FUNZIONE CARATTERISTICA DI UN FENOMENO ALEATORIO

In: « *Memorie della R. Accademia dei Lincei* », 1930, vol. IV, fasc. 5, pp. 86-133.

RELAZIONE

letta dal socio G. Castelnuovo relatore, a nome anche del socio T. Levi-Civita, sulla Memoria del Dott. B de Finetti dal titolo: *Funzione caratteristica di un fenomeno aleatorio.*

Nel Calcolo delle probabilità succede spesso di dover considerare un evento che può esser sottoposto a un numero illimitato di prove ed ha la stessa probabilità *a priori* di verificarsi in ciascuna prova; può darsi anzi che sia costante la probabilità *a priori* che l'evento si presenti un numero prestabilito di volte in n prove, qualunque sia l'ordine in cui le prove favorevoli si seguono. Non si esclude però che l'esito di una prova possa influire sulle successive, in guisa che le probabilità *a priori* debbano modificarsi quando si conoscano nuove circostanze accompagnanti l'evento. A siffatti eventi, che il De Finetti chiama *fenomeni aleatori*, è dedicata la maggior parte della Memoria sulla quale qui riferiamo.

L'Autore, definita in modo opportuno la funzione caratteristica di un fenomeno aleatorio, dimostra come il fenomeno venga completamente individuato, sia pure in modo astratto, dalla conoscenza di quella funzione. Egli ne profitta per estendere alla classe dei fenomeni aleatori proprietà che valgono per lo schema di Bernoulli, il quale costituisce un caso particolarissimo della classe suddetta.

La definizione sopra riportata di fenomeni aleatori lascia apparire il legame che lo studio di essi ha colla teoria delle probabilità *a posteriori* e delle probabilità delle ipotesi. Il De Finetti ne profitta per precisare qualche punto di queste teorie, ove una maggiore determinazione è possibile.

L'interesse dell'argomento trattato ed i metodi elevati di Analisi coi quali la ricerca è condotta ci inducono a proporre che il lavoro del De Finetti sia inserito tra le Memorie accademiche.

Funzione caratteristica di un fenomeno aleatorio

Memoria del dott. BRUNO de FINETTI

§ 1. — Argomento.

Il caso più generale cui si riferiscono le presenti ricerche è quello delle classi d'eventi equivalenti che sarà trattato nel Cap. IV; il problema praticamente più significativo che vi rientra è quello dei fenemeni aleatori, e forma l'argomento dei primi tre capitoli ([1]).

([1]) Tale problema fu oggetto d'una comunicazione dell'A. al Congresso Internazionale dei Matematici, Bologna 1928, Sez. IV-A, che può considerarsi un riassunto dei Cap. I-IIIdella presente memoria. Un riassunto brevissimo ed elementare vedasi in « Boll. dell'U. M. I. », 1929, N. 2.

Daremo nel primo la definizione precisa dei *fenomeni aleatori*, stabilendone varie proprietà anche concettualmente interessanti, e dimostrando come ogni fenomeno aleatorio si possa completamente individuare mediante la sua funzione caratteristica (¹). Nel secondo si studieranno quelle operazioni sulle funzioni caratteristiche che traducono analiticamente dei problemi di calcolo delle probabilità sui fenomeni aleatori. Nel terzo si applicheranno i metodi di calcolo così introdotti allo studio di due argomenti importanti del calcolo delle probabilità: le probabilità a posteriori e le probabilità delle ipotesi.

Il quarto capitolo mostrerà quali siano le condizioni caratteristiche delle classi d'eventi in cui — anche all'infuori del caso che siano prove di un medesimo fenomeno aleatorio — i metodi precedentemente studiati risultano applicabili, e ci condurrà a trattare anche un problema nuovo.

Seguirà in Appendice la dimostrazione di un teorema che occorre nel testo come lemma, ma ha notevole importanza anche di per sè e nella teoria delle variabili casuali. Esso assicura l'esistenza di una legge limite di probabilità quando la funzione caratteristica tende uniformemente a una funzione limite, e viene a colmare una lacuna nella teoria delle funzioni caratteristiche.

Capitolo Primo

I fenomeni aleatori.

§ 2. — Definizione di fenomeno aleatorio.

Un fenomeno di cui si può fare (o quanto meno si può concepire) un numero qualunque di prove lo diremo *fenomeno aleatorio* quando l'ordine in cui le prove favorevoli e sfavorevoli si alternano sia da attribuirsi al caso (²). Si esige cioè che (quali si siano n, h) le $\binom{n}{h}$ successioni di n prove di cui h favorevoli, successioni che differiscono tra loro solo per l'ordine, abbiano tutte uguale probabilità. Questa, in termini precisi, la proprietà caratteristica di quelli che abbiamo convenuto di definire *fenomeni aleatori*.

Sarà bene vedere con qualche esempio la portata di tale restrizione, e avere così un'idea chiara del campo di questa ricerca. Se si ha una moneta o un dado, e lo si lancia sempre allo stesso modo, non ci sarà nessun motivo di indole causale, nemmeno se della perfezione del pezzo non siamo sicuri, che possa influire sull'ordine in cui si alternano le prove favorevoli e sfavorevoli: l'ordine sarà dovuto al caso, e si ha quindi un fenomeno aleatorio secondo la data definizione. Lo stesso si dica per il problema della roulette, per le estrazioni da un'urna scelta a sorte in una collezione nota, e tutti i casi consimili.

(¹) Per il metodo della funzione caratteristica si vedano p. es. i trattati del Lévy e del Castelnuovo.

(²) È forse preferibile lasciare alla locuzione «fenomeno aleatorio» un significato affatto generico, e dire «fenomeni aleatori a prove equivalenti» quelli che qui consideriamo (per la ragione della nuova denominazione che propongo vedasi il Cap. IV). Nel presente lavoro, dedicato esclusivamente ai fenomeni aleatori «a prove equivalenti», tale specificazione può essere ad ogni modo sottintesa senza pericolo d'ambiguità.

Se invece si considera una successione di tiri al bersaglio di uno stesso tiratore, o la successione delle giornate piovose e non piovose, o delle giornate in cui il signore di rimpetto si rade la barba, tale condizione non si potrà ragionevolmente ritenere verificata, perchè nel primo caso si può prevedere dapprima un progressivo addestramento del tiratore, e poi il sopravvenire della stanchezza, ciò che rende probabile un addensamento delle prove favorevoli nel periodo di forma migliore, perchè i giorni piovosi saranno riuniti in periodi di piovosità più o meno lunghi, senza parlare poi della periodicità stagionale, e perchè il signore di rimpetto si raderà sempre a intervalli più o meno regolari.

