GIUSEPPE POMPILJ

SULLA MEDIA GEOMETRICA E SOPRA UN INDICE DI MUTABILITÀ CALCOLATI MEDIANTE UN CAMPIONE (*)

SUNTO: Si determinano la media e la varianza delle variabili casuali descritte dai due indici « media geometrica » e « indice di mutabilità » al variare del campione.

INTRODUZIONE

I.

Dico subito che i due parametri che figurano nel titolo di questo lavoro non hanno alcun rapporto tra di loro; si trovano qui riuniti solo perché il metodo con cui verranno studiati ricade, come preciserò meglio nel seguito di questa parte introduttiva, in un unico schema logico che inquadra tutta la statistica.

Dei due parametri la *media geometrica* non ha bisogno di presentazione, invece, circa l'*indice di mutabilità*, è necessario precisare che mi occuperò dell'indice lineare atto a misurare la mutabilità delle serie sconnesse introdotto da C. GINI fin dal 1912 al n. 82 della ormai classica memoria *Variabilità e mutabilità*, che trovasi riprodotta, insieme a varie altre, in un volume (1) che raccoglie i principali lavori di C. GINI sulla *variabilità* e la *concentrazione*.

Come si vede, i due indici si riferiscono il primo a fenomeni le cui modalità sono quantitative, il secondo a fenomeni le cui modalità sono qualitative. Il fatto naturalmente non è casuale, ma di proposito ho scelto due indici riferentesi ai due tipi diversi di fenomeni per mettere in evidenza le grandi analogie che passano, nell'ordine d'idee che

(*) Nota presentata dal socio nazionale FRANCESCO SEVERI.

(1) Ved. C. GINI, *Memorie di metodologia statistica*, Milano, Giuffrè, 1939, Vol. I, pagine 189-358.

esporrò, tra i metodi con cui vanno trattati i valori segnaletici riferentesi a fenomeni quantitativi (cioè a modalità quantitative) e quelli riferentesi a fenomeni qualitativi.

Approfittando di queste analogie, ed essenzialmente allo scopo di semplificare l'esposizione, nel seguito di questa parte introduttiva, salvo avviso in contrario, mi riferirò sempre a fenomeni quantitativi riserbandomi di far poi osservare come le stesse considerazioni valgano, *mutatis mutandis*, per i fenomeni qualitativi.

II.

Le considerazioni che seguono non hanno pretese di originalità; esse in massima parte sviluppano, illustrandone le varie conseguenze, vedute e concezioni esplicitamente esposte da G. Castelnuovo (2) fin dal lontano 1918. Tuttavia mi dilungherò su questo argomento perché ritengo non inutile raccogliere e coordinare tutte queste varie considerazioni che, almeno per quel che mi risulta, si trovano solo sparse in varie opere alcune delle quali citerò di volta in volta.

Si dice comunemente che la statistica studia i fenomeni collettivi o di massa; ma con ciò si vengono a confondere due aspetti diversi (3) della nostra scienza (4): la statistica pura e la statistica applicata.

La statistica pura non studia il fenomeno collettivo, ma la rilevazione numerica di esso, arrivando a risultati che hanno una validità astratta, come edificio logico coerente, anche se le conseguenze pratiche di queste deduzioni saranno poi in netto contrasto con successive indagini sul fenomeno medesimo, alla stessa stregua che la dinamica di Newton rimane valida, nel suo modello, nonostante la precessione del perielio di Mercurio.

La statistica applicata invece studia direttamente il fenomeno prima definendolo, cioè delimitandolo, poi escogitando il metodo migliore per farne il rilevamento ed infine interpretando i risultati astratti

(2) Ved. per esempio: G. Castelnuovo, *Calcolo della probabilità*, Bologna, Zanichelli. Prefazione.

(3) Ved. per esempio: A. C. Aitken, *Statistical mathematics*, Oliver & Boyd, Londra, 1942.

(4) Si è discusso a lungo per decidere se la statistica sia una scienza oppure no; ma credo che tutta la questione si riduca al significato più o meno comprensivo che si dà alla parola « scienza »; personalmente propendo a darle un significato tale che si possano chiamar scienze sia la matematica che la statistica.

forniti dalla statistica pura come misure, relazioni e leggi del particolare fenomeno in esame.

In sede di statistica pura comunemente si contrappongono *universo* e *campione* le cui relazioni dànno luogo a due diversi problemi:

PROBLEMA DIRETTO: conosciuto l'universo, studiare le proprietà dei campioni;

PROBLEMA INVERSO: conosciuto uno o più campioni, studiare l'universo.

Dei due problemi il più interessante è senza dubbio il secondo; e sembrava in molti casi risolto, ma, come ha fatto osservare C. GINI (5), le soluzioni proposte risentono del « peccato originale del calcolo delle probabilità » e pertanto non si reggono in piedi. La conclusione cui l'eminente statistico arriva è che il problema è in generale insolubile a meno che non s'introducano delle ipotesi che vanno di volta in volta determinate.

L'interesse del problema diretto risiede invece solo nel fatto che la sua soluzione è il presupposto necessario (ma non sufficiente) per affrontare lo studio del problema inverso. E posso aggiungere che esso è già virtualmente risolto, salvo alcune difficoltà di dettaglio più di natura matematica che statistica.

Sic stantibus rebus, mi soffermerò essenzialmente sul problema diretto che cercherò di presentare in uno schema astratto che inquadra tutta la statistica pura. Questa ricerca di un'impostazione completamente astratta rappresenta un punto essenziale nello sviluppo di ogni scienza perché, come scrive F. SEVERI (6), è soltanto in uno spirito d'astrazione e di generalità che si scoprono i riposti legami fra teorie dissimili e tra queste e i fatti. Del resto dall'ultimo scorcio del secolo scorso si sono avute riprove luminosissime dell'esattezza di questa concezione, al punto che il grande fisico M. PLANK pensava essere la

---

(5) Ved. C. GINI, *I pericoli della statistica*, Atti della I riunione scientifica della Soc. Italiana di Statistica, 1939; *A proposito dei testi di significatività*, Atti della VI e VII riunione scientifica della Soc. Italiana di Statistica, 1943; *I testi di significatività*, ibidem; *Sulla probabilità inversa nel caso di grandezze a distribuzione costante*, ibidem; C. GINI-G. LIVADA, *Sulla probabilità inversa nel caso di grandezze intensive ed in particolare sulle sue applicazioni a collaudi per masse a mezzo di campioni*, ibidem; C. GINI, *Osservazioni sulla comunicazione della dott. M. G. Geppert « Sul valore dei così detti testi di significatività »*, ibidem.

(6) Ved. F. SEVERI, *Matematica e matematici d'oggi*, Atti del Convegno matematico, Roma, 1942.

scienza tanto più vicina alla conoscenza dell'Universo, concepito come realtà obbiettiva, quanto più si allontana dall'originario carattere antropomorfico e intuitivo per elevarsi nell'astratto.

Nel caso della statistica pura questa astrazione si raggiunge immediatamente sostituendo all'universo la *variabile casuale* e al campione il suo esclusivo contenuto numerico che chiamerò *campione astratto* o anche *campione della variabile casuale*. Tale sostituzione è possibile in virtù del seguente principio fondamentale: *ad ogni fenomeno collettivo è associata una* VARIABILE CASUALE *che ne rappresenta l'aspetto quantitativo, ed ogni rilevamento del fenomeno consiste in più esperienze semplici che forniscono un* CAMPIONE *della variabile casuale stessa.*

Credo che questo principio non abbia bisogno di delucidazioni; esso riattacca direttamente la statistica pura al fondamentale concetto di variabile casuale. E anche se il nome di variabile casuale non è certamente felice, il concetto è di una tale potenza di sintesi da permettere, attraverso procedimenti snelli ed eleganti, di presentare i vari risultati in una forma chiara e rigorosa.

Inoltre si viene così automaticamente a definire il « principio di identità » per la statistica pura. L'esatta definizione di tale principio è essenziale tutte le volte che si voglia dare una trattazione organica-rigorosa di una qualunque scienza e quindi è opportuno soffermarsi su questo argomento almeno per quel che riguarda l'oggetto del nostro studio.

A tale scopo basta osservare che fenomeni diversi possono essere associati a variabili casuali eguali e che in tal caso, in base al principio fondamentale sopra enunciato, la statistica pura è costretta a considerare come eguali questi fenomeni diversi; e non sembri questa una contraddizione, perché il concetto di eguaglianza è un concetto relativo, in quanto si parla comunemente di eguaglianza solo tra enti diversi; così due stoffe sono eguali, rispetto al colore, se hanno lo stesso colore, due segmenti sono eguali, rispetto alla geometria metrica, se con un movimento si possono sovrapporre, due fenomeni collettivi sono eguali, rispetto alla statistica pura, se sono associati con una stessa variabile casuale.

Conviene ora osservare più da vicino la struttura di un campione.

Eseguiti N esperimenti semplici su una variabile casuale si ottengono N valori se la variabile casuale è semplice, oppure N coppie ordinate di valori se la variabile casuale è doppia e così via; l'insieme

di questi N valori, o coppie di valori ecc. . . . forma il campione relativo all'esperienza consistente negli N esperimenti.

Bisogna però tener presente che nello svolgimento della teoria si suppone che questi esperimenti siano *indipendenti* tra di loro o, più precisamente, che essi vengano eseguiti secondo lo schema di BERNOULLI. Questa ipotesi non è rispettata quando gli esperimenti si eseguono con lo schema dell'estrazione in blocco, come per esempio, accade quando per la statistica delle stature non si scelgono a caso N individui ma piuttosto un blocco di N individui.

D'altra parte lo schema di BERNOULLI è quello che offre in tutti i casi delle formule notevolmente più semplici che per di più, in generale, numericamente differiscono poco dalle analoghe dello schema di estrazione in blocco; ed allora non volendo rinunziare alla comodità di lavorare con formule relativamente semplici, dobbiamo trovare quale significato abbia il loro uso nel caso che il campione sia stato ottenuto con una estrazione in blocco. A tale scopo si ricordi che i due schemi coincidono se e solo se l'universo su cui si opera è formato da infiniti individui, pertanto solo in questo caso è perfettamente lecito, dal punto di vista logico, usare le formule dello schema bernoulliano per studiare un campione ottenuto con un'estrazione in blocco. Ben magro risultato questo perché in pratica di universi infiniti non se ne incontrano! Ma a questo punto ci soccorre un'osservazione di C. GINI (7): molte volte la statistica tende a studiare il fenomeno collettivo in un suo significato che si potrebbe dire più ampio, cioè tende a cogliere quello che si riscontrerebbe qualora gli individui dell'universo diventassero infinitamente numerosi venendo così a costituire un ipotetico universo-limite di cui l'universo dato deve essere considerato un campione ottenuto bernoullianamente; se ora da questo campione ne estraiamo un altro, anche con un'estrazione in blocco, rispetto all'universo-limite quest'altro campione sarà sempre ottenuto secondo lo schema di BERNOULLI. Spiego meglio questi concetti con un esempio: il fenomeno da studiare sia la statura dei maschi ventenni italiani; molte volte quello che interessa non è tanto il sapere come si comportano rispetto all'altezza gli italiani oggi ventenni, ma piuttosto come si comporta la razza di cui gli italiani oggi ventenni sono solo degli esemplari; ed ecco così nascer l'ipotetico universo-limite, formato da infiniti individui, che si potrebbe pensar costruito raccogliendo i

(7) Ved. C. GINI, op. cit. (1), p. 236.

risultati di un numero infinito di prove sulla variabile casuale associata al fenomeno «statura degli italiani ventenni»; per studiare questo fenomeno misuriamo le stature di un blocco di 10.000 individui, si costruisce così un campione che rispetto a tutti gli italiani oggi ventenni è stato ottenuto con un'estrazione in blocco ma che di fronte all'ipotetico universo-limite è stato invece ottenuto bernoullianamente. Si potrebbe poi finalmente ancora osservare che i 10.000 individui considerati sono relativamente pochi di fronte a tutti gli italiani oggi ventenni di modo che le formule dello schema di BERNOULLI differiranno numericamente assai poco dalle analoghe dello schema dell'estrazione in blocco.

Ma chi si soffermi a considerare più da vicino l'esempio, cui più volte si è ricorso, della statistica delle stature, nota subito il seguente fatto importantissimo: le misure delle N altezze sono inevitabilmente affette da errori, quindi non conviene formalizzarsi troppo su questi numeri ed, appena N sia un po' grande, è più opportuno, anche perché semplifica i calcoli, raggruppare in un'unica classe le misure che cadono in un certo intervallo, cioè, per esempio, racchiudere in un'unica classe le altezze che cadono tra 1.600 millimetri e 1.660 millimetri e, se queste altezze sono in numero di a, dire che, degli N individui misurati, a sono alti un metro e sessantatré centimetri (essendo 1.630 millimetri l'altezza media tra 1.600 millimetri e 1.660 millimetri).

Questa osservazione si generalizza (8). Anche se la variabile casuale semplice è continua il campione sarà sempre discreto, raccogliendosi in un'unica classe i valori che cadono in un certo intervallo. Quello che si è detto per una variabile casuale semplice vale anche per le componenti di una variabile casuale multipla di modo che possiamo dire che quando, con un numero un po' grande di prove, si costruisce un campione di una variabile casuale, si riuniscono in un'unica classe quei valori o coppie ordinate di valori ecc. . . . che cadono in un certo intervallo, o coppia ordinata d'intervalli (e quindi rettangolo o intervallo a due dimensioni) ecc. ecc.

E si osservi che se anche la variabile casuale è illimitata, cioè anche se le sue componenti possono assumere valori varianti da $-\infty$ a $+\infty$,

---

(8) Ved. per esempio: G. U. YULE - M. G. KENDALL, *An introduction to the theory of statistics*, Londra, Griffin, 1940, ed ancora: A. L. BOWLEY, *Elements of statistics*, Londra, King, 1937.

gli intervalli rimangono sempre in numero finito, raccogliendo in tal caso artificiosamente negli intervalli estremi, che poi saranno i più poveri, i valori al di sopra o al di sotto di certi limiti.

Con il raggruppamento convenzionale sopra esposto alla variabile casuale di partenza viene associata una variabile casuale discreta detta *variabile casuale divisa in intervalli*, la quale assume come valori i centri degli intervalli con certe probabilità che subito si calcolano mediante la variabile casuale di partenza.

