6.86	5.03	4.86	65.20	65.04	65.22	65.60	65.95	95 162	95 029	95 042	95 342	95 393
2	4.42	4.72	4.92	4.35	4.47	64.72	64.56	64.66	64.93	65.27	94 414	94 268	94 390	94 862	94 929
7	1.56	1.59	1.55	1.44	1.41	60.59	60.53	60.64	60.81	61.15	93 121	92 829	92 932	93 556	93 627
12	1.12	1.09	1.08	1.06	.91	55.95	55.88	55.99	56.14	56.46	92 547	92 264	92 383	93 021	93 132
17	2.02	2.04	1.94	1.84	1.67	51.33	51.26	51.36	51.50	51.78	91 886	91 611	91 751	92 408	92 585
22	3.02	2.84	2.77	2.66	2.33	46.93	46.85	46.92	47.04	47.26	90 770	90 523	90 711	91 403	91 690
27	3.35	3.38	3.25	3.13	3.00	42.64	42.53	42.58	42.68	42.84	89 339	89 139	89 376	90 103	90 509
32	3.78	3.80	3.87	3.69	3.61	38.35	38.25	38.28	38.36	38.50	87 783	87 571	87 831	88 605	89 053
37	4.44	4.37	4.56	4.38	4.28	34.07	33.97	34.02	34.07	34.20	86 044	85 840	86 044	86 870	87 351
42	5.83	5.73	5.93	5.69	5.49	29.86	29.74	29.83	29.85	29.95	83 940	83 780	83 898	84 781	85 320
47	7.91	8.08	8.09	7.80	7.70	25.78	25.65	25.76	25.74	25.83	81 208	81 053	81 113	82 079	82 665
52	11.03	11.40	11.08	11.03	10.78	21.86	21.75	21.86	21.81	21.88	77 605	77 357	77 460	78 466	79 082
57	16.33	16.52	16.16	16.44	15.91	18.15	18.07	18.14	18.10	18.14	72 709	72 369	72 596	73 500	74 222
62	24.61	24.93	24.86	24.97	24.86	14.74	14.67	14.73	14.71	14.72	65 942	65 571	65 846	66 595	67 375
67	37.19	38.52	37.87	38.14	38.12	11.68	11.66	11.71	11.69	11.70	56 875	56 321	56 641	57 23[illegible]	57 920
72	57.69	57.27	56.51	56.80	56.49	9.01	9.02	9.04	9.03	9.04	45 263	44 672	45 101	45 504	46 085
77	89.12	89.12	89.12	89.12	89.12	6.81	6.81	6.81	6.81	6.81	31 613	31 256	31 633	31 885	32 327
82	135.22	135.22	135.22	135.22	135.22	5.12	5.12	5.12	5.12	5.12	17 986	17 781	17 993	18 137	18 388
87	191.21	191.21	191.21	191.21	191.21	3.86	3.86	3.86	3.86	3.86	7 660	7 572	7 662	7 724	7 831
92	256.88	256.88	256.88	256.88	256.88	2.87	2.87	2.88	2.87	2.88	2 258	2 232	2 259	2 277	2 308

* Computation on this line are based on a correlation of q_0 in the United States Death Registration States of 1920 with deaths under 1 year per 1000 live births in New York State, exclusive of New York City.

TABLE III. — *Correlation between the Expectation of Life Among Industrial Policyholders of the Metropolitan Life Insurance Company (B) and White Males in the General U. S. Population (A)**, 1921-1929.

AGE	Symbol	Expectation of life for year specified											Standard error of computed value	Regression Equation **
		1921	1922	1923	1924	1925	1926	1927	1928	1929	1930			
											Observed	Computed		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
0	*B*	55.08	55.04	54.55	55.62	55.51	55.02	56.42	55.88	55.78				
	A	57.66	57.58	56.94	57.88	57.88	57.42	58.83	57.89	58.32	59.39	59.60	.15	$A = 3.504 - .0234\,t + .9799\,B$
10	*B*	52.73	52.36	51.69	52.62	52.38	52.10	52.62	52.34	52.20				
	A	55.58	55.08	54.56	54.99	54.84	54.50	55.08	54.26	54.38	55.07	55.06	.13	$A = 12.978 - .1006\,t + .7993\,B$
20	*B*	44.17	43.65	42.94	43.87	43.61	43.28	43.79	43.52	43.35				
	A	46.90	46.28	45.76	46.17	46.02	45.60	46.17	45.38	45.47	46.08	46.03	.13	$A = 12.218 - .1031\,t + .7746\,B$
30	*B*	36.23	35.70	34.96	35.78	35.56	35.17	35.69	35.41	35.26				
	A	38.57	37.90	37.43	37.76	37.60	37.16	37.67	36.95	37.04	37.57	37.46	.12	$A = 10.869 - .1027\,t + .7514\,B$
40	*B*	28.44	27.92	27.20	27.94	27.75	27.30	27.73	27.49	27.35				
	A	30.35	29.74	29.30	29.57	29.39	28.94	29.40	28.72	28.77	29.20	29.08	.11	$A = 9.564 - .1012\,t + .7149\,B$
50	*B*	20.97	20.54	19.89	20.63	20.39	20.05	20.41	20.23	20.08				
	A	22.51	21.97	21.58	21.84	21.66	21.26	21.66	21.09	21.12	21.47	21.26	.10	$A = 7.674 - .0976\,t + .6858\,B$
60	*B*	14.20	13.89	13.46	14.09	13.83	13.58	13.99	13.77	13.59				
	A	15.43	14.97	14.68	14.93	14.80	14.45	14.80	14.33	14.39	14.71	14.51	.09	$A = 4.790 - .0783\,t + .7208\,B$
70	*B*	8.81	8.54	8.30	8.79	8.70	8.53	8.82	8.78	8.63				
	A	9.64	9.37	9.15	9.39	9.29	8.99	9.31	8.87	8.96	9.24	8.92	.09	$A = 3.392 - .0805\,t + .6732\,B$
80	*B*	5.19	5.09	4.94	5.18	5.14	5.09	5.15	5.14	5.08				
	A	5.38	5.37	5.18	5.39	5.23	5.07	5.35	5.01	5.09	5.35	5.08	.09	$A = 1.326 - .0375\,t + .7638\,B$