Per decidere, in pratica, se un certo fenomeno si possa considerare fenomeno aleatorio, oppure no, basta pensare se un'eventuale regolarità o altra particolarità riscontrata nell'ordine della successione si attribuirebbe al caso (e allora si ha un fenomeno aleatorio) o si potrebbe ritenere dovuta a qualche circostanza connessa al fenomeno, in modo da far pensare che anche in un'altra uguale serie di prove sia probabile si rinnovi.

§ 3. — Sul concetto di fenomeno aleatorio.

Sarà opportuno insistere ancora sul concetto di fenomeno aleatorio, sia per giustificare la denominazione, che a taluno sembrerà troppo limitata, che per evitare possibili equivoci sul significato delle ricerche che seguiranno.

Il caso più semplice di fenomeno aleatorio è quello ben noto per cui in ogni prova si abbia una stessa probabilità p, indipendentemente dall'esito delle altre prove (V. § 13). All'infuori di questo caso la probabilità varia in seguito alla conoscenza dell'esito delle diverse prove che mano mano successivamente si acquisisce, ma varia in un senso ben diverso da quello preso in considerazione, ad es., nello schema di POISSON.

Una successione di eventi, per essere prove di un medesimo fenomeno aleatorio secondo la definizione usata, deve avvenire sempre nelle identiche condizioni; cambiandosi le condizioni in cui si verificano, dovranno considerarsi prove di fenomeni *diversi*. Se sappiamo, ad esempio, che una successione di estrazioni si fa alternativamente da due urne diverse, le estrazioni di posto dispari (essendo fatte dalla prima urna) sono prove di un fenomeno, le altre (dalla seconda) di un fenomeno distinto.

Il fatto che le condizioni debbano rimanere invariate non impedisce che varii la probabilità: essa varierà, in generale, nel modo che vedremo nel Cap. terzo. Ma varierà non perchè il *fatto* che alcune prove si siano verificate e altre no influisca sulla possibilità che altre prove si abbiano a verificare, ma perchè la *conoscenza* dell'esito di alcune prove ci permette di apprezzare meglio le condizioni in cui, fino dal principio, il fenomeno si svolge. Se inizialmente siamo indecisi fra diverse ipotesi, in seguito alla conoscenza dell'esito di alcune prove ci appariranno più probabili quelle maggiormente in accordo colla frequenza osservata, e per effetto di ciò anche la probabilità dovrà subire qualche variazione.

Il primo teorema che enunceremo nel prossimo § dimostra che le condizioni prese a definire i fenomeni aleatori conducono a considerare i problemi di probabilità discussi abbastanza ampiamente dal LÉVY (*op. cit.*, Ch. V. n. 34), e trattati ad esempio dal BERTRAND nel Cap. sulle probabilità a posteriori (*Calcul des Probabilités*, Ch. VII; specialmente i nn. 121, 122, 136).

Il modo di concepire tali problemi è però, nei citati Autori e in tutti gli altri molti che se ne occuparono, completamente diverso dal nostro. Sarà bene richiamarlo per evitare possibili equivoci e per lumeggiare le ragioni per cui ci sembra necessario partire da un'impostazione radicalmente diversa.

Quelli che abbiamo definito col nome di « fenomeni aleatori » sono i fenomeni le cui prove sarebbero, nell'ordinaria terminologia, *indipendenti e con probabilità costante ma incognita*. Ora, il parlare di *probabilità incognite* è, secondo la concezione soggettiva della probabilità ([1]), cosa priva di senso, e in ogni caso oscura e capziosa. Parlare di prove *indipendenti*, nel caso che ci interessa, è per lo meno improprio, dato che in seguito all'esito delle prove precedenti la probabilità delle prove successive si modifica. Secondo la concezione soggettiva è senz'altro erroneo ([2]). Occorre poi considerare la probabilità che la probabilità supposta costante ma incognita p assuma questo o quel valore, o, in altre parole, la « legge di probabilità » di p, e non si vede quale contenuto positivo potrebbe avere tale frase ([3]).

Ad ogni modo, dall'uso che si fa ordinariamente della nozione, non certo troppo chiara e felice, di fenomeno « a prove indipendenti e con probabilità costante ma incognita » risulta che da essa si ritiene di poter dedurre che, se sappiamo che sono state fatte n prove e m di esse sono risultate favorevoli, tutti i modi possibili in cui le prove favorevoli e sfavorevoli si possono alternare fra loro ci appaiono ugualmente probabili. Si deduce cioè la proprietà che definisce i nostri « fenomeni aleatori », e cioè una proprietà perfettamente sensata e significativa.

Inversamente, il Teor. I che enunceremo nel prossimo §, o, meglio, la [20] del § 10, mostrano che, ammesso la solita concezione abbia senso, un fenomeno aleatorio è per l'appunto un fenomeno « a prove indipendenti con probabilità costante ma incognita p », ove la Φ si interpreti come « legge di probabilità (funzione di ripartizione) della probabilità incognita p » ([4]). Ciò che costituisce la giustificazione *formale*, ma *solamente formale*, della solita impostazione, che rimane sempre, concettualmente, per lo meno discutibile.

([1]) Per un'esposizione completa della mia concezione della probabilità si veda il lavoro « *Probabilismo. Saggio critico sulla teoria delle probabilità e sul valore della scienza* », che sarà pubblicato nella rivista filosofica « Logos ». Un breve riassunto vedasi in « Boll. U. M. I. », dicembre 1930. I fondamenti matematici della teoria delle probabilità secondo tale punto di vista saranno sviluppati in altri lavori di cui alcuni in preparazione. Una breve nota sull'argomento v. in Rend. Lincei, novembre 1930.

Il n. 22 del citato *Probabilismo* è interamente dedicato all'esame critico della nostra definizione di « fenomeno aleatorio » e di quella usuale di « fenomeno a prove indipendenti con probabilità costante ma incognita ».

([2]) Si potrebbe però introdurre il concetto di « *indipendenza subordinata* », che traduce in forma precisa l'idea male espressa dalla locuzione criticata. Me ne occuperò in altro lavoro.

([3]) All'infuori dei casi in cui si conosca a priori una classe di eventi subordinatamente a ciascuno dei quali il fenomeno abbia prove indipendenti con probabilità costante e *nota* p. Come ad esempio per le estrazioni da un'urna scelta a caso in una collezione nota.

([4]) Cfr. l'ultimo capoverso del § 10.

§ 4. — Riassunto dei risultati.