E non è questa la sola nuova variabile casuale che tali generi di considerazioni introducono, ché questo concetto di divisione in k intervalli $C_1, C_2, \ldots, C_k$ porta automaticamente in giuoco la *variabile casuale delle prove ripetute a più alternative* che nasce considerando come evento E_i $(i = 1, 2, \ldots, k)$ il fatto che in un esperimento semplice si sia presentato un valore dell'intervallo C_i.

Alla luce di questa nuova veduta il campione costruito con la nostra esperienza (consistente in N esperimenti semplici) assume il significato di una particolare determinazione della variabile casuale k-pla delle prove ripetute N volte secondo lo schema di Bernoulli e la nostra esperienza diventa un esperimento semplice su questa nuova variabile casuale.

Osservi il lettore che questo non è solo il punto d'incontro tra la variabile casuale semplice e quella multipla, tra la variabile casuale discreta e quella continua, ma è anche il terreno dove le modalità quantitative si unificano alle qualitative essendo la variabile casuale delle prove ripetute del tutto indifferente al fatto che gli intervalli rappresentino classi di qualità o di quantità; ed allora chiaramente comprenderà come questa concezione rappresenti quella che si può dire la chiave di volta di tutta la statistica pura, perché il grande potere di unificazione, che tale concezione possiede, non solo dà ai concetti e ai metodi la massima generalità, ma per di più riconduce gli schemi ed i problemi più disparati ed anche più incerti ad un unico concetto chiaro, solido, maneggevole e profondamente noto: la variabile casuale delle prove ripetute a più alternative secondo lo schema di Bernoulli.

III.

Molte altre cose e importantissime, scendendo in maggiori dettagli, si potrebbero dire su questi argomenti che vorrei chiamar generali, ma, per rispetto di quelle esigenze economiche che debbono presiedere alla distribuzione delle parti di un qualsiasi lavoro, mi fermo qui avendo già raggiunto il mio fine di esporre, per sommi capi, quei concetti che debbono guidare nell'impostazione e nello svolgimento dei vari capitoli della statistica.

Per avvicinarmi all'argomento che in particolar modo è l'oggetto di questo lavoro, cioè lo studio di due indici, mi soffermerò ad esaminare le esigenze metodologiche che dalle precedenti riflessioni, discendono per chi voglia trattare la teoria di un indice assegnato. Beninteso tralascerò tutto quello che rientra nell'ambito della « statistica applicata » e questo non perché l'argomento non sia di per se stesso importante e degno di studio, ma bensì perché mi porterebbe lontano dallo scopo di questa analisi che si limita alla ricerca e alla elencazione dei vari argomenti che bisognerebbe svolgere per dare, in sede di « statistica pura », la trattazione completa di un indice.

Avendo distinto i due concetti di variabile casuale e di campione ed avendo precisato che il primo concetto è il presupposto del secondo, è subito chiaro come la teoria astratta di un indice si debba sviluppare in due parti distine e ordinatamente susseguentesi: « l'indice per la variabile casuale » prima, « l'indice per il campione » poi.

Ma si presenta spontanea un'altra divisione in due parti il cui susseguirsi è pure fissato *a priori:* prima « l'indice nel caso che la variabile casuale associata al fenomeno sia generica » e poi « l'indice nel caso che la variabile casuale associata al fenomenosia di un tipo particolare » (per esempio: un tipo del Pearson); distinguerò i due argomenti dicendo: « l'indice in generale » e « l'indice in particolari distribuzioni ».

Ne segue che la trattazione di un indice, in sede di statistica pura, risulta, fin da una prima analisi superficiale, scissa in quattro parti che si possono, secondo lo specchietto seguente, in due diversi modi ordinare:

I) *L'indice per la variabile casuale:*
  a) *in generale;*
  b) *in particolari distribuzioni.*

I) *L'indice in generale:*
  a) *per la variabile casuale;*
  b) *per il campione.*

II) *L'indice per il campione:*

c) *in generale;*
d) *in particolari distribuzioni.*

II) *L'indice in particolari distribuzioni:*

c) *per la variabile casuale;*
d) *per il campione.*

Mi atterrò nel seguito all'ordinamento di sinistra; inoltre avverto che da questo punto in poi invece di scrivere « variabile casuale » abbrevierò con: v. c.

Nella prima parte (l'indice per la v. c.: in generale), dopo aver definito l'indice il cui valore indicherò con i, se ne dànno le eventuali generalizzazioni, se ne studiano le proprietà e sopratutto se ne determinano le relazioni con altri eventuali parametri.

Nella seconda parte (l'indice per la v. c.: nel caso di particolari distribuzioni) si calcolano anzitutto i valori che assume l'indice per queste particolari v. c. e poi si determinano le relazioni tra l'indice ed altri parametri.

Come si è visto, quel che riguarda « l'indice per la v. c. », si sbriga in poche parole; un po' più complesso si presenterà invece lo studio de « l'indice per il campione ».

Anzitutto occorrerà definire l'indice per il campione, il che presenta subito una difficoltà dato che i due concetti di v. c. e di campione non sono logicamente omogenei figurando nel primo le probabilità e nel secondo le frequenze, il che fa sì che la definizione dell'indice per la v. c. non porti come conseguenza logica quella dell'indice per il campione.

Tuttavia se si tien conto del significato che hanno gli indici in statistica, tale ostacolo si supera sostituendo, nella definizione originaria dell'indice, alle probabilità della v. c. le frequenze del campione, o, in altre parole, considerando a sua volta il campione come un universo (nel qual caso lo chiamerò: campione–universo).

Se si vuole, da un punto di vista logico puro tale sostituzione è del tutto arbitraria, pur essendo consigliata da esigenze d'indole pratica, e credo sia utile tener presente l'arbitrarietà di tale atto per conservare, in alcuni casi particolari, la libertà d'azione di variare *leggermente*, per fini pratici, la definizione stessa; ma su questo argomento ritornerò fra poche righe quando avrò introdotto altri concetti neces-

sari per l'analisi che stiamo svolgendo, per ora mi basta mettere in luce l'opportunità di usare due simboli diversi per indicare l'indice del campione universo e quello del campione; precisamente per il primo indice scriverò i^* e per il secondo $_{,}i$; il che non toglie che in generale porremo: $_{,}i = i^*$.

Osserviamo incidentalmente che non è affatto detto che la nostra definizione ci permetta di ottenere senza fatica il valore $_{,}i$ che assume l'indice su un determinato campione; pertanto bisognerà subito ricercare quegli accorgimenti e quegli artifici che rendono il più possibile spedito e semplice il calcolo di $_{,}i$, almeno per particolari strutture del campione.

Fissiamo ora la nostra attenzione sull'indice i^*, calcolato per il campione-universo, e immaginiamo di far variare il gruppo di N prove che hanno costruito il campione; ovviamente la struttura stessa del campione varia e con esso il valore i^* dell'indice che pertanto descrive una nuova v. c. $\mathfrak{z}^*$ la quale viene così ad imporsi alla nostra attenzione.

Se supponiamo, come molto spesso accade, che questa v. c. $\mathfrak{z}^*$ si possa, con buona approssimazione, riguardare continua, la mèta verso cui fin dall'inizio deve tendere la ricerca resta automaticamente fissata: determinare la funzione di ripartizione di questa v. c.; ché, una volta in possesso di questa funzione, tutto il resto della ricerca può essere trattato in modo relativamente semplice e spedito.

Ma purtroppo non sempre si riesce a soddisfare questo, anche se encomiabile, desiderio; e qualche volta la funzione di ripartizione della v. c. $\mathfrak{z}^*$, per colpa delle eccessive difficoltà che s'incontrano, resta nel limbo delle aspirazioni insoddisfatte. In tal caso facendo di necessità virtù, bisogna limitare i nostri programmi, accontentandoci di venire a sapere solo quello che più interessa; cioè da un lato la media della v. c. $\mathfrak{z}^*$ e dall'altro il grado di attendibilità dell'indice calcolato sul campione. Nei riguardi dell'attendibilità avverto che tutto si riduce al calcolo di un conveniente indice di variabilità. Ma quale tipo di indice, uno « scostamento medio » o una « differenza media » ? La questione è presto risolta perché, siccome vogliamo sapere l'attendibilità dell'indice, cioè conoscere una misura di come i^* si scosta da i, s'impone la scelta di uno « scostamento medio »; e, sia per semplicità di calcoli sia perché, in molti casi, al crescere del numero N delle prove

la v. c. $\mathfrak{z}^*$ tende ad essere normale, conviene servirsi di uno scostamento quadratico, e precisamente di:

$$\sigma^2 = M([\mathfrak{z}^* - i]^2)$$

L'introduzione della v. c. $\mathfrak{z}^*$ permette di precisare una questione poco prima lasciata sospesa.

Se per il nostro indice succede che sia $M(\mathfrak{z}^*) = i$, il precedente scostamento quadratico σ^2 diventa la « varianza » $\mu^2_{\mathfrak{z}^*}$ della v. c. $\mathfrak{z}^*$, indice quest'ultimo assai utile ai fini dello studio della v. c. medesima; inoltre, come si può sperimentalmente constatare, molte formule e relazioni di questo caso particolare si semplificano notevolmente; onde, se con una semplice correzione, si può variare i^* in modo che la media della nuova v. c. sia proprio l'indice i, conviene porre $_{,}i$ eguale a quest'ultimo indice corretto.

Un esempio di questo modo di procedere viene sviluppato nel terzo paragrafo a proposito dell'indice di mutabilità g_U; infatti l'indice di mutabilità g^*_U del campione-universo descrive la v. c. $\mathfrak{G}^*$ la cui media è (9):

$$M(\mathfrak{G}^*) = \frac{N-1}{N} g_U$$

e allora come indice del campione si prende:

$$_{,}g_U = \frac{N}{N-1} g^*_U$$

di modo che l'indice $_{,}g_U$, al variare del gruppo di N prove, descrive la v. c. $\mathfrak{G}$ la cui media vale esattamente: g_U.

Dopo queste precisazioni sulla definizione di $_{,}i$, che può pertanto alcune volte differire da i^*, conviene considerare, anziché la $\mathfrak{z}^*$, la v. c. $\mathfrak{z}$ descritta da $_{,}i$ al variare del gruppo di N prove. E possiamo ancora ripetere che, se non si riesce a calcolare la funzione di ripartizione della v. c. $\mathfrak{z}$ descritta dall'indice del campione, bisogna almeno cercar di calcolare la sua media e il suo scostamento quadratico da i. Ma a questo punto bisognerebbe spiegare come mai si possono calcolare la media e la varianza di una v. c. di cui non si conosce la funzione di ripartizione; non entrerò in tali dettagli tanto più che leggendo questa nota si vedrà chiaramente il procedimento a cui si ricorre,

(9) Tale media è stata già calcolata da C. Gini, loc. cit. (1), pag. 351.

tuttavia mi limiterò ad osservare due cose, e cioè in primo luogo che in questi problemi ha un ruolo importantissimo quella v. c. delle prove ripetute di cui ho già parlato, ed in secondo luogo che, in generale, la funzione di ripartizione della v. c. non è del tutto incognita ma solo espressa mediante integrali che non si sanno calcolare.

Molte volte accanto al nostro indice i si considerano sopra il campione altri parametri che con l'indice stesso vengono combinati allo scopo di ottenere valori segnalatici di qualche particolare aspetto del fenomeno.

Questo porta a studiare la v. c. multipla descritta dal nostro indice e dagli altri parametri al variare del campione; e, al solito, se non si riuscirà a determinare la funzione di densità di questa nuova v. c. multipla, dovremo limitarci a calcolare i relativi coefficienti di correlazione.

Quando però si parla di scostamento dell'indice calcolato sul campione dal suo vero valore, bisogna tener presente un fatto notevole che sempre si presenta quando il campione è stato diviso in intervalli; in tale ipotesi il « vero valore » di cui si discorre è quello che assume l'indice per la v. c. divisa in intervalli, valore che potrebbe esser diverso da quello che invece assume per la v. c. di partenza.

Che relazioni passano tra questi due diversi valori dell'indice?

Ecco un altro interessante campo d'indagini il cui scopo è soprattutto quello di determinare, almeno sotto alcune ipotesi generali, delle semplici correzioni da applicare al nostro indice per attenuar l'influenza della divisione in intervalli.

Passiamo ora a discutere la quarta ed ultima parte della trattazione di un indice in sede di statistica pura (l'indice per il campione: in particolari distribuzioni).

In questa parte della trattazione naturalmente ricompaiono molte delle questioni già accennate per la parte precedente; ma ora si ha il vantaggio che, non trovandoci più di fronte al caso generale, si possono sfruttare, ai fini della ricerca, tutte le particolarità delle v. c. in esame. Naturalmente se vogliamo avvantaggiarci di questa situazione non dobbiamo considerare campioni divisi in intervalli ma tener separate le N determinazioni che formano il campione.

L'interesse di questa ricerca non risiede solo nella precisazione di ciò che accade quando le v. c. di partenza sono di un tipo particolare

notevole ed interessante, ma anche su un altro fatto su cui sarà bene soffermarsi un po'.

Quando si studia l'indice per un campione proveniente da una generica v. c., molte volte, per rendere i risultati conclusivi, bisogna introdurre l'ipotesi che il campione sia stato ottenuto con un gran numero di prove, di modo che si può dire che la statistica è molte volte disarmata di fronte a quei campioni, che chiamerò *campioni poveri*, ottenuti con un esiguo numero di prove. Ma se questo accade quando nulla si sa o si vuol supporre sulla v. c. di partenza, non è più così se s'introduce l'ipotesi che essa sia di un determinato tipo, e se ne sono avuti esempi luminosissimi con i parametri «t» di STUDENT e «z» di FISHER che permettono di studiare i campioni poveri quando la v. c. di partenza è normale (10).