* Death registration states as constituted in 1920.
** t = number of years difference between year of estimate and 1925.

TABLE IV. — *Method of Computing Correlations between the Expectation of Life among Industrial Policyholders of Metropolitan Life Insurance Company (B) and White Males in the General U. S. Population* (A), 1921-1929.*

AGE	Symbol	Trend equation for period 1921–1929 **	Deviations in Values of (e^0_x) from Corresponding Trend Values									Regression Equation between δ and δ'
			Symbol	1921	1922	1923	1924	1925	1926	1927	1928	1929	
0	*B*	$B' = 55.4333 + .1410\,t$	δ	+.21	+.03	—.60	+.33	+.08	—.55	+.70	+.02	—.22	
	A	$A' = 57.8222 + .1148\,t$	δ'	+.30	+.10	—.65	+.17	+.06	—.52	+.78	—.28	+.04	$\delta' = .9789\ \delta$
10	*B*	$B' = 52.3378 - .0140\,t$	δ	+.34	—.02	—.68	+.27	+.04	—.22	+.31	+.04	—.08	
	A	$A' = 54.8122 - .1118\,t$	δ'	+.32	—.07	—.48	+.07	+.07	—.20	+.49	—.22	+.02	$\delta' = .7993\ \delta$
20	*B*	$B' = 43.5756 - .0427\,t$	δ	+.42	—.05	—.72	+.25	+.03	—.25	+.30	+.07	—.05	
	A	$A' = 45.9722 - .1362\,t$	δ'	+.38	—.10	—.48	+.06	+.05	—.24	+.47	—.18	+.04	$\delta' = .7746\ \delta$
30	*B*	$B' = 35.5289 - .0650\,t$	δ	+.44	—.02	—.70	+.19	+.03	—.29	+.29	+.08	—.01	
	A	$A' = 37.5644 - .1515\,t$	δ'	+.40	—.12	—.44	+.04	+.04	—.25	+.41	—.16	+.08	$\delta' = .7514\ \delta$
40	*B*	$B' = 27.6800 - .0872\,t$	δ	+.41	—.02	—.65	+.17	+.07	—.29	+.22	+.07	+.02	
	A	$A' = 29.3533 - .1635\,t$	δ'	+.34	—.10	—.38	+.05	+.04	—.25	+.37	—.14	+.07	$\delta' = .7149\ \delta$
50	*B*	$B' = 20.3544 - .0672\,t$	δ	+.35	—.02	—.60	+.21	+.04	—.24	+.19	+.08	—.01	
	A	$A' = 21.6322 - .1437\,t$	δ'	+.30	—.09	—.34	+.06	+.03	—.23	+.32	—.11	+.06	$\delta' = .6858\ \delta$
60	*B*	$B' = 13.8222 - .0375\,t$	δ	+.23	—.04	—.44	+.23	+.01	—.20	+.24	+.06	—.08	
	A	$A' = 14.7533 - .1053\,t$	δ'	+.26	—.10	—.28	+.07	+.05	—.20	+.26	—.11	+.06	$\delta' = .7208\ \delta$
70	*B*	$B' = 8.6556 + .0130\,t$	δ	+.21	—.08	—.33	+.15	+.04	—.14	+.14	+.09	—.08	
	A	$A' = 9.2189 - .0717\,t$	δ'	+.13	—.06	—.21	+.10	+.07	—.16	+.23	—.13	+.03	$\delta' = .6732\ \delta$
80	*B*	$B' = 5.1111 + .0007\,t$	δ	+.08	—.02	—.17	+.07	+.03	—.02	+.04	+.03	—.03	
	A	$A' = 5.2300 - .0370\,t$	δ'	—	+.03	—.12	+.12	—	—.12	+.19	—.11	+.01	$\delta' = .7638\ \delta$

* Death registration states as constituted in 1920.
** t = number of years difference between any specified year and 1925.

PROF. CORRADO GINI, *Direttore responsabile.*
Roma — «L'Universale» Tipografia Poliglotta.