Se di un certo fenomeno aleatorio si fanno n prove, il numero di quelle che risultano favorevoli è ovviamente una variabile casuale X_n capace di assumere soltanto i valori $0, 1, \ldots, n$; se indichiamo con $\omega_h^{(n)}$ la probabilità che il fenomeno considerato si verifichi h volte su n prove, la variabile casuale X_n è caratterizzata dalle probabilità $\omega_0^{(n)}, \omega_1^{(n)}, \ldots, \omega_n^{(n)}$ colle quali può assumere i diversi valori possibili. Nel caso particolare e ben noto in cui il fenomeno abbia una probabilità costante p nota a priori, si sa che $\omega_h^{(n)} = \binom{n}{h} p^h (1-p)^{n-h}$, ma nel caso generale di cui ci occupiamo le $\omega_h^{(n)}$ godono di un larghissimo grado d'arbitrarietà.

Un fenomeno aleatorio ci definirà dunque una successione di variabili casuali $X_1, X_2, \ldots, X_n, \ldots$ che è chiaro debbano risultare tra loro interdipendenti. Tale interdipendenza si traduce analiticamente in una relazione differenziale ricorrente che lega le loro funzioni caratteristiche $\psi_1, \psi_2, \ldots, \psi_n, \ldots$ e che costituisce la base di questa ricerca.

Riassumiamo qui brevemente i risultati dei prossimi §§.

Si dimostra che al crescere di n la funzione $\psi_n\left(\frac{t}{n}\right) = \sum_0^n{}_h\, \omega_h^{(n)}\, e^{i\frac{h}{n}t}$ tende uniformemente in ogni regione finita alla funzione intera $\psi(t) = \sum_0^\infty{}_h\, \omega_h^{(h)} \frac{i^h\, t^h}{h\,!}$, che è quella appunto che si dirà, per definizione, la « funzione caratteristica » del fenomeno aleatorio. Nota la ψ si ricavano tutte le ψ_n e di conseguenza tutte le $\omega_h^{(n)}$, e ciò giustifica bene la denominazione.

L'integrale da $-\infty$ a $+\infty$ della funzione $\frac{e^{it} - e^{-i\xi t}}{i\,t}\, \psi(t)$ esiste sempre, è reale, mai decrescente al crescere di ξ, ed uguale rispettivamente a 0 e 2π per $\xi < 0$ e $\xi > 1$; di conseguenza esiste una variabile casuale di cui $\psi(t)$ è la funzione caratteristica, la corrispondente funzione di ripartizione è

$$\Phi(\xi) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{e^{it} - e^{-i\xi t}}{i\,t}\, \psi(t)\, d\,t,$$

ed è $\Phi(\xi) = 0$ per $\xi < 0$, $\Phi(\xi) = 1$ per $\xi > 1$.

Da tali risultati scendono due teoremi importanti:

I) La probabilità che la frequenza su n prove sia contenuta entro limiti assegnati ξ_1 e ξ_2 tende a $\Phi(\xi_2) - \Phi(\xi_1)$ al crescere di n;

(1) Nella mia nota *Sui passaggi al limite nel calcolo delle probabilità* (« Rend. R. Ist. Lombardo », 1930) ho dimostrato che tale locuzione è erronea: non si può parlare della probabilità che *tutte* le frequenze dopo l'n-*esima* siamo comprese fra ξ_1 e ξ_2, ma solo del limite, per $m \to \infty$, della probabilità che lo siano tutte le frequenze dall'n-*esima* alla $(n+m)$*esima*. Lo stesso rilievo si tenga presente a proposito di analoghe espressioni usate nel seguito, come ad es. nei §§ 11 e 34.

II) La probabilità che tutte le frequenze dopo l'n-esima [1] siano comprese fra limiti assegnati ξ_1 e ξ_2 tende a $\underset{\xi=\xi_2}{lims}\,\Phi(\xi) - \underset{\xi=\xi_1}{limd}\,\Phi(\xi)$ al crescere di n.

Perchè, data una $\psi(t)$, possa esistere un fenomeno aleatorio di cui essa sia funzione caratteristica occorre e basta che la corrispondente funzione di ripartizione $\Phi(\xi)$ (naturalmente reale e mai decrescente) sia nulla per $\xi < 0$ e $= 1$ per $\xi > 1$.

§ 5. — La relazione ricorrente fondamentale.

Tra le $\omega_h^{(n)}$ dovrà sussistere la relazione

$$[1] \qquad \omega_k^{(m)} = \sum_k^{n-m+k}{}_h\, \omega_h^{(n)} \frac{\binom{h}{k}\binom{n-h}{m-k}}{\binom{n}{m}} \qquad (m \leq n)$$

perchè $\binom{h}{k}\binom{n-h}{m-k} : \binom{n}{m}$ è la probabilità che su m prove, prese tra n di cui h favorevoli, le favorevoli siano k, quando tutte le combinazioni sono ugualmente probabili [1]. In particolare, per $m = n - 1$:

$$[2] \qquad n\,\omega_k^{(n-1)} = (n-k)\,\omega_k^{(n)} + (k+1)\,\omega_{k+1}^{(n)},$$

e ponendo

$$[3] \qquad \Omega_n(z) = \sum_0^n{}_h\, \omega_h^{(n)} z^h$$

tutte le [2] per $k = 0, 1, \ldots, n-1$ si riassumono nella relazione differenziale ricorrente

$$[4] \qquad n\,\Omega_{n-1}(z) = n\,\Omega_n(z) + (1-z)\,D\,\Omega_n(z).$$

Infatti

$$n\,\Omega_{n-1}(z) = n \sum_0^{n-1}{}_k\, \omega_k^{(n-1)} z^k = \sum_0^{n-1}{}_k\,(n-k)\,\omega_k^{(n)} z^k + \sum_0^{n-1}{}_k\,(k+1)\,\omega_{k+1}^{(n)} z^k =$$
$$= \sum_0^n{}_k\,(n-k)\,\omega_k^{(n)} z^k + \sum_0^n{}_k\, k\,\omega_k^{(n)} z^{k-1} = n \sum_0^n{}_k\, \omega_k^{(n)} z^k + (1-z) \sum_0^n{}_k\, k\,\omega_k^{(n)} z^{k-1} =$$
$$= n\,\Omega_n(z) + (1-z)\,D\,\Omega_n(z).$$

Osserviamo che la funzione caratteristica $\psi_n(t)$ non è altro se non

$$[5] \qquad \psi_n(t) = \Omega_n(e^{it})$$

da cui

$$[5'] \qquad \Omega_n(z) = \psi_n(-i \log z)$$

e quindi la [4] — essendo $D\,\psi_n(t) = i\,e^{it}\,D\,\Omega(e^{it})$ — potrebbe scriversi mediante le ψ_n:

[1] Cfr. ad es. CASTELNUOVO, *Op. cit.*, Vol. I, p. 15.