Riepilogando, dalla discussione precedente discende che i vari argomenti da svolgere, nella trattazione di un indice in sede di statistica pura, possono essere elencati nei dodici punti del sommario seguente:

I) L'INDICE PER LA V. C.

a) *in generale:*

1° definizione dell'indice e sue generalizzazioni;

2° proprietà dell'indice e sue relazioni con altri parametri;

b) *in particolari distribuzioni:*

3° valori che assume l'indice per particolari v. c.;

4° relazioni, per queste particolari v. c., tra l'indice ed altri parametri;

II) L'INDICE PER IL CAMPIONE.

c) *in generale:*

5° definizione dell'indice per il campione;

6° metodi pratici per il calcolo dell'indice;

7° studio della v. c. descritta dall'indice al variare del campione;

8° studio della v. c. multipla descritta dal nostro indice e da altri eventuali parametri al variare del campione;

9° studio dell'influenza della divisione in celle sul valore dell'indice;

(10) Ved. per esempio G. U. YULE - M. G. KENDALL, op. cit., (8).

d) *in particolari distribuzioni:*

10° studio della v. c. descritta dall'indice al variare del campione nel caso che la v. c. di partenza appartenga a particolari tipi;

11° studio della v. c. multipla descritta dal nostro indice e da altri eventuali parametri al variare del campione, nel caso che la v. c. di partenza appartenga a particolari tipi;

12° contributi alla teoria dei campioni poveri.

Quando si passa dai fenomeni a modalità quantitative a quelli a modalità qualitative, quel che sin qui si è detto non sembrerebbe più applicabile perché fin dall'inizio viene a mancare la v. c. associata al fenomeno. Tuttavia questa difficoltà è più apparente che sostanziale perché la v. c. sta ad esprimere la legge teorica che governa il fenomeno, legge la cui esistenza si può ammettere anche per i fenomeni qualitativi; e in realtà tutto si rimette a posto associando ad un fenomeno le cui modalità sono qualitative una MUTABILE CASUALE il cui concetto definisco nel n. 8 della presente nota.

Si tratta, è vero, di un puro formalismo, ma in tal modo sostituendo alla v. c. associata al fenomeno la m. c. (abbreviazione di « mutabile casuale ») tutto quello che si è detto per i fenomeni quantitativi si può adattare, con lievi ed evidenti modifiche, ai fenomeni qualitativi, e pertanto il sommario già determinato per lo studio degli indici che intervengono nella teoria dei fenomeni quantitativi serve pure, *mutatis mutandis*, come base per svolgere la teoria di un indice che si presenta nello studio dei fenomeni a modalità qualitative.

IV.

In questa nota illustro con pratici esempi alcuni dei punti sopra esposti applicandoli, per quel che riguarda i fenomeni quantitativi, alla « media geometrica » (§ II) e, per quel che riguarda i fenomeni qualitativi, all'« indice di mutabilità » introdotto da C. GINI (§ III).

Dico subito che tocco solo alcuni (11) dei punti programmatici che ho raccolto nel sommario precedente soffermandomi in particolare

(11) L'applicazione completa del programma sopra esposto è stata fatta, in occasione della sua tesi di laurea, da B. MICHETTI il quale se ne è servito per ripresentare in modo organico la teoria della « differenza lineare media ».

a determinare la media e lo scostamento della v. c. descritta dai due indici al variare del campione.

Per tale scopo ho ristabilito (§ I) alcune formule già praticamente note sui momenti della v. c. delle prove ripetute a parecchie alternative e sui successivi momenti degli scarti delle v. c. descritte dai *momenti* calcolati su un campione al variare del campione stesso, soffermandomi inoltre a valutare l'ordine di grandezza di tali successivi momenti.

V.

Chiuderò questa introduzione proponendo un criterio per la classificazione degli indici.

In statistica è invalso l'uso di caratterizzare con appositi termini alcuni aspetti generali dei valori segnaletici in uso; così, per esempio, si parla di *coefficiente* di correlazione, di *rapporto* di concentrazione ecc. dove le parole « coefficiente », « rapporto », . . . stanno ad indicare particolari proprietà del valore segnaletico in questione.

Ma sfogliando i vari trattati di statistica si nota subito come vi sia un notevole disaccordo sul significato da darsi a queste parole; disaccordo derivante dal fatto che non viene mai, in modo esplicito e chiaro, enunciato il naturale criterio in base al quale si fa la classificazione, la quale invece, tenendo presente l'interpretazione geometrica di una v. c. ([12]) resta automaticamente appoggiata alla maniera di comportarsi dei valori segnaletici al variare del sistema di riferimento.

D'altra parte il possesso di un mezzo così espressivo che permette di ricordare, mediante una semplice nomenclatura, alcune notevoli proprietà dei valori segnaletici in discussione, è tanto importante ed apporta una tale economia di esposizione che credo valga la pena di discutere brevemente, come farò, tale questione prima ancora di cominciar a trattare l'argomento che interessa la presente nota.

Per ragionare in un caso comprensivo e nello stesso tempo semplice farò tale classificazione nel caso di una v. c. doppia (cioè distendibile su un *piano* cartesiano ad assi ortogonali); ma non proporrò una apposita terminologia per contraddistinguere i vari tipi di valori segnaletici che verrano in tal modo messi in luce, lasciando ad altri più capaci il compito di trovare un nome espressivo per ognuno dei casi che si presentano.

---

(12) Ved. per esempio: G. CASTELNUOVO, op. cit. ([1]) ed anche: G. POMPILJ: *Sulla regressione*, « Rend. di Mat. pura ed applicata », 1946.

Come è noto gli elementi caratteristici di un sistema cartesiano ad assi ortogonali sono tre: l'origine, le unità di misura e l'orientamento degli assi; di modo che si possono fare, su un tale sistema cartesiano, le seguenti otto operazioni:

I – lasciare *tutto* immutato;
II – cambiare solo *le unità di misura*;
III – cambiare solo *l'origine*;
IV – cambiare solo *la direzione degli assi*;
V – cambiare solo *le unità di misura e l'origine*;
VI – cambiare solo *le unità di misura e la direzione degli assi*;
VII – cambiare solo *l'origine e la direzione degli assi*;
VIII – cambiare *tutto*.

Ciò posto, volendo scendere a più dettagliati particolari, indicherò provvisoriamente con le prime otto lettere maiuscole dell'alfabeto latino quegli indici che rimangono invariati per una delle otto operazioni sopra elencate; per indicare gli eventuali sottocasi userò lettere con apici:

1° A: un qualunque valore segnaletico che, pertanto, cambia generalmente valore quando si varia il sistema di riferimento. In particolare si chiamano *omogenei di grado* $[r, s]$ quegli A che, quando si variano le sole unità di misura prendendo rispettivamente, per nuove unità dei due assi, $1/h$ e $1/k$ delle vecchie, non subiscono altra alterazione che quella di essere moltiplicati per il monomio: $h^r k^s$.

2° B: valore segnaletico che non cambia quando si variano le sole unità di misura; cioè A omogeneo di grado [0,0]. Accanto ai B si possono considerare i B′ che non cambiano valore solamente quando si variano in egual proporzione ($h = k$) le unità di misura dei due assi.

3° C: valore segnaletico che non cambia quando si varia solo l'origine. In particolare si possono considerare i C *omogenei di grado* $[r, s]$.

4° D: valore segnaletico che non cambia quando si varia solo l'orientamento degli assi. In particolare si possono considerare i D *omogeni di grado* $[r, s]$.

5° E: valore segnaletico che non cambia quando si variano solamente le unità di misura e l'origine; cioè C omogeneo di grado [0,0]. Accanto agli E si possono considerare i C omogenei di grado $[r, -r]$ che indicherò con E′.

6° F: valore segnaletico che non cambia quando si variano solamente le unità di misura e l'orientamento degli assi; cioè D omogeneo

di grado [0,0]. Accanto agli F si possono considerare gli F′, cioè D omogenei di grado $[r, -r]$.

7° G: valore segnaletico che non cambia quando si varino solamente l'origine e l'orientamento degli assi. In particolare si possono considerare i G *omogeni di grado* $[r, s]$.

8° H: valore segnaletico che non cambia comunque si varino l'origine, le unità di misura e l'orientamento degli assi; cioè G omogeneo di grado [0,0]. Accanto agli H si possono considerare gli H′ cioè i G omogenei di grado $[r, -r]$.

È chiaro che per le v. c. semplici interessano solo i seguenti quattro tipi: A, B, C, E.

Due parole infine sulle convenzioni di scrittura. Indicherò le v. c. con caratteri rondi maiuscoli e con le stesse lettere corsive minuscole i valori assunti dalle v. c. medesime; tale convenzione costringe ad indicare gli indici sul campione con lettere (latine) corsive minuscole, ed allora, per uniformità di scrittura, conviene usare gli stessi caratteri corsivi minuscoli per indicare gli analoghi indici sulla v. c. o m. c.

Le m. c. a loro volta saranno indicate con caratteri mozzi larghi maiuscoli di modo che, per esempio, sulla m. c. $\mathbf{U}$ si ha l'indice di mutabilità $g_{\mathbf{U}}$, su un campione della stessa m. c. si ha l'indice ${}_{\mathscr{S}}g_{\mathbf{U}}$ che, al variare del campione, descrive la v. c. $\mathscr{G}$.

§ I. — PREMESSE

1. – La v. c. delle prove ripetute a parecchie alternative, secondo lo schema di Bernoulli. In questo primo paragrafo, anche allo scopo di introdurre i simboli, raccoglierò varie formule già praticamente note, dimostrandone alcune che serviranno nel seguito.

In un esperimento si presenti necessariamente uno dei k eventi incompatibili $E_1, E_2, \ldots, E_k$ con probabilità costanti $p_1, p_2, \ldots, p_k$ $\left(\sum_1^k p_i = 1\right)$; ed inoltre il risultato di ogni singolo esperimento non abbia alcuna influenza nel risultato degli altri (schema di Bernoulli).

Eseguito, in queste condizioni, un gruppo di N esperimenti, l'evento E_1 si sia presentato a_1 volte, l'evento E_2 a_2 volte, . . . , l'evento E_k a_k volte $\left(\sum_1^k a_i = N\right)$.

Al variare del gruppo delle N prove la k —pla ordinata $(a_1, a_2, \ldots, a_k)$ varia e descrive una v. c. $(\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_k)$ che assume la determinazione $(a_1, a_2, \ldots, a_k)$ con probabilità [13]:

$$P = \frac{N!}{a_1!\, a_2! \cdots a_k!}\, p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k}$$

Ogni singolo valore a_i a sua volta descrive una v. c. semplice $\mathfrak{a}_i$ di media Np_i.

Indichiamo con $\mathfrak{N}_i$ la v. c. scarto (dalla media) della v. c. $\mathfrak{a}_i$; cioè la v. c. che assume il valore:

$$n_i = a_i - Np_i$$

quando la v. c. $\mathfrak{a}_i$ prende la determinazione a_i.

Come è noto [14], posto $q_i = 1 - p_i$, si hanno le relazioni:

$$[1] \quad \begin{cases} M(\mathfrak{N}_i) = 0 \;;\; M(\mathfrak{N}_i^2) = Np_iq_i \;;\; M(\mathfrak{N}_i^3) = Np_iq_i(q_i - p_i) \;; \\ M(\mathfrak{N}_i^4) = Np_iq_i[1 + 3(N-2)p_iq_i]. \end{cases}$$

Ma oltre che i precedenti valori medi occorre calcolare anche tutti i momenti misti, fino a quelli del quarto ordine, della v. c. k-pla $(\mathfrak{N}_1, \mathfrak{N}_2, \cdots, \mathfrak{N}_k)$; a tale scopo occorre premettere un

Lemma: *La regressione della v. c. $\mathfrak{N}_i$ rispetto alla v. c. r-pla $(\mathfrak{N}_{j_1}, \mathfrak{N}_{j_2}, \cdots, \mathfrak{N}_{j_r})$ è lineare.*

Supponiamo che la v. c. r-pla $(\mathfrak{N}_{j_1}, \mathfrak{N}_{j_2}, \cdots, \mathfrak{N}_{j_r})$ abbia assunto la determinazione $(n_{j_1}, n_{j_2}, \cdots, n_{j_r})$. Ciò significa che gli r eventi $E_{j_1}, E_{j_2}, \cdots, E_{j_r}$ si sono complessivamente presentati in:

$$A = N(p_{j_1} + \cdots + p_{j_r}) + n_{j_1} + \cdots + n_{j_r}$$

delle N prove, mentre nei rimanenti:

$$B = N - A = N(1 - p_{j_1} - \cdots - p_{j_r}) - n_{j_1} - \cdots - n_{j_r}$$

esperimenti si è presentato uno degli altri $k - r$ eventi.

---

(13) Ved. per esempio G. Castelnuovo, op. cit. (2), vol. I, pag. 50.
(14) Ved. per esempio G. Castelnuovo, op. cit., (2), vol. I, pagg. 51 e 53.

Per le rimanenti B prove la probabilità che si presenti l'evento E_i è ovviamente:

$$p'_i = \frac{p_i}{1 - p_{j_1} - \cdots - p_{j_r}}$$

quindi, nelle nostre condizioni, la media della v. c. $\mathfrak{N}_i$ è data da:

$$m_i(n_{j_1} \cdots n_{j_r}) = B\, p'_j - N p_i =$$

$$= \left[N(1 - p \quad - \cdots - p_{j_r}) - n_{j_1} - \cdots - n_{j_r}\right] \frac{p_i}{1 - p_{j_1} - \cdots - p_{j_r}} - N p_i =$$

$$= - \frac{p_i}{1 - p_{j_1} - \cdots - p_{j_r}} (n_{j_1} + \cdots + n_{j_r}) .$$

Si tratta pertanto di regressione lineare c. v. d.