[4′] $$n\,\psi_{n-1}(t) = n\,\psi_n(t) - i\,(e^{-it} - 1)\,D\,\psi_n(t).$$

Benchè le Ω_n non abbiano un significato immediato come le funzioni caratteristiche, pure la forma [4] è preferibile perchè si presta meglio allo sviluppo dei calcoli.

Derivando la [4] m volte si ha

[6] $$n\,D^m\Omega_{n-1}(z) = (n-m)\,D^m\Omega_n(z) + (1-z)\,D^{m+1}\Omega_n(z)$$

da cui interessa ricavare i valori delle derivate per $z=1$. È $n\,D^m\,\Omega_{n-1}(1) = (n-m)\,D^m\,\Omega_n(1)$, ossia, moltiplicando per $(n-m-1)!\,m!/n!$:

[7] $$\frac{D^m\Omega_{n-1}(1)}{\binom{n-1}{m}} = \frac{D^m\Omega_n(1)}{\binom{n}{m}}$$

da cui

[8] $$\frac{D^m\,\Omega_n(1)}{\binom{n}{m}} = D^m\Omega_m(1).$$

Ora

$$D^m\Omega_m(z) = D^m\sum_0^m{}_h\,\omega_h^{(n)}\,z^h = \omega_m^{(m)}\,D^m z^m = m!\,\omega_m^{(m)}$$

e quindi

[9] $$\frac{1}{m!}\,D^m\Omega_n(1) = \binom{n}{m}\,\omega_m^{(m)}.$$

§ 6. — Tendenza al limite dei polimoni Ω_n.

Possiamo ora sviluppare $\Omega_n(1+z)$ nella forma

[10] $$\Omega_n(1+z) = \sum_0^n{}_h\,\frac{D^h\,\Omega_n(1)}{h!}\,z^h = \sum_0^n{}_h\binom{n}{h}\,\omega_h^{(h)}\,z^h$$

e dimostrare che, al crescere di n, $\Omega_n\left(1+\frac{z}{n}\right)$ tende uniformemente in ogni regione finita alla funzione intera

[11] $$\Omega(1+z) = \sum_0^\infty{}_h\,\omega_h^{(h)}\,\frac{z^h}{h!}.$$

Che Ω sia funzione intera risulta dal fatto che

$$|\,\omega_h^{(h)}| = \omega_h^{(h)} \leq 1\,,\ |\,\Omega(1+z)\,| \leq \sum_0^\infty{}_h\,\omega_h^{(h)}\,\frac{|\,z\,|^h}{h!} \leq \sum_0^\infty{}_h\,\frac{|\,z\,|^h}{h!} = e^{|z|}\,.$$

Dimostriamo ora che, fissati z ed ε, possiamo sempre determinare N tale che per $n > N$ sia sempre

$$|\,\Omega(1+z)-\Omega_n\left(1+\frac{z}{n}\right)|<\varepsilon.$$

$$|\,\Omega(1+z)-\Omega_n\left(1+\frac{z}{n}\right)|=|\sum_0^\infty{}_h\,\omega_h^{(h)}\frac{z^h}{h\,!}-\sum_0^n{}_h\frac{\binom{n}{h}}{n^h}\,\omega_h^{(h)}\,z^h\,|\leq$$

$$\leq\sum_0^n{}_h\,\omega_h^{(h)}\,|z|^h\left|\frac{1}{h\,!}-\frac{\binom{n}{h}}{n^h}\right|+\sum_{n+1}^\infty{}_h\,\omega_h^{(h)}\frac{|z|^h}{h\,!}\leq$$

$$\leq\sum_0^n{}_h\,|z|^h\left|\frac{1}{h\,!}-\frac{\binom{n}{h}}{n^h}\right|+\sum_{n+1}^\infty{}_h\frac{|z|^h}{h\,!}.$$

Ma

$$\frac{\binom{n}{h}}{n^h}=\frac{n\,!}{h\,!\,(n-h)\,!\,n^h}\leq\frac{1}{h\,!}$$

perchè

$$\frac{n\,!}{(n-h)\,!\,n^h}\leq 1$$

$$(=\text{per}\quad h=0,1,\quad<\text{per}\quad h>1)$$

e quindi

$$\left|\frac{1}{h!}-\frac{\binom{n}{h}}{n^h}\right|=\frac{1}{h!}-\frac{\binom{n}{h}}{n^h}\qquad,\text{e}$$

$$\sum_0^n{}_h\,|z|^h\left|\frac{1}{h!}-\frac{\binom{n}{h}}{n^h}\right|+\sum_{n+1}^\infty{}_h\frac{|z|^h}{h\,!}=\sum_0^n{}_h\frac{|z|^h}{h\,!}+$$

$$+\sum_{n+1}^\infty{}_h\frac{|\,z\,|^h}{h\,!}-\sum_0^n{}_h\binom{n}{h}\left(\frac{|\,z\,|}{n}\right)^h=e^{|z|}-\left(1+\frac{|\,z\,|}{n}\right)^n.$$

Per determinare il massimo di questa funzione basta osservare che essa è crescente con $|\,z\,|$ (infatti la derivata è positiva) e quindi in una regione finita in cui $|\,z\,|\leq a$

[12] $$\left|\Omega\,(1+z)-\Omega_n\left(\frac{z}{n}\right)\right|\leq e^a-\left(1+\frac{a}{n}\right)^n.$$

Anzi risulta che la tendenza al limite è uniforme per tutte le possibili successioni di polimoni Ω_n. Possiamo enunciare quindi in modo completo: *fissati* a *ed* ε, *si può sempre determinare* N *tale che per ogni* n $>$ N *sia*

$$\left|\Omega\,(1+z)-\Omega_n\left(1+\frac{z}{n}\right)\right|<\varepsilon$$

***in tutto il cerchio* $|\,\mathrm{z}\,|<\mathrm{a}$ *e per qualunque funzione* Ω.**

È anche facile vedere che per z reale positivo $\Omega_n\left(1+\frac{z}{n}\right)$ tende al limite crescendo.

§ 7. — Tendenza al limite delle funzioni caratteristiche.