Dalla prima proprietà caratteristica delle funzioni di regressione [15] si ottiene infine l'identità tra momenti:

$$[2] \quad \begin{cases} M(\mathfrak{N}_i \mathfrak{N}_{j_1}^{t_1} \mathfrak{N}_{j_2}^{t_2} \cdots \mathfrak{N}_{j_r}^{t_r}) = \\ = - \dfrac{p_i}{1 - p_{j_1} - \cdots - p_{j_r}} \Big[M(\mathfrak{N}_{j_1}^{t_1+1} \mathfrak{N}_{j_2}^{t_2} \cdots \mathfrak{N}_{j_r}^{t_r}) + \\ \qquad + M(\mathfrak{N}_{j_1}^{t_1} \mathfrak{N}_{j_2}^{t_2+1} \cdots \mathfrak{N}_{j_r}^{t_r}) + \cdots + M(\mathfrak{N}_{j_1}^{t_1} \mathfrak{N}_{j_2}^{t_2} \cdots \mathfrak{N}_{j_r}^{t_r+1}) \Big] \end{cases}$$

da cui, particolarizzando i vari indici, si ottengono subito le seguenti formule:

$$[3] \quad \begin{cases} M(\mathfrak{N}_i \mathfrak{N}_j) = - \dfrac{p_i}{1 - p_j} M(\mathfrak{N}_j^2) = - N p_i p_j \\ M(\mathfrak{N}_i \mathfrak{N}_j^2) = - \dfrac{p_i}{1 - p_j} M(\mathfrak{N}_j^3) = - N p_i p_j (q_j - p_j) \\ M(\mathfrak{N}_i \mathfrak{N}_j \mathfrak{N}_l) = - \dfrac{p_i}{1 - p_j - p_l} [M(\mathfrak{N}_j^2 \mathfrak{N}_l) + M(\mathfrak{N}_j \mathfrak{N}_l^2)] = 2 N p_i p_j p_l \\ M(\mathfrak{N}_i \mathfrak{N}_j^3) = - \dfrac{p_i}{1 - p_j} M(\mathfrak{N}_j^4) = - N p_i p_j [1 + 3(N - 2) p_j q_j] \end{cases}$$

Per completare il calcolo di tutti i momenti del quarto ordine bisogna ancora determinare i seguenti valori medi:

$$M(\mathfrak{N}_i^2 \mathfrak{N}_j^2) \quad , \quad M(\mathfrak{N}_i \mathfrak{N}_j^2 \mathfrak{N}_l) \quad , \quad M(\mathfrak{N}_i \mathfrak{N}_j \mathfrak{N}_l \mathfrak{N}_r) .$$

---

(15) Ve. per esempio G. POMPILJ, loc. cit. [12].

A tale scopo consideriamo l'evento E consistente nel fatto che si presenti o l'evento E_i o l'evento E_j; la probabilità dell'evento E è ovviamente $p_i + p_j$ mentre la probabilità che non si presenti è: $1 - p_i - p_j$; indichiamo poi con $\mathfrak{N}$ lo scarto della v. c. descritta dal numero delle volte che si presenta l'evento E nelle N prove.

L'ultima delle [1] ci fornisce il seguente valore per il quarto momento della v. c. $\mathfrak{N}$:

$$[4] \quad M(\mathfrak{N}^4) = N(p_i + p_j)(1 - p_i - p_j)[1 + 3(N-2)(p_i + p_j)(1 - p_i - p_j)$$

D'altra parte sussiste l'identità:

$$\mathfrak{N} = \mathfrak{N}_i + \mathfrak{N}_j$$

e quindi:

$$M(\mathfrak{N}^4) = M([\mathfrak{N}_i + \mathfrak{N}_j]^4) = M(\mathfrak{N}_i^4) + 4M(\mathfrak{N}_i^3 \mathfrak{N}_j) + 6M(\mathfrak{N}_i^2 \mathfrak{N}_j^2) + \\ + 4M(\mathfrak{N}_i \mathfrak{N}_j^3) + M(\mathfrak{N}_j^4).$$

Eguagliando il secondo membro di quest'ultima formula a quello della [4] e tenendo conto delle identità di [1] e [3] si ha:

$$[5] \qquad M(\mathfrak{N}_i^2 \mathfrak{N}_j^2) = N p_i p_j [1 + (N-2)(q_i q_j + 2 p_i p_j)]$$

Tenuto conto sia della [5] sia dell'ultima delle [3] e particolarizzando la [2] si ha poi:

$$M(\mathfrak{N}_i \mathfrak{N}_j^2 \mathfrak{N}_i) = -\frac{p_i}{1 - p_j - p_i}[M(\mathfrak{N}_j^2 \mathfrak{N}_i) + M(\mathfrak{N}_j^2 \mathfrak{N}_j^2)] = \\ = -N(N-2) p_i p_j p_i (1 - 3 p_j)$$

$$M(\mathfrak{N}_i \mathfrak{N}_j \mathfrak{N}_i \mathfrak{N}_r) = -\frac{p_i}{1 - p_j - p_i - p_r}[M(\mathfrak{N}_j^2 \mathfrak{N}_i \mathfrak{N}_r) + M(\mathfrak{N}_j \mathfrak{N}_i^2 \mathfrak{N}_r) + \\ + M(\mathfrak{N}_j \mathfrak{N}_i \mathfrak{N}_r^2)] = 3N(N-2) p_i p_j p_i p_r$$

Per comodità del lettore riunisco in un'unica tabella le espressioni dei vari momenti fin qui incontrati; espressioni queste che, in tutti gli sviluppi che seguono, avranno una decisiva importanza.

$$
(6)\quad \begin{cases}
M(\mathfrak{A}_i) = 0; \\
M(\mathfrak{A}_i^2) = N p_i q_i \quad , \quad M(\mathfrak{A}_i \mathfrak{A}_j) = -N p_i p_j; \\
M(\mathfrak{A}_i^3) = N p_i q_i (q_i - p_i) \quad , \quad M(\mathfrak{A}_i \mathfrak{A}_j^2) = -N p_i p_j (q_j - p_j), \\
\qquad M(\mathfrak{A}_i \mathfrak{A}_j \mathfrak{A}_t) = 2 N p_i p_j p_t; \\
M(\mathfrak{A}_i^4) = N p_i q_i [1 + 3(N-2) p_i q_i], \; M(\mathfrak{A}_i \mathfrak{A}_j^3) = -N p_i p_j [1 + 3(N-2) p_j q_j], \\
\qquad M(\mathfrak{A}_i^2 \mathfrak{A}_j^2) = N p_i p_j [1 + (N-2)(q_i q_j + 2 p_i p_j)], \\
\qquad M(\mathfrak{A}_i \mathfrak{A}_j^2 \mathfrak{A}_t) = -N(N-2) p_i p_j p_t (1 - 3 p_j), \\
\qquad M(\mathfrak{A}_i \mathfrak{A}_j \mathfrak{A}_t \mathfrak{A}_r) = 3 N (N-2) p_i p_j p_t p_r.
\end{cases}
$$

2. – Le v. c. « momento ». Sia $\mathfrak{X}$ una v. c. semplice discreta (o divisa in intervalli) che assuma i valori reali (e finiti) $x_1, \ldots, x_k$ con probabilità $p_1, \ldots, p_k$. Eseguiamo sulla v. c. $\mathfrak{X}$ un certo numero N di prove secondo lo schema di Bernoulli; si ottengono in tal modo N valori di cui a_1 coincidono con x_1, a_2 con $x_2, \ldots, a_k$ con x_k. Questi N valori formano quello che si chiama un « campione » della v. c. $\mathfrak{X}$ mentre la k-pla ordinata $(a_1, a_2, \ldots, a_k)$ rappresenta una determinazione della v. c. delle prove ripetute già studiata nel numero precedente.

Per definire i momenti sul campione, in base alle convenzioni esposte nell'introduzione, dobbiamo anzitutto considerare il campione come « universo » nel qual caso lo chiamerò campione-universo.

Indicando con m_r il momento r-mo della v. c. $\mathfrak{X}$ e con m_r^* il momento r-mo del campione-universo si hanno le definizioni:

$$m_r = \sum_1^k {}_i\, x_i^r p_i$$

$$m_r^* = \frac{1}{N} \sum_1^k {}_i\, x_i^r a_i$$

Al variare del campione il momento m_r^* del campione-universo varia e descrive la v. c.:

$$\mathfrak{M}_r^* = \frac{1}{N} \sum_1^k {}_i\, x_i^r \mathfrak{A}_i = m_r + \frac{1}{N} \sum_1^k {}_i\, x_i^r \mathfrak{A}_i$$

e, per le [6], si ha subito:

$$M(\mathfrak{M}_r^*) = m_r + \frac{1}{N} \sum_i x_i^r M(\mathfrak{A}_i) = m_r.$$

Si è così dimostrato che la media della v. c. $\mathfrak{M}_r^*$ è proprio m_r; di modo che possiamo definire il momento r-mo ${}_r m_r$ del campione eguagliandolo all'analogo momento del campione-universo:

$$ {}_r m_r = m_r^* = \frac{1}{N} \sum_i x_i^r a_i $$

e quindi definire la « v. c. $\mathfrak{M}_r$ *momento r-mo del campione* » eguagliandola alla v. c. $\mathfrak{M}_r^*$; cioè

$$ \mathfrak{M}_r = \mathfrak{M}_r^* = m_r + \frac{1}{N} \sum_i x_i^r \mathfrak{A}_i $$

Indicherò con $\mathfrak{s}_r$ la v. c. scarto della v. c. $\mathfrak{M}_r$, cioè:

$$ \mathfrak{s}_r = \frac{1}{N} \sum_i x_i^r \mathfrak{A}_i . $$

Interessa per il seguito calcolare i momenti delle v. c. $\mathfrak{s}_r$, e precisamente calcolerò i momenti, fino a quelli del quarto ordine inclusi, della v. c. quadrupla $(\mathfrak{s}_r, \mathfrak{s}_u, \mathfrak{s}_s, \mathfrak{s}_t)$. I valori di alcuni di questi momenti sono già noti [16] e per di più sono stati calcolati per una via che sostanzialmente coincide con quella che seguirò in questo numero, tuttavia rifarò tutti i calcoli allo scopo di mostrare quale semplicità si raggiunge impiegando sistematicamente e nel modo più esplicito i momenti delle v. c. $\mathfrak{A}_i$ preventivamente calcolati.

Bisogna anzitutto determinare i momenti misti fino al quarto ordine formati dalle v. c. scarto di tipo $\mathfrak{s}_r$ ed $\mathfrak{A}_i$:

$$ \mathrm{M}(\mathfrak{s}_r \mathfrak{A}_j) = \frac{1}{N} \left[\sum_i^{i \neq j} x_i^r \mathrm{M}(\mathfrak{A}_i \mathfrak{A}_j) + x_j^r \mathrm{M}(\mathfrak{A}_j^2) \right] = p_j (x_j^r - m_r) , $$

$$ \begin{aligned} \mathrm{M}(\mathfrak{s}_r \mathfrak{A}_j \mathfrak{A}_i) &= \frac{1}{N} \left[\sum_i^{i \neq j, i} x_i^r \mathrm{M}(\mathfrak{A}_i \mathfrak{A}_j \mathfrak{A}_i) + x_j^r \mathrm{M}(\mathfrak{A}_j^2 \mathfrak{A}_i) + x_i^r \mathrm{M}(\mathfrak{A}_j \mathfrak{A}_i^2) \right] = \\ &= p_j p_i (2 m_r - x_j^r - x_i^r) , \end{aligned} $$

$$ \mathrm{M}(\mathfrak{s}_r \mathfrak{A}_j^2) = \frac{1}{N} \left[\sum_i^{i \neq j} x_i^r \mathrm{M}(\mathfrak{A}_i \mathfrak{A}_j^2) + x_j^r \mathrm{M}(\mathfrak{A}_j^3) \right] = p_j (q_j - p_j)(x_j^r - m_r) , $$

$$ \begin{aligned} \mathrm{M}(\mathfrak{s}_r \mathfrak{s}_u \mathfrak{A}_j) &= \frac{1}{N} \left[\sum_i^{i \neq j} x_i^r \mathrm{M}(\mathfrak{s}_u \mathfrak{A}_i \mathfrak{A}_j) + x_j^r \mathrm{M}(\mathfrak{s}_u \mathfrak{A}_j^2) \right] = \\ &= \frac{p_j}{N} \left[2 m_r m_u - m_{r+u} + x_j^{r+u} - m_u x_j^r - m_r x_j^u \right] , \end{aligned} $$

$$ \mathrm{M}(\mathfrak{s}_r^2 \mathfrak{A}_j) = \frac{p_j}{N} \left[x_j^{2r} - m_{2r} - 2 m_r (x_j^r - m_r) \right] , $$

(16) Ved. per esempio: G. U. Yule-M. G. Kendall, op. cit. (9).

$$\mathrm{M}(\delta_r \mathfrak{A}_j \mathfrak{A}_t \mathfrak{A}_s) = \frac{1}{\mathrm{N}}\Big[\sum_i^{i+j,t,s} x_i^r \mathrm{M}(\mathfrak{A}_i \mathfrak{A}_j \mathfrak{A}_t \mathfrak{A}_s) + x_j^r \mathrm{M}(\mathfrak{A}_j^2 \mathfrak{A}_t \mathfrak{A}_s) +$$

$$+ x_t^r \mathrm{M}(\mathfrak{A}_j \mathfrak{A}_t^2 \mathfrak{A}_s) + x_s^r \mathrm{M}(\mathfrak{A}_j \mathfrak{A}_t \mathfrak{A}_s^2)\Big] =$$

$$= (\mathrm{N}-2)\, p_j p_t p_s (3 m_r - x_j^r - x_t^r - x_s^r),$$

$$\mathrm{M}(\delta_r \mathfrak{A}_j \mathfrak{A}_t^2) = \frac{1}{\mathrm{N}}\Big[\sum_i^{i+j,t} x_i^r \mathrm{M}(\mathfrak{A}_i \mathfrak{A}_j \mathfrak{A}_t^2) + x_j^r \mathrm{M}(\mathfrak{A}_j^2 \mathfrak{A}_t^2) + x_t^r \mathrm{M}(\mathfrak{A}_j \mathfrak{A}_t^3)\Big] =$$

$$= p_j p_t [(\mathrm{N}-2)\{(1-p_t) x_j^r - 2 p_t x_t^r - (1-3p_t) m_r\} + x_j^r - x_t^r]$$

$$\mathrm{M}(\delta_r \mathfrak{A}_j^3) = \frac{1}{\mathrm{N}}\Big[\sum_i^{i+j} x_i^r \mathrm{M}(\mathfrak{A}_i \mathfrak{A}_j^3) + x_j^r \mathrm{M}(\mathfrak{A}_j^4)\Big] = p_j [1 + 3(\mathrm{N}-2) p_j q_j][x_j^r - m_r],$$

$$\mathrm{M}(\delta_r \delta_u \mathfrak{A}_j \mathfrak{A}_t) = \frac{1}{\mathrm{N}}\Big[\sum_i^{i+j,t} x_i^r \mathrm{M}(\delta_u \mathfrak{A}_i \mathfrak{A}_j \mathfrak{A}_t) + x_j^r \mathrm{M}(\delta_u \mathfrak{A}_j^2 \mathfrak{A}_t) + x_t^r \mathrm{M}(\delta_u \mathfrak{A}_j \mathfrak{A}_t^2)\Big] =$$

$$= \frac{p_t p_j}{\mathrm{N}}[(\mathrm{N}-2)\{3 m_r m_u - m_{r+u} - m_u (x_j^r + x_t^r) - m_r (x_j^u + x_t^u)\} +$$

$$+ (\mathrm{N}-1)\{x_j^u x_t^r + x_j^r x_t^u\} - x_t^{r+u} - x_j^{r+u}],$$

$$\mathrm{M}(\delta_r^2 \mathfrak{A}_j \mathfrak{A}_t) = \frac{p_j p_t}{\mathrm{N}}[(\mathrm{N}-2)\{3 m_r^2 - m_{2r} - 2 m_r (x_j^r + x_t^r)\} + 2(\mathrm{N}-1) x_j^r x_t^r - x_j^{2r} - x_t^{2r}],$$

$$\mathrm{M}(\delta_r \delta_u \mathfrak{A}_j^2) = \frac{1}{\mathrm{N}}\Big[\sum_i^{i+j} x_i^r \mathrm{M}(\delta_u \mathfrak{A}_i \mathfrak{A}_j^2) + x_j^r \mathrm{M}(\delta_u \mathfrak{A}_j^3)\Big] =$$

$$= \frac{p_j}{\mathrm{N}}[\{1 + (\mathrm{N}-2)(1-p_j)\} m_{r+u} + \{1 + 2(\mathrm{N}-2) p_j\}$$

$$(x_j^{r+u} - x_j^u m_r - x_j^r m_u) - (\mathrm{N}-2)(1-3p_j) m_r m_u],$$

$$\mathrm{M}(\delta_r^2 \mathfrak{A}_j^2) = \frac{p_j}{\mathrm{N}}[\{1 + (\mathrm{N}-2)(1-p_j)\} m_{2r} + \{1 + 2(\mathrm{N}-2) p_j\}(x_j^{2r} - 2 x_j^r m_r) +$$

$$- (\mathrm{N}-2)(1-3p_j) m_r^2],$$

$$\mathrm{M}(\delta_r \delta_u \delta_z \mathfrak{A}_j) = \frac{1}{\mathrm{N}}\Big[\sum_i^{i+j} x_i^r \mathrm{M}(\delta_u \delta_z \mathfrak{A}_i \mathfrak{A}_j) + x_j^r \mathrm{M}(\delta_u \delta_z \mathfrak{A}_j^2)\Big] =$$

$$= \frac{p_j}{\mathrm{N}^2}[(\mathrm{N}-2)(3 m_r m_u m_z - m_{r+u} m_z - m_{r+z} m_u - m_{u+z} m_r +$$

$$- m_r m_u x_j^z - m_r m_z x_j^u - m_u m_z x_j^r) + (\mathrm{N}-1)(m_{r+u} x_j^z + m_{r+z} x_j^u +$$

$$+ m_{u+z} x_j^r) - m_{r+u+z} + x_j^{r+u+z} - x_j^{r+u} m_z - x_j^{r+z} m_u - x_j^{u+z} m_r],$$

$$\mathrm{M}(\delta_r^2 \delta_u \mathfrak{A}_j) = \frac{p_j}{\mathrm{N}^2}[(\mathrm{N}-2)(3 m_r^2 m_u - m_{2r} m_u - 2 m_{r+u} m_r - m_r^2 x_j^u +$$

$$- 2 m_r m_u x_j^r) + (\mathrm{N}-1)(m_{2r} x_j^u + 2 m_{r+u} x_j^r) + x_j^{2r+u} +$$

$$- x^{2r} m_u - 2 x_j^{r+u} m_r - m_{2r+u}],$$

$$\mathrm{M}(\delta_r^3 \mathfrak{A}_j) = \frac{p_j}{\mathrm{N}^2}[3(\mathrm{N}-2)(m_r^3 - m_{2r} m_r - m_r^2 x_j^r) + 3(\mathrm{N}-1) m_{2r} x_j^r +$$

$$+ x_j^{3r} - 3 m_r x_j^{2r} - m_{3r}].$$

Possiamo ora calcolare i momenti, fino al quarto ordine, della v. c. quadrupla $(\delta_r, \delta_u, \delta_z, \delta_t)$:

$$M(\delta_r) = \frac{1}{N}\sum_i x_i^r M(\mathfrak{A}_i) = 0 .$$

$$M(\delta_r \delta_u) = \frac{1}{N}\sum_i x_i^r M(\delta_u \mathfrak{A}_i) = \frac{1}{N}\left[m_{r+u} - m_r m_u\right],$$

$$M(\delta_r^2) = \frac{1}{N}\left[m_{2r} - m_r^2\right],$$

$$M(\delta_r \delta_u \delta_z) = \frac{1}{N}\sum_i x_i^r M(\delta_u \delta_z \mathfrak{A}_i) = \frac{1}{N^2}\left[m_{r+u+z} - m_{r+u} m_z - m_{r+z} m_u + \right.$$
$$\left. - m_{u+z} m_r + 2 m_r m_u m_z\right],$$

$$M(\delta_r^2 \delta_u) = \frac{1}{N^2}\left[m_{2r+u} - 2 m_{r+u} m_r - m_{2r} m_u + 2 m_r^2 m_u\right],$$

$$M(\delta_r^3) = \frac{1}{N^2}\left[m_{3r} - 3 m_{2r} m_r + 2 m_r^3\right],$$

$$M(\delta_r \delta_u \delta_z \delta_t) = \frac{1}{N}\sum_i x_i^r M(\delta_u \delta_z \delta_t \mathfrak{A}_i) = \frac{1}{N^2}\left[3 m_r m_u m_z m_t + m_{r+u} m_{z+t} + \right.$$
$$+ m_{r+z} m_{u+t} + m_{r+t} m_{z+u} - m_{r+u} m_z m_t - m_{r+z} m_u m_t - m_{r+t} m_u m_z +$$
$$\left. - m_{u+z} m_r m_t - m_{u+t} m_z m_r - m_{z+t} m_r m_u\right] + \frac{1}{N^3}\left[m_{r+u+z+t} + \right.$$
$$+ 2 m_{r+u} m_z m_t + 2 m_{r+z} m_u m_t + 2 m_{r+t} m_u m_z + 2 m_{u+z} m_r m_t +$$
$$+ 2 m_{u+t} m_r m_z + 2 m_{z+t} m_r m_u - 6 m_r m_u m_z m_t - m_{r+u} m_{z+t} +$$
$$- m_{r+z} m_{u+t} - m_{r+t} m_{u+z} - m_{r+u+z} m_t - m_{r+u+t} m_z - m_{r+z+t} m_u +$$
$$\left. - m_{u+z+t} m_r\right],$$

$$M(\delta_r^2 \delta_u \delta_z) = \frac{1}{N^2}\left[3 m_r^2 m_u m_z + m_{2r} m_{u+z} + 2 m_{r+u} m_{r+z} - 2 m_{r+u} m_r m_z + \right.$$
$$\left. - 2 m_{r+z} m_r m_u - m_{2r} m_u m_z - m_{u+z} m_r^2\right] + \frac{1}{N^3}\left[m_{2r+u+z} + \right.$$
$$+ 4 m_{r+u} m_r m_z + 4 m_{r+z} m_r m_u + 2 m_{2r} m_u m_z + 2 m_{u+z} m_r^2 +$$
$$- 6 m_r^2 m_u m_z - m_{2r} m_{u+z} - 2 m_{r+u} m_{r+z} - m_{2r+u} m_z +$$
$$\left. - m_{2r+z} m_u - 2 m_{r+u+z} m_r\right],$$

$$M(\delta_r^2 \delta_u^2) = \frac{1}{N^2}\left[3 m_r^2 m_u^2 + m_{2r} m_{2u} + 2 m_{r+u}^2 - 4 m_{r+u} m_r m_u - m_{2r} m_u^2 + \right.$$
$$\left. - m_{2u} m_r^2\right] + \frac{1}{N^3}\left[m_{2r+2u} + 8 m_{r+u} m_r m_u + 2 m_{2r} m_u^2 + 2 m_{2u} m_r^2 + \right.$$
$$\left. - 6 m_r^2 m_u^2 - m_{2r} m_{2u} - 2 m_{r+u}^2 - 2 m_{2r+u} m_u - 2 m_{2u+r} m_r\right],$$

$$M(\delta_r^3 \delta_u) = \frac{3}{N^2}\left[m_r^3 m_u + m_{r+u} m_{2r} - m_{r+u} m_r^2 - m_{2r} m_r m_u\right] + \frac{1}{N^3}\left[m_{3r+u} + \right.$$
$$\left. + 6 m_{r+u} m_r^2 + 6 m_{2r} m_r m_u - 6 m_r^3 m_u - 3 m_{2r} m_{u+r} - 3 m_{2r+u} m_r - m_{3r} m_u\right],$$

$$M(\delta_r^4) = \frac{3}{N^2}\left[m_r^4 + m_{2r}^2 - 2 m_{2r} m_r^2\right] + \frac{1}{N^3}\left[m_{4r} + 12 m_{2r} m_r^2 - 6 m_r^4 - 3 m_{2r}^2 - 4 m_{3r} m_r\right].$$

3. — Altro metodo per il calcolo dei successivi momenti della v. c. $\mathcal{S}_r$. Non mi soffermerò a trarre qualche ovvia conseguenza dai momenti fin qui calcolati; voglio invece far osservare che la strada seguita nel numero precedente ci fornisce i successivi valori medi, uno alla volta, senza però darci alcuna indicazione, rigorosamente dimostrata, sull'ordine di grandezza dei momenti non ancora calcolati. Per la risoluzione di tale problema si presenta invece più utile un diverso metodo di calcolo, metodo di cui in questo numero mi servirò limitandomi a ricalcolare solo i primi quattro momenti della v. c. $\mathcal{S}_r$. Attraverso le considerazioni che seguono non solo si riuscirà a decidere sull'ordine di grandezza dei successivi momenti, questione che praticamente si poteva considerar risolta, ma soprattutto, ed è per questo che mi soffermo su tali sviluppi, si stabilirà incidentalmente una particolare formuletta approssimata che sarà utile nel successivo paragrafo.

Anzitutto si ricordi che per la v. c. $\mathfrak{X}$ (n. 2) si definiscono le seguenti funzioni generatrici [17]:

— *funzione generatrice dei momenti*:

$$F(\alpha) = M(e^{\alpha\mathfrak{X}}) = 1 + m_1\alpha + m_2\frac{\alpha^2}{2} + \cdots + m_r\frac{\alpha^r}{r!} + \cdots$$

— *funzione generatrice dei seminvarianti*:

$$L(\alpha) = \log F(\alpha) = \lambda_1\alpha + \lambda_2\frac{\alpha^2}{2} + \cdots + \lambda_r\frac{\alpha^r}{r!} + \cdots$$

dove il parametro λ_1 è eguale alla media m_1 della v. c. $\mathfrak{X}$ e gli altri parametri $\lambda_2, \ldots, \lambda_r, \ldots$, introdotti da T. M. Thiele, si chiamano *seminvarianti* e sono omogenei di grado eguale all'indice e del tipo C precisato nell'ultima parte dell'introduzione. L'espressione dei seminvarianti mediante i momenti $\overline{m}_r$ rispetto alla media, come è noto, è la seguente:

$$\begin{aligned}
\lambda_2 &= \overline{m}_2 \\
\lambda_3 &= \overline{m}_3 \\
\lambda_4 &= \overline{m}_4 - 3\,\overline{m}_2^2 \\
\lambda_5 &= \overline{m}_5 - 10\,\overline{m}_2\,\overline{m}_3 \,. \\
&\ldots\ldots\ldots\ldots\ldots\ldots
\end{aligned}$$

(17) Ved. per esempio: A. C. Aitken, op. cit. (1). Si osservi che, non avendo introdotta alcuna ipotesi sulla v. c. $\mathfrak{X}$, nulla possiamo dire circa la convergenza degli sviluppi in serie di potenze delle funzioni generatrici; d'altra parte mi servirò di tali sviluppi solo per calcolare le relazioni tra i loro coefficienti e quindi l'ipotesi della convergenza non è necessaria, ved.: E. Borel, *Leçons sur les séries divergentes*, Parigi, Gauthier-Villars, 1901.

Riguardo a tali espressioni ci interessa osservare che:

I – si presentano come una somma di cui ogni termine è fornito dal prodotto di momenti rispetto alla media tali che la somma dei rispettivi gradi dà il grado d'omogeneità del seminvariante;

II – tutti i termini dello sviluppo di λ_{2r-1}, contengono almeno un momento d'ordine dispari ed ancora, oltre il termine: $\overline{m}_{2r-1}$, v'è, a meno di un coefficiente numerico, un termine del tipo: $\overline{m}_s^{r-3}\,\overline{m}_3$

III – tra i termini dell'espressione di λ_{2r}, sempre a meno di un coefficiente numerico, ve n'è uno del tipo: $\overline{m}_s^r$, ed inoltre vi è il termine: $\overline{m}_{2r}$.