Il risultato di maggiore interesse è quello analogo per le funzioni caratteristiche. Dimostreremo infatti che

$$\psi_n\left(\frac{t}{n}\right)=\Omega_n\left(e^{i\frac{t}{n}}\right)$$

tende a

[13] $$\psi(t)=\Omega(1+i\,t)=\sum_0^\infty{}_h\,\omega_h^{(h)}\,\frac{i^h\,t^h}{h\,!}.$$

Si ha infatti

$$\left|\Omega_n\left(e^{\frac{t}{n}}\right)-\Omega_n\left(1+\frac{t}{n}\right)\right|\leq M_n\left|e^{\frac{t}{n}}-\left(1+\frac{t}{n}\right)\right|$$

ove M_n sia il massimo modulo della derivata di Ω_n sul segmento $\left(e^{\frac{t}{n}},1+\frac{t}{n}\right)$, che è compreso, per $|\,t\,|<\tau$, nel cerchio di centro 1 e raggio $e^{\frac{\tau}{n}}-1$; la disuguaglianza vale in tutto il campo $|\,t\,|<\tau$ se M_n rappresenta il massimo modulo della derivata nel cerchio predetto.

Calcoliamolo:

$$D\Omega_n(1+t)=\sum_0^n{}_h\,h\binom{n}{h}\omega_h^{(h)}\,t^{h-1}=n\sum_1^n{}_h\binom{n-1}{h-1}\omega_h^{(h)}\,t^{h-1}=n\sum_0^{n-1}{}_k\binom{n-1}{k}\omega_{k+1}^{(k+1)}t^h$$

$$|\,D\,\Omega_n(1+t)\,|\leq\sum_0^{n-1}{}_k\binom{n-1}{k}\omega_{k+1}^{(k+1)}\,|\,t\,|^k=n\sum_0^{n-1}{}_k\frac{(n-1)\,!}{k\,!\,(n-1-k)\,!\,n^k}\,\omega_k^{(k)}\,|\,n\,t\,|^k\leq$$

$$\leq n\sum_0^{n-1}{}_k\,\omega_k^{(k)}\,\frac{|\,n\,t\,|^k}{k\,!}<n\,e^{n\,|\,t\,|}.$$

Quindi per $|\,t\,\leq e^{\frac{\tau}{n}}-1$ è $|\,D\,\Omega_n(1+t)\,|<n\,e^{n\left(e^{\frac{\tau}{n}}-1\right)}<n\,e^{2\,\tau}$ (purchè $n>\tau$).

Infatti la funzione $\frac{1}{x}\,(e^x-1)$ è crescente con x, e quindi, se $n>\tau$:

$$n\left(e^{\frac{\tau}{n}}-1\right)=\tau\,.\,\frac{n}{\tau}\left(e^{\frac{\tau}{n}}-1\right)<\tau\,(e-1)<2\,\tau.$$

È poi:

$$\left|e^{\frac{t}{n}}-\left(1+\frac{t}{n}\right)\right|=\left|\sum_{2}^{\infty}{}_h \frac{t^h}{n^h\,h\,!}\right|\leq\sum_{2}^{\infty}{}_h \frac{|\,t\,|^h}{n^h\,h\,!}<\frac{1}{n^2}\sum_{2}^{\infty}{}_h \frac{|\,t\,|^h}{h\,!}<\frac{e^{|t|}}{n^2}<\frac{e^{\tau}}{n^2}.$$

Quindi

$$\left|\Omega_n\left(e^{\frac{t}{n}}\right)-\Omega_n\left(1+\frac{t}{n}\right)\right|<n\,e^{2\,\tau}.\frac{e^{\tau}}{n^2}=\frac{e^{3\,\tau}}{n},$$

e poichè

$$|\,\Omega_n\left(1+\frac{t}{n}\right)-\Omega\,(1+t)\,|<e^{\tau}-\left(1+\frac{\tau}{n}\right)^n$$

si ha

$$\left|\Omega\,(1+t)-\Omega_n\left(e^{\frac{t}{n}}\right)\right|<\frac{1}{n}e^{3\,\tau}+e^{\tau}-\left(1+\frac{\tau}{n}\right)^n$$

ossia

$$[14]\qquad \left|\psi\,(t)-\psi_n\left(\frac{t}{n}\right)\right|<\frac{1}{n}\,e^{3\,\tau}\;+e^{\tau}-\left(1+\frac{\tau}{n}\right)^n\qquad(\text{per } n>\tau).$$

Ne risulta che *in ogni campo finito* ($|\,t\,|<\tau$), *assegnato comunque* ε, *è possibile determinare* N *in modo che per ogni* $\mathrm{n}>\mathrm{N}$ *si abbia in tutto il campo* $\left|\,\psi\,(t)-\psi_{\mathrm{n}}\left(\frac{\mathrm{t}}{\mathrm{n}}\right)\right|<\varepsilon$, *qualunque sia la funzione caratteristica di fenomeno aleatorio* $\psi\,(t)$.

§ 8. — Funzione caratteristica di un fenomeno aleatorio.

Abbiamo così dimostrato che $\psi_n\left(\frac{t}{n}\right)$ tende uniformemente alla $\psi\,(t)$, che è quella che abbiamo definita funzione caratteristica del fenomeno aleatorio.

Osserviamo che $\psi_n\left(\frac{t}{n}\right)$ è la funzione caratteristica della frequenza su n prove; possiamo dire: la funzione caratteristica della frequenza su n prove tende uniformemente verso una funzione limite al crescere di n. E ricordando che i coefficienti dello sviluppo della funzione caratteristica sono i successivi momenti, si ha che il momento m-esimo della frequenza su n prove, ossia il valor probabile della potenza m-esima della frequenza su n prove, tende a $\omega_m^{(m)}$ quando $n\rightarrow\infty$.

Il valor-probabile-limite della potenza m-esima *della frequenza per un numero indefinitamente crescente di prove è uguale al valore della probabilità che* m *prove siano tutte favorevoli.*

Le [11], [10], [5] provano chiaramente che la ψ basta a determinare completamente tutte le ψ_n, e quindi le $\omega_h^{(n)}$, come s'era asserito.

§ 9. — Funzione di ripartizione di un fenomeno aleatorio.

Dimostreremo in Appendice che se la funzione caratteristica di una legge variabile di probabilità tende uniformemente a una funzione limite, questa funzione è funzione caratteristica di una legge fissa, e la funzione di ripartizione della legge variabile tende alla funzione di ripartizione della legge fissa a meno tutt'al più che in un insieme numerabile di punti. Nel caso che ci interessa vedremo inoltre che quest'eccezione non si presenta mai.