Sulle due funzioni generatrici si dimostrano poi i seguenti due teoremi [18]

a) se $\mathfrak{X}_1$ *ed* $\mathfrak{X}_2$ *sono due v. c. indipendenti e se* $F_1(\alpha)$, $L_1(\alpha)$, $F_2(\alpha)$, $L_2(\alpha)$ *sono le relative funzioni generatrici dei momenti e dei seminvarianti, indicando con* $F(\alpha)$ *ed* $L(\alpha)$ *le funzioni generatrici per la v. c. somma* $\mathfrak{X}_1 + \mathfrak{X}_2$ *si ha*:

$$F(\alpha) = F_1(\alpha) \cdot F_2(\alpha)$$
$$L(\alpha) = L_1(\alpha) + L_2(\alpha)$$

b) *se* $F(\alpha)$ *ed* $L(\alpha)$ *sono le funzioni generatrici per la v. c.* $\mathfrak{X}$, *le analoghe funzioni per la v. c.* $k\mathfrak{X}$ *sono*:

$$F(k\alpha) \quad , \quad L(k\alpha).$$

Ciò posto si consideri la v. c. $\mathfrak{X}^r$ e per essa si definiscano:

i successivi momenti: $m_s^{(r)} = m_{sr}$

i successivi momenti rispetto alla media: $\overline{m}_s^{(r)}$

i successivi seminvarianti: $\lambda_s^{(r)}$

In particolare si ha:

$$\overline{m}_2^{(r)} = m_2^{(r)} - \left(m_1^{(r)}\right)^2 = m_{2r} - m_r^2$$
$$\overline{m}_3^{(r)} = m_3^{(r)} - 3\, m_1^{(r)}\, m_2^{(r)} + 2\left(m_1^{(r)}\right)^3 = m_{3r} - 3\, m_r\, m_{2r} + 2\, m_r^3$$
$$\overline{m}_4^{(r)} = m_4^{(r)} - 4\, m_1^{(r)}\, m_3^{(r)} + 6\left(m_1^{(r)}\right)^2 m_2^{(r)} - 3\left(m_1^{(r)}\right)^4 =$$
$$= m_{4r} - 4\, m_r\, m_{3r} + 6\, m_r^2\, m_{2r} - 3\, m_r^4$$

. . . . . .

$$\lambda_2^{(r)} = \overline{m}_2^{(r)} = m_{2r} - m_r^2$$
$$\lambda_3^{(r)} = \overline{m}_3^{(r)} = m_{3r} - 3\, m_r\, m_{2r} + 2\, m_r^3$$
$$\lambda_4^{(r)} = \overline{m}_4^{(r)} - 3\left(\overline{m}_2^{(r)}\right)^2 = m_{4r} + 12\, m_r^2\, m_{2r} - 4\, m_r\, m_{3r} - 3\, m_{2r}^2 - 6\, m_r^4$$

. . . . . .

(18) Ved. per esempio: A. C. AITKEN, op. cit. (2).

Indichiamo poi con $F^{(r)}(\alpha)$ ed $L^{(r)}(\alpha)$ le due funzioni generatrici per la v. c. $\mathfrak{X}^r$, di modo che:

$$F^{(r)}(\alpha) = 1 + m_1^{(r)}\alpha + m_2^{(r)}\frac{\alpha^2}{2} + \cdots$$

$$L^{(r)}(\alpha) = \lambda_1^{(r)}\alpha + \lambda_2^{(r)}\frac{\alpha^2}{2} + \cdots$$

Si osservi ora che la v. c. $\mathfrak{M}_r$ può essere definita anche nella seguente maniera [19]:

$$\mathfrak{M}_r = \frac{1}{N}\left(\mathfrak{X}_1^r + \mathfrak{X}_2^r + \cdots + \mathfrak{X}_N^r\right)$$

dove: $\mathfrak{X}_1$ è la v. c. $\mathfrak{X}$ alla prima prova, $\mathfrak{X}_2$ è la v. c. $\mathfrak{X}$ alla seconda prova, ..., $\mathfrak{X}_N$ è la v. c. $\mathfrak{X}$ alla N-ma prova; ed è chiaro che, nelle nostre ipotesi, le v. c. $\mathfrak{X}_i$ sono indipendenti tra di loro.

Tenuto conto di ciò e ricordando i due precedenti teoremi sulle funzioni generatrici si ha che la funzione generatrice dei momenti della v. c. $\mathfrak{M}_r$ è data da:

$$F(\alpha) = \left[F^{(r)}\left(\frac{\alpha}{N}\right)\right]^N$$

mentre per la funzione generatrice dei seminvarianti si ha:

$$[7] \qquad L(\alpha) = N\,L^{(r)}\left(\frac{\alpha}{N}\right) = \lambda_1^{(r)}\alpha + \frac{\lambda_2^{(r)}}{N}\frac{\alpha^2}{2} + \frac{\lambda_3^{(r)}}{N^2}\frac{\alpha^3}{3!} + \frac{\lambda_4^{(r)}}{N^3}\frac{\alpha^4}{4!} + \cdots$$

Ricordando ora le espressioni dei seminvarianti di T. M. Thiele in funzione dei momenti rispetto alla media si ha, oltre la già nota relazione:

$$M(\mathfrak{M}_r) = \lambda_1^{(r)} = m_r,$$

il seguente gruppo di relazioni tra i momenti della v. c. $\mathfrak{S}_r$ ed i momenti della v. c. $\mathfrak{X}$:

$$M(\mathfrak{S}_r^2) = \frac{\lambda_2^{(r)}}{N} = \frac{1}{N}\left[m_{2r} - m_r^2\right]$$

$$M(\mathfrak{S}_r^3) = \frac{\lambda_3^{(r)}}{N^2} = \frac{1}{N^2}\left[m_{3r} - 3\,m_r\,m_{2r} + 2\,m_r^3\right]$$

$$M(\mathfrak{S}_r^4) - 3\left[M(\mathfrak{S}_r^2)\right]^2 = \frac{\lambda_4^{(r)}}{N^3} = \frac{1}{N^3}\left[m_{4r} + 12\,m_r^2\,m_{2r} - 4\,m_r\,m_{3r} - 3\,m_{2r}^2 - 6\,m_r^4\right]$$

---

(19) Ved. per esempio G. Castelnuovo, op. cit., (8).

Da queste relazioni si ricavano immediatamente i valori medi già calcolati nel numero precedente.

Questo metodo di calcolo permette di dimostrare che, *al divergere di* N, *i valori medi* $M(\delta_r^{2s-1})$ *e* $M(\delta_r^{2s})$ *tendono a zero come* N^{-s}.

Dimostrerò il precedente teorema per induzione supponendolo vero per i momenti, della v. c. δ_r, d'ordine inferiore a $2s-1$ o $2s$.

Per le cose dette si hanno le due relazioni:

$$[8]\quad \begin{cases} M(\delta_r^{2s-1}) + \sum a_{i_1 i_2 \cdots i_k i_k} [M(\delta_r^{i_1})]^{i_1} \cdots [M(\delta_r^{i_k})]^{i_k} = \dfrac{\lambda_{2s-1}^{(r)}}{N^{s-1}} \\ M(\delta_r^{2s}) + \sum b_{z_1 j_1 \cdots z_l j_l} [M(\delta_r^{z_1})]^{j_1} \cdots [M(\delta_r^{z_l})]^{j_l} = \dfrac{\lambda_{2s}^{(r)}}{N^{s-1}} \end{cases}$$

dove:

$$[9]\quad \begin{cases} t_1 i_1 + t_2 i_2 + \cdots + t_k i_k = 2s-1 & (2s-1 > t_1, t_2, \cdots, t_k) \\ z_1 j_1 + z_2 j_2 + \cdots + z_l j_l = 2s & (2s > z_1, z_2, \cdots, z_l) \end{cases}$$

e le a e b sono coefficienti numerici, indipendenti da N, che non interessa conoscere.

Posto:

$$t_r = 2\tau_r - \vartheta_r \quad (\vartheta_r = 0,1) \quad , \quad z_r = 2\zeta_r - \eta_r \quad (\eta_r = 0,1)$$

le precedenti relazioni [9] diventano:

$$2\sum_1^k{}_r\, \tau_r i_r = 2s - 1 + \sum_1^k{}_r\, \vartheta_r i_r$$

$$2\sum_1^l{}_r\, \zeta_r j_r = 2s + \sum_1^l{}_r\, \eta_r j_r$$

dove:

$$\sum_1^k{}_r\, \vartheta_r i_r \geq 1 \quad , \quad \sum_1^l{}_r\, \eta_r j_r \geq 0$$

di modo che si ottengono finalmente le diseguaglianze:

$$[10]\qquad \sum_1^k{}_r\, \tau_r i_r \geq s \quad , \quad \sum_1^l{}_r\, \zeta_r j_r \geq s.$$

D'altra parte, avendo ammesso il teorema per i momenti d'ordine inferiore, si ha che il prodotto di valori medi che figura come termine nella prima delle [8], se N è grande, è dell'ordine di grandezza di $N^{-\sum_1^k{}_r \tau_r i_r}$ mentre il prodotto di valori medi che figura come termine a primo

membro della seconda delle [8], sempre se N è grande, è dell'ordine di grandezza di $N^{-\frac{1}{2}\Sigma_r t_r j_r}$; di modo che le diseguaglianze [10] assicurano che, se N è sufficientemente grande, l'ordine di grandezza dei due momenti $M(\delta_r^{2s-1})$ e $M(\delta_r^{2s})$ è *non superiore* a N^{-s}; per affermare poi che tale ordine di grandezza è proprio N^{-s} basta ricordare che nella prima delle [8] figura un termine del tipo: $[M(\delta_r^2)]^{s-1} M(\delta_r^3)$, mentre nella seconda figura un termine del tipo: $[M(\delta_r^2)]^s$, ed è chiaro che ambedue questi termini sono dell'ordine di grandezza di N^{-s}.

Ciò posto basta, finalmente, osservare che il teorema è vero per i primi momenti, già da noi calcolati, per concludere che esso è vero qualunque sia s.

Introducendo il solito simbolo « μ^2 » per indicare la « *varianza* » di una v. c., si ha dalle formule già stabilite in questo e nel precedente numero:

$$\mu^2_{\mathfrak{M}_r} = M(\delta_r^2) = \frac{1}{N}[m_{2r} - m_r^2].$$

Tenuto conto di ciò e ricordato che $\lambda_1^{(r)} = m_r$, il precedente sviluppo [7] della funzione generatrice dei seminvarianti per la v. c. $\mathfrak{M}_r$ si può scrivere:

$$L(\alpha) = m_r\alpha + \mu^2_{\mathfrak{M}_r}\frac{\alpha^2}{2} + \frac{\lambda_3^{(r)}}{N^2}\frac{\alpha^3}{3!} + \cdots$$

da cui segue che, quando N è sufficientemente grande in modo che siano trascurabili i termini dell'ordine di grandezza di N^{-2}, sussiste la formula approssimata:

$$L(\alpha) \simeq m_r\alpha + \mu^2_{\mathfrak{M}_r}\frac{\alpha^2}{2}.$$

Da quest'ultima formula, con un errore *relativo* dell'ordine di grandezza di N^{-2}, segue:

$$[11] \qquad F(\alpha) = e^{L(\alpha)} \simeq e^{m_r\alpha + \mu^2_{\mathfrak{M}_r}\frac{\alpha^2}{2}}.$$

D'altra parte tale formula approssimata era da prevedersi perché (20)

$$e^{m_r\alpha + \mu^2_{\mathfrak{M}_r}\frac{\alpha^2}{2}}$$

(20) Ved. per esempio A. C. AITKEN, op. cit. (3).

è la funzione generatrice dei momenti per la v. c. normale di media m_r e varianza $\mu^2_{\mathfrak{M}_r}$, cioè della v. c. che, come è noto [21], approssima la v. c. $\mathfrak{M}_r$ quando N è grande.

Dalla [11] si ottiene in particolare, se N è sufficientemente grande la formula approssimata:

$$M\left(e^{\mathfrak{S}_1}\right) \simeq e^{\frac{1}{2}\frac{\mu^2_{\mathfrak{X}}}{N}} \tag{12}$$

§ II. – LA MEDIA GEOMETRICA

4. – La media geometrica della v. c. divisa in intervalli. Ragioniamo senz'altro sulla v. c. divisa in intervalli. Sia $\mathfrak{X}$ la v. c. che assume i valori reali positivi (e finiti): $x_1, \ldots, x_k$ con probabilità $p_1, \ldots, p_k$.

Mediante la v. c. $\mathfrak{X}$ definiamo la v. c.:

$$\mathfrak{S} = \log \mathfrak{X}$$

che pertanto assume i valori: $\log x_1, \ldots, \log x_k$ con probabilità $p_1, \ldots, p_k$ e, più precisamente, assume il valore $\log x_i$ quando la v. c. $\mathfrak{X}$ assume il valore x_i.

Contraddistinguerò con un apice i vari momenti della v. c. $\mathfrak{S}$ così definita; in particolare si avrà:

$$m'_1 = \sum_i^k p_i \log x_i$$

$$\mu^2_{\mathfrak{S}} = \overline{m}'_2 = m'_2 - (m'_1)^2$$

È noto che la media geometrica m_g della v. c. $\mathfrak{X}$ è definita da:

$$m_g = x_1^{p_1} x_2^{p_2} \cdots x_k^{p_k} = \prod_i^k x_i^{p_i}$$

ed è chiaro che tale media geometrica è legata alla media m'_1 della v. c. $\mathfrak{S}$ da una qualunque delle due immediate relazioni:

$$m_g = e^{m'_1} \quad , \quad m'_1 = \log m_g$$

(21) Ved. per esempio: G. Castelnuovo, op. cit. (2).

5. – La media geometrica del campione–universo. Eseguite N prove sulla v. c. $\mathfrak{X}$ si sia ottenuto: a_1 volte il valore x_1, a_2 volte il valore x_2, ..., a_k volte il valore x_k; con:

$$a_1 + \cdots + a_k = N .$$

Questi N valori x_i formano un «campione» della v. c. $\mathfrak{X}$.

Se consideriamo questo campione come un universo si definisce una nuova v. c. $\mathfrak{X}^*$ che assume i k valori $x_1 \ldots x_k$ con probabilità

$$\frac{a_1}{N}, \ldots, \frac{a_k}{N}.$$

La media geometrica di questa nuova v. c. $\mathfrak{X}^*$, o, come diremo, la media geometrica del campione–universo è data da:

$$m_g^* = \prod_1^k{}_i\, x_i^{\frac{a_i}{N}} = \sqrt[N]{x_1^{a_1} \cdots x_k^{a_k}} .$$

Al variare del campione il parametro m_g^* varia e descrive la v. c.:

$$\mathfrak{M}^* = \prod_1^k{}_i\, x_i^{\frac{\mathfrak{a}_i}{N}}$$

di cui ci proponiamo di calcolare la media.

A tale scopo osserviamo che con lo stesso gruppo di N prove si costruisce anche un campione della v. c. $\mathfrak{S} = \log \mathfrak{X}$, campione che considerato come universo, definisce la v. c. $\mathfrak{S}^* = \log \mathfrak{X}^*$.

Come già si è visto (n. 2) la media del campione–universo coincide con la media del campione di modo che, riferendoci senz'altro a quest'ultima, possiamo scrivere:

$${}_{\mathfrak{S}}m'_1 = \frac{1}{N} \sum_1^k{}_i\, a_i \log x_i .$$

Al variare del campione la media ${}_{\mathfrak{S}}m'_1$ descrive la v. c.:

$$\mathfrak{M}'_1 = \frac{1}{N} \sum_1^k{}_i\, \mathfrak{a}_i \log x_i = m'_1 + \mathfrak{s}'_1$$

già studiata, in generale, nel precedente paragrafo.

Ed è chiaro che tra le due v. c. $\mathfrak{M}_g^*$ ed $\mathfrak{M}'_1$ passa la ovvia relazione:

[13] $$\mathfrak{M}_g^* = e^{\mathfrak{M}'_1} = e^{m'_1 + \mathfrak{s}'_1} = m_g\, e^{\mathfrak{s}'_1}.$$

D'altra parte si ha:

$$e^{\mathfrak{s}'_1} = 1 + \mathfrak{s}'_1 + \frac{1}{2}\mathfrak{s}'^2_1 + \cdots + \frac{1}{r!}\mathfrak{s}'^r_1 + \cdots$$

e quindi, in definitiva, si ottiene:

$$\mathfrak{M}_g^* = m_g \sum_0^\infty {}_r \frac{1}{r!} \delta_1'^r$$

da cui finalmente segue:

$$[14] \qquad \mathrm{M}(\mathfrak{M}_g^*) = m_g \mathrm{M}\left(e^{\delta_1'}\right) = m_g \sum_0^\infty {}_r \frac{1}{r!} \mathrm{M}(\delta_1'^r) .$$

Ricordiamo ora le formule trovate nel precedente paragrafo e scriviamo in forma esplicita i momenti della v. c. δ_1' che compaiono a secondo membro della [14]:

$$\mathrm{M}(\mathfrak{M}_g^*) = m_g \left(1 + \frac{1}{2\mathrm{N}} \mu_\delta^2 + \frac{1}{3!\,\mathrm{N}^2} \overline{m}_3' + \frac{1}{4!\,\mathrm{N}^3} \left[\overline{m}_4' + 3(\mathrm{N}-1)\mu_\delta^4\right] + \cdots\right)$$

Va però osservato che i puntini della precedente formula non rappresentano termini che, per comodità, non si sono scritti perché facilmente ottenibili con formule ricorrenti, ma piuttosto termini che non abbiamo ancora calcolato; d'altra parte anche ammesso che si riesca, cosa impossibile, a calcolare tutti i momenti della v. c. δ_1' si otterrebbe una formula troppo complessa che non si saprebbe come utilizzare.

Tutto ciò mostra come sia necessario ricorrere a formule approssimate e a tale scopo ci sarà utile l'ultima formula del precedente paragrafo; supposto infatti che N sia sufficientemente grande la [12] ci permette di scrivere:

$$[15] \qquad \mathrm{M}(\mathfrak{M}_g^*) = m_g \mathrm{M}\left(e^{\delta_1'}\right) \simeq m_g e^{\frac{1}{2}\frac{\mu_\delta^2}{\mathrm{N}}} .$$

6. – La media geometrica del campione. Dalla precedente formula approssimata [15] si ha che se N è *grandissimo* sussiste la relazione:

$$\mathrm{M}(\mathfrak{M}_g^*) \simeq m_g$$

mentre se N è semplicemente grande risulta più corretta la [15] stessa.

Ricordiamo ora che, per quanto si è visto nell'introduzione, la media geometrica ${}_g m_g$ del campione dovrebbe descrivere una v. c. la cui media è il più possibile prossima al valore teorico m_g; di modo che si sarebbe indotti a porre:

$$[16] \qquad {}_g m_g = e^{-\frac{1}{2}\frac{\mu_\delta^2}{\mathrm{N}}} m_g^* ;$$

si osservi però che il fattore correttivo che figura nella [16], se N è piccolo, non ha alcun significato mentre, se N è grandissimo, è tanto

prossimo ad 1 che è praticamente inutile considerarlo, di modo che la precedente formula [16] avrebbe significato solo nel caso che N sia grande ma non troppo; a tutto questo si aggiunga poi che nel fattore correttivo figura la varianza μ_{ξ}^{2} che, tutto sommato, non si conosce, perché siamo in possesso solo di ${}_{s}\mu_{\xi}^{2}$ calcolata sul campione.

Tutte queste considerazioni inducono a porre senz'altro:

$$[17] \qquad {}_{s}m_{g} = m_{g}^{*} = \sqrt[N]{x_{1}^{n_{1}} \cdots x_{k}^{n_{k}}}$$

7. – *La v. c.* $\mathfrak{M}_{g}$. Al variare del campione la media geometrica ${}_{s}m_{g}$, definita dalla [17], varia e descrive una v. c. che indicherò con $\mathfrak{M}_{g}$; si ha immediatamente:

$$\mathfrak{M}_{g} = \mathfrak{M}_{g}^{*} .$$

Per quel che si è visto nei numeri precedenti il valor medio della v. c. $\mathfrak{M}_{g}$ non coincide con la media geometrica m_{g} della v. c. $\mathfrak{X}$ ed inoltre si riesce a valutare tale valor medio solo nell'ipotesi che il numero N delle prove sia sufficientemente grande in modo da poter trascurare senza danno i termini dell'ordine di grandezza di N^{-2}. Tale ipotesi, cioè che il numero N sia grande, sarà sempre presente in tutto il resto di questo paragrafo anche se non tornerò più a dichiararla esplicitamente.

In base alla [15] si ottiene:

$$M(\mathfrak{M}_{g}) \simeq m_{g}\, e^{\frac{1}{2}\frac{\mu_{\xi}^{2}}{N}} > m_{g}$$

il che appunto mostra che *la media geometrica del campione è in media maggiore della media geometrica della v. c.*

Interessa ora indagare sul grado di attendibilità della media geometrica del campione.

Una prima misura sintetica di tale attendibilità si può ricercare attraverso un indice che misuri lo scostamento medio della v. c. $\mathfrak{M}_{g}$ dal valore teorico m_{g}. A tale scopo definiamo la v. c. *scostamento* (22):

$$\delta_{g} = \mathfrak{M}_{g} - m_{g} = m_{g}\left(e^{\delta_{1}} - 1\right)$$

(22) Non si confonda lo *scostamento* δ_{g} con gli *scarti* δ_{r} definiti nel precedente paragrafo. Lo scarto è uno scostamento dalla media.

e prendiamo il valor medio del quadrato della v.c. δ_g come indice di scostamento; indicherò tale indice con σ^2 (23). Si ha:

$$\sigma^2 = \mathrm{M}(\delta_g^2) = m_g^2 \left[\mathrm{M}\left(e^{2\delta_i'}\right) - 2\,\mathrm{M}\left(e^{\delta_i'}\right) + 1\right] \simeq$$

$$\simeq m_g^2 \left[e^{2\frac{\mu_S^2}{N}} - 2\,e^{\frac{1}{2}\frac{\mu_S^2}{N}} + 1\right] =$$

$$= m_g^2 \left[\left(1 + \frac{2}{N}\mu_S^2 + \frac{2}{N^2}\mu_S^4 + \cdots\right) - \left(2 + \frac{1}{N}\mu_S^2 + \frac{1}{4N^2}\mu_S^4 + \cdots\right) + 1\right] =$$

$$= m_g^2 \left[\frac{1}{N}\mu_S^2 + \frac{7}{4N^2}\mu_S^4 + \cdots\right] \simeq \frac{1}{N} m_g^2 \mu_S^2;$$

e si noti che la seconda approssimazione, cioè quella che ci porta a scrivere:

$$\sigma^2 \simeq \frac{1}{N} m_g^2 \mu_S^2, \tag{18}$$

non rappresenta nulla di nuovo perché già si è convenuto di trascurare i termini dell'ordine di grandezza di N^{-2}.

Possiamo osservare incidentalmente che, *al divergere di* N, *i valori medi* $\mathrm{M}(\delta_g^{2r-1})$ *e* $\mathrm{M}(\delta_g^{2r})$ *tendono a zero come* N^{-r}. Si ha infatti:

$$\mathrm{M}(\delta_g^r) = m_g^r \sum_{0}^{r}{}_j (-1)^j \binom{r}{j} \mathrm{M}\left(e^{(r-j)\delta_i'}\right) =$$

$$= m_g^r \sum_{0}^{r}{}_j (-1)^j \binom{r}{j} \sum_{0}^{\infty}{}_i \frac{(r-j)^i}{i!} \mathrm{M}(\delta_i'^i) =$$

$$= m_g^r \sum_{0}^{\infty}{}_i \frac{1}{i!} \mathrm{M}(\delta_i'^i) \sum_{0}^{r}{}_j (-1)^j \binom{r}{j} (r-j)^i$$

Per dimostrare il nostro assunto basta provare che:

$$\sum_{0}^{r}{}_j (-1)^j \binom{r}{j} (r-j)^i \begin{cases} = 0 & \text{se } i < r \\ \neq 0 & \text{se } i \geq r \end{cases} \tag{19}$$

Se infatti la [19] è vera si può scrivere:

$$\mathrm{M}(\delta_g^r) = m_g^r \sum_{r}^{\infty}{}_i \frac{1}{i!} \mathrm{M}(\delta_i'^i) \sum_{0}^{r}{}_j (-1)^j \binom{r}{j} (r-j)^i$$

ed allora, ricordando il teorema sull'ordine di grandezza dei successivi momenti della v. c. δ_i' (n. 3), si conclude che $\mathrm{M}(\delta_g^r)$, al divergere di N, tende a zero come $\mathrm{M}(\delta_i'^r)$ cioè (n. 3) come N^{-s} se $r = 2s - 1$ oppure $r = 2s$.

---

(23) Se lo scostamento è dalla media l'indice σ^2 coincide ovviamente con la varianza μ^2.

Tutto si riduce dunque a dimostrare la relazione [19] che in questa sede si potrebbe senz'altro considerar come nota; tuttavia, siccome non saprei a quale opera rimandare chi desiderasse controllar tale formula, mi soffermerò a dimostrarla il più rapidamente possibile.

A tale scopo si ponga:

$$F_0(x) = (x-1)^r = \sum_0^r{}_j (-1)^j \binom{r}{j} x^{r-j}$$

$$F_1(x) = x F_0(x) = \sum_0^r{}_j (-1)^j \binom{r}{j} (r-j) x^{r-j}$$

. . . . . .

$$F_i(x) = x F_{i-1}(x) = \sum_0^r{}_j (-1)^j \binom{r}{j} (r-j)^i x^{r-j}$$

. . . . . .

e si osservi che per $x = 1$ il polinomio $F_0(x)$ si annulla insieme alle sue derivate d'ordine inferiore ad r mentre la sua derivata d'ordine r vale $r!$, di modo che ciò accade anche per il polinomio $F_1(x)$ di cui però, per $x = 1$, si annullano solo le derivate d'ordine inferiore a $r - 1$ mentre quella d'ordine $r - 1$ vale $r!$, e così via per gli altri polinomi $F_i(x)$ i quali, se $i < r$, si annullano per $x = 1$ mentre: $F_r(1) = r!$. D'altra parte si ha

$$F_i(1) = \sum_0^r{}_j (-1)^j \binom{r}{j} (r-j)^i$$

di modo che si è dimostrato la parte essenziale della [19]. Resta solo da controllare che $F_i^{\prime}(1) \neq 0$ se $i > r$; a tale scopo basta osservare che il polinomio, di grado, r $F_i(x)$, se $i < r$, si annulla $r - i$ volte per $x = 1$ e semplicemente sia nell'origine che in altri $i - 1$ punti di ascissa positiva e minore di uno; ciò posto, $F_r(x)$ si annulla (semplicemente) nell'origine ed in altri $r - 1$ punti le cui ascisse x_i soddisfano alle diseguaglianze:

$$0 < x_1 < x_2 < \cdots < x_{r-1} < 1 ,$$

ed allora $F_{r+1}(x) = x F_r^{\prime}(x)$ si annulla (semplicemente) nell'origine e, per il teorema di Rolle, in altri $r - 1$ punti le cui ascisse $x_i^{\prime}$ soddisfano alle diseguaglianze:

$$0 < x_1^{\prime} < x_1 < x_2^{\prime} < x_2 < \cdots < x_{r-1}^{\prime} < x_{r-1} < 1 ,$$

e così via; il che appunto dimostra che il polinomio, di grado r, $F_i(x)$, se $i < r$, non si annulla per $x = 1$.

Dopo questa digressione sull'ordine di grandezza dei successivi momenti della v. c. δ_g, ritorniamo al problema, che più ci interessa, dell'attendibilità della media geometrica fornita da un campione.

Si è già trovato un primo indice sintetico di tale attendibilità: lo scostamento

$$\sigma^2 = M(\delta_g^2) \simeq \frac{1}{N} m_g^2 \mu_{\mathfrak{S}}^2.$$

Tale indice σ^2 di scostamento fornisce già un primo notevole risultato in quanto dimostra che, *al divergere del numero* N *delle prove, la media geometrica* $_g m_g$ *del campione diventa sempre più attendibile.*

Ma se vogliamo un risultato ancor più significativo, se cioè vogliamo determinare la funzione di densità della v. c. $\mathfrak{M}_g$ dobbiamo ricorrere, come faremo, ad altro genere di considerazioni.

È noto (24) che, nelle nostre ipotesi, la v. c. δ_i' assume un valore dell'intervallo:

$$-\vartheta \frac{\mu_{\mathfrak{S}}}{\sqrt{N}} \vdash\!\!-\!\!\dashv \vartheta \frac{\mu_{\mathfrak{S}}}{\sqrt{N}}$$

con probabilità circa eguale a:

0,6826	se	$\vartheta = 1$
0,9544	se	$\vartheta = 2$
0,9973	se	$\vartheta = 3$.