Indicando quindi $\Phi_n(\xi)$ la funzione di ripartizione corrispondente a $\psi_n(t)$:

[15] $$\Phi_n(\xi) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{it} - e^{-i\xi t}}{it}\, \psi_n(t)\, dt$$

e $\Phi(\xi)$ la funzione di ripartizione corrispondente a $\psi(t)$:

[16] $$\Phi(\xi) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{it} - e^{-i\xi t}}{it}\, \psi(t)\, dt$$

si ha che

[17] $$\lim_{n=\infty} \Phi_n(n\xi) = \Phi(\xi).$$

$\Phi_n(n\xi)$ è la probabilità che la frequenza su n prove sia minore di ξ, più eventualmente (e cioè per $n\xi$ intero) la metà delle probabilità che la frequenza sia esattamente uguale a ξ. La [17] esprime quindi il Teor. I del § 4, nel cui enunciato la precedente specificazione è sottintesa, ciò che va tenuto presente.

§ 10. — Delle funzioni di ripartizione.

Dalla $\Phi(\xi)$ si può ricavare [1] la $\psi(t)$:

[18] $$\psi(t) = \int_0^1 e^{it\xi}\, d\Phi = e^{it} - it\int_0^1 e^{it\xi}\, \Phi(\xi)\, d\xi$$

e quindi essa pure basta a determinare interamente tutte le ψ_n e $\omega_h^{(n)}$:

[19] $$\Omega_n(1+z) = \int_0^1 (1+z\xi)^n\, d\Phi = (1+z)^n - nz\int_0^1 (1+z\xi)^{n-1}\, \Phi(\xi)\, d\xi$$

[20] $$\omega_h^{(n)} = \binom{n}{h}\int_0^1 \xi^h (1-\xi)^{n-h}\, d\Phi = \binom{n}{h}\int_0^1 (h-n\xi)\, \xi^{h-1} (1-\xi)^{n-h-1}\, \Phi(\xi)\, d\xi$$

e con ciò a caratterizzare completamente il fenomeno aleatorio.

Per dimostrare la [19] basta sviluppare $(1+z\xi)^n$ e integrare i singoli termini ricordando che il momento h-esimo $\int_0^1 \xi^h\, d\Phi$ è uguale a $\omega_h^{(h)}$, e confrontare

(1) Cfr. i citati trattati nel LÉVY e del CASTELNUOVO. Gli integrali in $d\Phi$ sono sempre integrali di STIELTJES.

colla [10]. Si ha così $\Omega_n(z) = \int_0^1 (1 - \xi + z\,\xi)^n \, d\,\Phi$, e sviluppando secondo le potenze di z si ottengono come coefficienti le espressioni delle $\omega_h^{(n)}$ date dalle [20]. Esse permettono di esprimere le $\omega_h^{(n)}$ mediante le $\omega_h^{(h)}$ (relazione che si poteva dedurre direttamente dalla [1] per induzione)

[20′] $$\omega_h^{(n)} = \binom{n}{h} \sum_0^{n-h}{}_i (-1)^i \binom{n-1}{i} \omega_{h+i}^{(h+i)}$$

Mostriamo ora che condizione necessaria e sufficente perchè una funzione ψ possa essere funzione caratteristica di un fenomeno aleatorio è che la $\Phi(\xi)$ corrispondente (reale, mai decrescente) sia nulla per $\xi < 0$ e $= 1$ per $\xi > 1$. Che la condizione sia necessaria scende ovviamente dal fatto che per $\xi < 0$, $\xi > 1$ si ha sempre $\Phi_n(n\,\xi) = 0$, $\Phi_n(n\,\xi) = 1$. Ma lo si può constatare direttamente, e non sarà inutile rendersi ragione dell'asserto per tale via.

Supponiamo, $\Phi(-\varepsilon) = \alpha$ $(\varepsilon > 0,\ \alpha > 0)$. Vedremo che $\psi(t) = \int_0^1 e^{it\xi} \, d\,\Phi$ non può essere funzione caratteristica di un fenomeno aleatorio perchè alcune delle $\omega_h^{(n)}$ risultano certamente negative. È infatti.

$$\omega_1^{(n)} = n \int_0^1 \xi (1 - \xi)^{n-1} \, d\,\Phi$$

e, per n dispari,

$$\frac{\omega_1^{(n)}}{n} = \int_0^\alpha \xi (1 - \xi)^{n-1} \, d\,\Phi + \int_\alpha^1 \xi (1 - \xi)^{n-1} \, d\,\Phi < -\alpha\varepsilon(1 + \varepsilon)^{n-1} + (1 - \alpha).$$

e quindi $\omega_1^{(n)} < 0$ per n sufficientemente grande: assurdo.

Analogamente si dimostra che non può essere

$$\Phi(1 + \varepsilon) = 1 - \beta \quad (\varepsilon > 0, \beta > 0).$$

La condizione è sufficiente, perchè, qualunque sia Φ, la successione di funzioni $\Omega_n(1 + z) = \int_0^1 (1 - z\,\xi)^n \, d\,\Phi$ sodisfa la [4], $\Omega_n(1) = 1$, e quindi le

$$\omega_h^{(n)} = \binom{n}{h} \int_0^1 \xi^h (1 - \xi)^{n-h} \, d\,\Phi$$

sodisfano la [1], $\sum_0^n{}_h\, \omega_h^{(n)} = 1$, e sono tutte positive (o al più nulle) se per $\xi > 0$, $\xi > 1$ è rispettivamente $\Phi = 0$, $\Phi = 1$. Ed è ovvio che nulla vieta, in tali condizioni, di considerare la $\omega_h^{(n)}$ come probabilità del verificarsi in h su n prove di un qualche fenomeno.

Non contentandosi di un ragionamento così astratto, si potrà mostrare come, data una funzione di ripartizione Φ, uguale rispettivamente a 0 e 1 per $\xi > 0$ e $\xi > 1$, si possa dare un esempio di fenomeno aleatorio di cui essa è la funzione di ripartizione [1]. Si tratti di determinare a sorte un punto di una circonferenza, di cui archi uguali abbiamo probabilità uguali (l'esperienza si può imaginare realizzata con una roulette). La circonferenza è divisa in due parti, che diremo una rossa e una nera. Noi ignoriamo la lunghezza dell'arco rosso, che indicheramo con p (presa la circonferenza come unità di misura), ma abbiamo elementi per valutare la probabilità a priori delle diverse ipotesi che si possono fare sul valore di p: lo potremo supporre ad esempio determinato a sorte mediante un'esperienza preventiva, in cui si presenti una variabile casuale X capace di assumere i valori tra 0 e 1. Basta che la X segua la legge di probabilità Φ perchè, determinando $p = X$, si definisca un fenomeno aleatorio di cui Φ è la funzione di ripartizione.