Ed allora, in base alla [13], si può affermare che la media geometrica del campione soddisfa, con le stesse probabilità alle diseguaglianze:

$$[20] \qquad m_g e^{-\vartheta \frac{\mu_{\mathfrak{S}}}{N}} \leq {}_g m_g \leq m_g e^{\vartheta \frac{\mu_{\mathfrak{S}}}{N}}.$$

La [20] si può generalizzare dicendo che la media geometrica $_g m_g$ del campione soddisfa alle diseguaglianze:

$$[21] \qquad m_g e^{\vartheta_1 \frac{\mu_{\mathfrak{S}}}{\sqrt{N}}} \leq {}_g m_g \leq m_g e^{\vartheta_2 \frac{\mu_{\mathfrak{S}}}{\sqrt{N}}} \qquad (\vartheta_1 < \vartheta_2)$$

con la stessa probabilità con cui la v. c. S_i' assume un valore dell'intervallo:

$$[22] \qquad \vartheta_1 \frac{\mu_{\mathfrak{S}}}{\sqrt{N}} \vdash\!\!-\!\!\dashv \vartheta_2 \frac{\mu_{\mathfrak{S}}}{\sqrt{N}} \qquad (\vartheta_1 < \vartheta_2)$$

---

(24) Ved. per esempio: G. CASTELNUOVO, op. cit. (2).

Se il numero N delle prove è tanto grande da poter trascurare i termini dell'ordine di grandezza di N^{-1}, le diseguaglianze [21] sono praticamente equivalenti a:

$$[23] \qquad m_g + \vartheta_1 \frac{m_g \mu_{\mathfrak{Z}}}{\sqrt{N}} \leq m_g \leq m_g + \vartheta_2 \frac{m_g \mu_{\mathfrak{Z}}}{\sqrt{N}}$$

ed inoltre la probabilità con cui la v. c. S_i assume un valore dell'intervallo [22] se $|\vartheta_1|$ e $|\vartheta_2|$ non sono molto grandi è bene approssimata dell'integrale (25)

$$[24] \qquad \frac{1}{\sqrt{2\pi}} \int_{\vartheta_1}^{\vartheta_2} e^{-\frac{1}{2}z^2} \, dz ,$$

si può cioè affermare che, *quando i termini dell'ordine di grandezza di* N^{-1} *sono trascurabili*, il valore m_g assunto dalla v. c. $\mathfrak{M}_g$ soddisfa alla [23] con probabilità bene approssimata dalla [24], o, in altre parole, che, nelle nostre ipotesi, *la v. c.* $\mathfrak{M}_g$ *è approssimata dalla v. c. normale di media*: m_g *e varianza*: $\frac{m_g^2 \mu_{\mathfrak{Z}}^2}{N}$.

Osserviamo che la varianza della v. c. normale approssimante coincide con lo scostamento σ^2 già calcolato in precedenza; d'altra parte, nelle nostre ipotesi, si ha:

$$M(\mathfrak{M}_g) \simeq m_g$$

da cui

$$\mu^2 \mathfrak{M}_g \simeq \sigma^2$$

di modo che si potrebbe anche dire che, al divergere di N, la v. c. $\mathfrak{M}_g$ tende uniformemente alla v. c. normale con la stessa media e la stessa varianza.

§ III. – L'INDICE DI MUTABILITÀ

8. – La mutabile casuale. Anche se le modalità qualitative di un carattere si possono concepire come variabili con continuità (per esempio: colore dei capelli), in sede di rilevamento (statistico) del campione siamo costretti a separare artificialmente queste varie modalità in un certo numero di classi incompatibili facendo quello che diremo una *divisione in intervalli*; così, per esempio, divideremo le varie modalità

(25) Ved. per esempio: G. Castelnuovo, op. cit. (8).

del colore dei capelli in quattro intervalli: *biondi, rossi, castani, neri* attribuendo con un qualche criterio (il più delle volte soggettivo) all'uno intervallo o all'altro tutte le possibili gradazioni intermedie.

Ciò premesso segue che ogni serie si può considerare, usando un termine di C. Gini, come sconnessa.

Riepilogando si può dire che un certo carattere può presentare una delle k modalità incompatibili: $u_1, \dots, u_k$; e sappiamo ormai cosa si intende con la frase: « il carattere in esame ha presentato la modalità u_i ».

Consideriamo ora l'evento E_i consistente nel fatto che in un esperimento il carattere in questione assuma la modalità u_i; questo evento si presenterà con una certa probabilità p_i, ed inoltre, dato che gli eventi E_i sono incompatibili e uno di essi si deve sempre presentare in ogni esperimento, si ha la relazione:

$$p_1 + p_2 + \dots + p_k = 1 .$$

Possiamo ora condensare tutto ciò dicendo che al carattere in questione è associta una « *mutabile casuale* » *che assume le modalità* (incompatibili) $u_1, u_2, \dots, u_k$ *con probabilità* $p_1, p_2, \dots, p_k$ *tali che* $\sum_1^k p_i = 1$; indicherò tale mutabile casuale con: $\mathbf{U}$.

L'introduzione del concetto di « mutabile casuale » è, se si vuole, un puro formalismo che non aggiunge nulla a quanto si è già osservato, ma ho ritenuto opportuno far ciò per simmetrizzare i fenomeni a modalità qualitative con quelli a modalità quantitative; in altre parole la « mutabile casuale » (nel seguito abbrevierò con: m. c.) è l'equivalente, per i fenomeni qualitativi, delle « variabili casuali » già introdotte per i fenomeni quantitativi.

Si osservi che, stante la nostra definizione, la m. c. corrisponde alla v. c. divisa in intervalli e quindi sarebbe pleonastico dire « m. c. divisa in intervalli ». Tuttavia credo sia opportuno qualche volta usar quest'ultima espressione sia per ricordare l'arbitrarietà della divisione in intervalli, sia per conservare una frase simmetrica all'analoga: « v. c. divisa in intervalli ».

Circa l'arbitrarietà, precedentemente menzionata, della divisione in intervalli, si osservi che, se il fenomeno qualitativo è per sua natura continuo, la frase: « m. c. associata al fenomeno » non ha un preciso significato fino a che non si sia precisato in qual modo si è fatto la divisione in intervalli.

9. – L'INDICE DI MUTABILITÀ DELLA M. C. DIVISA IN INTERVALLI. Per misurare la variabilità dei fenomeni sia quantitativi che qualitativi C. GINI [26] ha proposto vari indici tra cui uno, adatto per i fenomeni qualitativi, che propongo di chiamare senz'altro « *indice di mutabilità* »; di tale indice mi occuperò in questo terzo paragrafo cominciando, secondo il solito, a definirlo per la m. c. divisa in intervalli.

Sia $\mathbf{U}$ una m. c. che assume le modalità $u_1, u_2, \ldots, u_k$ con probabilità $p_1, p_2, \ldots, p_k$; posto $q_i = 1 - p_i$ si definisce con C. GINI, per questa m. c., il seguente « indice di mutabilità »:

$$g_{\mathbf{U}} = \sum_i^k p_i q_i = 1 - \sum_i^k p_i^2 . \tag{24}$$

È chiaro che l'indice di mutabilità $g_{\mathbf{U}}$ vale zero se la m. c. $\mathbf{U}$ è costante, cioè assume una sola modalità con probabilità: 1, invece è massimo (a parità di k) se tutte le modalità hanno la stessa probabilità di presentarsi, cioè se:

$$p_1 = p_2 = \cdots = p_k = \frac{1}{k}$$

e in tal caso si ha:

$$g_{\mathbf{U}} = \frac{k-1}{k} .$$

10. – L'INDICE DI MUTABILITÀ DEL CAMPIONE–UNIVERSO. Eseguiti N esperimenti sulla m. c. $\mathbf{U}$, la modalità u_i si sia presentata a_i volte $(i = 1, 2, \ldots, k)$; in tal modo si è costruito un campione della m. c..

Consideriamo tale campione come un universo, venendo così a definire una nuova m. c. il cui indice di mutabilità è, ovviamente;

$$g^*_{\mathbf{U}} = \frac{1}{N^2} \sum_i^k a_i (N - a_i) .$$

Al variare del campione l'indice $g^*_{\mathbf{U}}$ varia e descrive la v. c.:

$$\mathfrak{g}^* = \frac{1}{N^2} \sum_i^k \mathfrak{a}_i (N - \mathfrak{a}_i) \tag{25}$$

di cui vogliamo calcolare la media.

(26) Ved. per esempio.: C. GINI op. cit. (1).

La precedente formula [25], introducendo gli scarti $\mathfrak{A}_i$ si può scrivere:

$$\mathfrak{S}^* = g_U - \frac{2}{N} \sum_i p_i \mathfrak{A}_i - \frac{1}{N^2} \sum_i \mathfrak{A}_i^2$$

da cui:

$$M(\mathfrak{S}^*) = g_U - \frac{2}{N} \sum_i p_i M(\mathfrak{A}_i) - \frac{1}{N^2} \sum_i M(\mathfrak{A}_i^2) ;$$

e finalmente, ricordando le formule [6] del n. 1, si ha:

$$[26] \qquad M(\mathfrak{S}^*) = g_U - \frac{1}{N} \sum_i p_i q_i = \frac{N-1}{N} g_U .$$

11. – L'INDICE DI MUTABILITÀ DEL CAMPIONE. In base allo schema già discusso nell'introduzione, tenuto conto della [26], possiamo ora definire l'indice di mutabilità ${}_{c}g_U$ del campione ponendo:

$$[27] \qquad {}_{c}g_U = \frac{N}{N-1} g_U^* = \frac{1}{N(N-1)} \sum_i a_i (N - a_i) .$$

12. – LA V. C. $\mathfrak{G}$. Al variare del campione l'indice di mutabilità ${}_{c}g_U$ varia e descrive una v. c. che indicherò con $\mathfrak{G}$. Tenuto conto della definizione [27] dell'indice ${}_{c}g_U$ si ottiene la seguente espressione per la v. c. $\mathfrak{G}$:

$$[28] \qquad \mathfrak{G} = \frac{N}{N-1} \mathfrak{S}^* = g_U + \frac{1}{N-1} \left[g_U - 2 \sum_i p_i \mathfrak{A}_i - \frac{1}{N} \sum \mathfrak{A}_i^2 \right]$$

da cui.

$$M(\mathfrak{G}) = \frac{N}{N-1} M(\mathfrak{S}^*) = g_U$$

come appunto volevamo ottenere.

Bisogna ora indagare sul grado di attendibilità dell'indice ${}_{c}g_U$ calcolato sul campione. Per questo scopo è utile conoscere la varianza della v. c. $\mathfrak{G}$, in quanto si può affermare che l'indice di mutabilità del campione è tanto più attendibile quanto più la varianza è piccola.

Indichiamo con $\delta_{\mathfrak{G}}$ lo scarto della v. c. $\mathfrak{G}$, cioè:

$$\delta_{\mathfrak{G}} = \mathfrak{G} - g_U = \frac{1}{N-1} \left[g_U - 2 \sum_i p_i \mathfrak{A}_i - \frac{1}{N} \sum_i \mathfrak{A}_i^2 \right]$$

indichiamo poi con $\mu_{\mathfrak{G}}^2$ la varianza della v. c. $\mathfrak{G}$ definita, come è nota, da:

$$\mu_{\mathfrak{G}}^2 = M(\delta_{\mathfrak{G}}^2) .$$

Per determinare il valore di $\mu_{\mathfrak{S}}^2$ conviene anzitutto calcolare i seguenti valori medi:

$$M(\mathfrak{S}_{\mathfrak{S}}\,\mathfrak{A}_j) = \frac{1}{N-1}\Big[g_U\,M(\mathfrak{A}_j) - 2\sum_i^{(+)} p_i\,M(\mathfrak{A}_i\,\mathfrak{A}_j) - 2\,p_j\,M(\mathfrak{A}_j^2) +$$
$$-\frac{1}{N}\sum_i^{(+)} M(\mathfrak{A}_i^2\,\mathfrak{A}_j) - \frac{1}{N}\,M(\mathfrak{A}_j^3)\Big] = 2\,p_j\,(q_j - g_U)$$

$$M(\mathfrak{S}_{\mathfrak{S}}\,\mathfrak{A}_j^2) = \frac{1}{N-1}\Big[g_U\,M(\mathfrak{A}_j^2) - 2\sum_i^{(+)} p_i\,M(\mathfrak{A}_i\,\mathfrak{A}_j^2) - 2\,p_j\,M(\mathfrak{A}_j^3) +$$
$$-\frac{1}{N}\sum_i^{(+)} M(\mathfrak{A}_i^2\,\mathfrak{A}_j^2) - \frac{1}{N}\,M(\mathfrak{A}_j^4)\Big] =$$
$$= 2\,p_j\,[(q_j - 2\,p_j)\,(q_j - g_U) - p_j\,q_j]\,.$$

Mediante questi valori medi possiamo finalmente calcolare la varianza della v. c. $\mathfrak{S}$; e precisamente:

$$\mu_{\mathfrak{S}}^2 = M(\mathfrak{S}_{\mathfrak{S}}^2) = \frac{1}{N-1}\Big[g_U\,M(\mathfrak{S}_{\mathfrak{S}}) - 2\sum_i p_i\,M(\mathfrak{A}_i\,\mathfrak{S}_{\mathfrak{S}}) - \frac{1}{N}\sum_i M(\mathfrak{A}_i^2\,\mathfrak{S}_{\mathfrak{S}})\Big] =$$
$$= \frac{4}{N}\Big[g_U\,(1 - g_U) - \sum_i p_i^2\,q_i\Big] + \frac{2}{N\,(N-1)}\Big[2\sum p_i^2\,q_i - g_U\,(1 - g_U)\Big]\,.$$

Si osservi, incidentalmente, che quando tutte le probabilità sono eguali, cioè:

$$p_1 = p_2 = \cdots = p_k = \frac{1}{k}\,,$$

la varianza, sopra calcolata, della v. c. $\mathfrak{S}$ vale:

$$\mu_{\mathfrak{S}}^2 = \frac{2}{N\,(N-1)} \cdot \frac{k-1}{k^2}$$

cioè, al divergere del numero N delle prove, tende a zero come N^{-2}; mentre, per valori generici delle probabilità p_i, la stessa varianza tende a zero come N^{-1}.