§ 11. — Teorema di Cantelli pei fenomeni aleatori.

Il secondo teorema del § 4 relativo alla probabilità che tutte le frequenze da un certo n in poi appartengano a un dato intervallo si può considerare come l'estensione al caso di qualunque fenomeno aleatorio del noto teorema di CANTELLI. Anche la dimostrazione si può ricondurre a quella del caso trattato da CANTELLI.

Premettiamo il seguente *lemma. Se* $\Phi(\xi_1) = 0$, *e* $\xi_2 < \xi_1$, *la probabilità che una almeno fra le frequenze dall'*n-esima *in poi siano inferiori a* ξ_2 *tende a* 0 *col crescere di* n.

Infatti la probabilità che la frequenza r-esima sia $\frac{h}{r}$ è

$$\omega_h^{(r)} = \binom{r}{h} \int_0^1 \xi^h (1-\xi)^{r-h}\, d\,\Phi,$$

e poichè $\xi^h (1-\xi)^{r-h}$ è massimo per $\xi = \frac{h}{r}$ e poi decrescente, nell'ipotesi che per $0 < \Phi < 1$ sia $\xi_1 \leq \xi \leq 1$, se $\frac{h}{r} < \xi_2 < \xi_1$, è in tutto l'intervallo d'integrazione $\xi^h (1-\xi)^{r-h} \leq \xi_1{}^h (1-\xi_1)^{r-h}$ e quindi

$$\omega_h^{(r)} \leq \binom{r}{h} \xi_1^h (1-\xi_1)^{r-h}.$$

La probabilità che la frequenza r-esima sia inferiore a ξ_2 è

$$q_r = \sum_{\frac{h}{r} < \xi_2} \omega_h^{(r)} \leq \bar{q}_r = \sum_{\frac{h}{r} < \xi_2} \binom{r}{h} \xi_1^h (1-\xi_1)^{r-h}$$

---

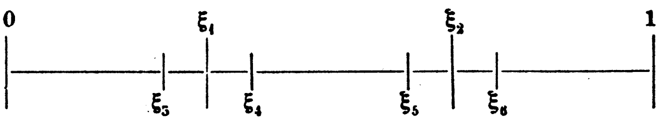
[1] Cfr. LÉVY, *op. cit.*, Ch. V, N. 34 (pag. 88).

dove $\bar{q}_r$ è dunque la probabilità che sia inferiore a ξ_2 la frequenza r-esima di un fenomeno a probabilità costante ξ_1.

Ora il CANTELLI ha dimostrato appunto(¹) che $\sum_r \bar{q}_r$ è convergente, e quindi la probabilità che almeno una frequenza a partire dall'n-esima sia inferiore a ξ_2, certo minore di $\sum_n^\infty{}_r \bar{q}_r$, tende a zero quando $n \to \infty$. Dalla $q_r \leq \bar{q}_r$ scende che a maggior ragione l'asserto vale nel nostro caso.

Analogamente, *se* $\Phi(\xi_1) = 1$, *e* $\xi_2 > \xi_1$, *la probabilità che una almeno delle frequenze dall'*n-esima *in poi sia superiore a* ξ_2 *tende a zero col crescere di* n.

Osserviamo poi che se $\Phi(\xi)$ è la funzione di ripartizione di una variabile casuale X, e suddividiamo comunque il campo dei valori che essa può assumere, è sempre $\Phi(\xi) = \lambda_1 \Phi^{(1)}(\xi) + \lambda_2 \Phi^{(2)}(\xi) + \ldots + \lambda_{(n)} \Phi^{(n)}(\xi)$, ove λ_i e la probabilità che X cada nella suddivisione i-esima, e $\Phi^{(i)}(\xi)$ la funzione di ripartizione della X subordinatamente a tale ipotesi. La stessa scomposizione si può operare per $\Phi(\xi)$ funzione di ripartizione di un fenomeno aleatorio, e, come vedremo al § 22, essa è suscettibile della medesima interpretazione. Grazie a tale artificio potremo cercare il limite, per $n \to \infty$, della probabilità che un certo fenomeno aleatorio, la cui funzione di ripartizione indichiamo con $\Phi(\xi)$, abbia tutte le frequenze dell'n-esima in poi comprese fra due valori assegnati ξ_1 e ξ_2 $(0 < \xi_2 < \xi_1 < 1)$, riconducendosi ai casi considerati nel lemma. Suddividiamo l'intervallo (0, 1) in cui può cadere la frequenza anzitutto in tre parti separate tra loro da ξ_1 e ξ_2, e poi, fissato ad arbitrio ε, isolando a destra e e sinistra di ξ_1 come di ξ_2 degli intervallini con probabilità minore di ε.



Precisamente, sceglieremo $\xi_3 < \xi_1$ in modo che $\lim_{\xi = \xi_1} s\, \Phi(\xi) - \Phi(\xi_3) < \varepsilon$ $\xi_4 > \xi_1$ in modo che $\Phi(\xi_4) - \lim_{\xi = \xi_1} d\, \Phi(\xi) < \varepsilon$, e analogamente ξ_5 e ξ_6 ai lati di ξ_2. Supporremo naturalmente $0 < \xi_3 < \xi_1 < \xi_4 < \xi_5 < \xi_2 < \xi_6 < 1$. Al nostro fenomeno aleatorio possiamo sostituirne uno formato dalla combinazione eventuale di altri 9, di cui il primo ha probabilità $\Phi(\xi_3)$ e ξ limitata fra 0 e ξ_3 (cioè: la sua funzione di ripartizione, fuori di tale intervallo, è rispettivamente uguale a 0 e 1), l'ultimo ha probabilità $1 - \Phi(\xi_6)$, e ξ limitata fra ξ_6 e 1, quello centrale ha probabilità $\Phi(\xi_5) - \Phi(\xi_4)$ e ξ limitata fra ξ_4 e ξ_5, due corrispondono ai valori ξ_1 e ξ_2, hanno probabilità uguale al salto (eventuale) di Φ in ξ_1 e ξ_2 (differenza fra i limiti destro e sinistro), quattro corrispondono agli intervallini a fianco di ξ_1 e ξ_2, e hanno, tutti insieme, probabilità minore di $4\,\varepsilon$.

Nel primo e nell'ultimo caso, per il lemma dimostrato, la probabilità che

(¹) Cfr. ad es. CASTELNUOVO, *op. cit.*, Vol. I, p. 78, dove si trovano anche citati i lavori originali del CANTELLI. Il caso trattato da tale Autore è, a dir vero, più generale di quello riportato dal CASTELNUOVO, ma noi qui non abbiamo bisogno che di riferirci a quel caso particolare, dovendo estenderlo in tutt'altro senso.

tutte le frequenze dopo l'n-*esima* siano comprese fra ξ_1 e ξ_2 è nulla. Nei casi corrispondenti a ξ_1, ξ_2 essa è pure nulla, per l'altro teorema di CANTELLI che afferma essere nulla la probabilità che la frequenza non oscilli infinite volte intorno alla probabilità (1). Nel caso centrale la probabilità che tutte le frequenze dopo l'n-*esima* siano comprese tra ξ_1 e ξ_2 tende a 1 per $n \to \infty$, perchè tendono a zero le probabilità che anche una sola superi ξ_2 o non raggiunga ξ_1: potremo fissare N in modo che, per ogni $n > N$, detta probabilità differisca da 1 per meno di ε. Gli altri casi hanno probabilità minore di $4\,\varepsilon$; quindi la probabilità che il fenomeno aleatorio abbia tutte le frequenze dopo l'n-*esima* comprese tra ξ_1 e ξ_2 differisce, se $n > N$, per meno di $4\,\varepsilon + [\Phi(\xi_3) - \Phi(\xi_4)]\,\varepsilon < 5\,\varepsilon$, da $\Phi(\xi_3) - \Phi(\xi_4)$, e quindi differisce da $\underset{\xi=\xi_2}{\mathit{lims}}\,\Phi(\xi) - \underset{\xi=\xi_1}{\mathit{limd}}\,\Phi(\xi)$ per meno di $7\,\varepsilon$. Ciò che dimostra il teorema: la probabilità che tutte le frequenza dopo l'n-*esima* siano comprese fra limiti assegnati ξ_1 e ξ_2 tende a $\underset{\xi_1=\xi_2}{\mathit{lims}}\,\Phi(\xi) - \underset{\xi=\xi_1}{\mathit{limd}}\,\Phi(\xi)$ al crescere di n.

§ 12. — Il limite della frequenza.

Non è lecito, a rigore, parlare di una successione illimitata di prove di un fenomeno aleatorio: le prove potranno essere in numero comunque grande, sempre finito. Supponiamo per un momento di poterlo considerare infinito per enunciare in forma un po' artificiosa, ma più semplice e suggestiva, alcuni teoremi sulle frequenze di cui vedremo subito dopo l'esatta interpretazione.

In una successione illimitata di prove di un fenomeno aleatorio si ha sempre la probabilità 1 che esista il limite della frequenza, e, ammesso che esista, tale limite è una variabile casuale che segue la legge di probabilità che ha $\psi(t)$ come funzione caratteristica e $\Phi(\xi)$ come funzione di ripartizione. La probabilità che tale limite sia $=\xi_1$ è quindi il salto $\underset{\xi=\xi_1}{\mathit{limd}}\,\Phi(\xi) - \underset{\xi=\xi_2}{\mathit{limd}}\,\Phi(\xi)$, la probabilità che sia compreso fra ξ_1 e ξ_2 (estremi esclusi) è $\underset{\xi=\xi_2}{\mathit{lims}}\,\Phi(\xi) - \underset{\xi=\xi_1}{\mathit{limd}}\,\Phi(\xi)$. In altre parole si potrebbe dire: tanto il massimo che il minimo limite della frequenza sono variabili casuali che seguono la legge di probabi-

(1) Tale teorema non sussiste evidentemente nell'ipotesi — finora esclusa — $\xi_1 = 0$ o $\xi_2 = 1$. Però in tal caso si vede subito che è nulla la probabilità che anche una sola prova su un numero comunque grande sia favorevole (risp. sfavorevole); il teorema che stabiliremo varrà sempre colla restrizione che se uno degli estremi dell'intervallo è 0 o 1 è necessario avvertire che tale estremo è *escluso*, mentre negli altri casi è indifferente considerare l'intervallo con o senza estremi. La probabilità che tutte le frequenze dopo l'n *esima* siano comprese tra 0 e ξ_2, *zero incluso*, tende, per $n \to \infty$, a $\underset{\xi=\xi_2}{\mathit{lims}}\,\Phi(\xi) = \underset{\xi=\xi_2}{\mathit{lims}}\,\Phi(\xi) - \underset{\xi=0}{\mathit{lims}}\,\Phi(\xi)$ anzichè a $\underset{\xi=\xi_2}{\mathit{lims}}\,\Phi(\xi) = \underset{\xi=0}{\mathit{limd}}\,\Phi(\xi)$, e analogamente per l'intervallo da ξ_1 a 1, *uno incluso*, tende a $1 - \underset{\xi=1}{\mathit{limd}}\,\Phi(\xi) - \underset{\xi=\xi_1}{\mathit{limd}}\,\Phi(\xi)$ anzichè a $\underset{\xi=1}{\mathit{lims}}\,\Phi(\xi) - \underset{\xi=\xi_1}{\mathit{limd}}\,\Phi(\xi)$.

lità $\Phi(\xi)$; la differenza fra il massimo e minimo limite è una variabile casuale che ha probabilità 1 di essere nulla, e probabilità nulla di assumere un valore diverso da zero, benchè i numeri positivi non superiori ad uno siano anch'essi valori effettivamente possibili.

In forma precisa, e cioè interpretando l'enunciato senza la considerazione artificiosa di una successione illimitata di prove, potremo affermare: assegnati ε e θ comunque piccoli, è sempre possibile determinare N così grande che, per quante prove si facciamo dopo l'N-*esima*, non differisca mai da 1 per più di θ la probabilità che la differenza della massima e della minima tra le frequenze dall'N-*esima* in poi sia minore di ε. Mentre l'altra parte dell'enunciato equivale al teorema di CANTELLI del precedente §.

Per dimostrare il nuovo teorema, dividiamo l'intervallo $(0, 1)$ in un numero finito n di parti di lunghezza $< \varepsilon$, escludendo di prendere come punti di divisione i punti di discontinuità di $\Phi(\xi)$ (che formano un aggregato al più numerabile: e la condizione è quindi lecita). Nell'intervallo i-*esimo* (ξ_{i-1}, ξ_i) prendiamo N_i così grande che la probabilità che tutte le frequenze dall'N_i-*esima* in poi siano comprese tra ξ_{i-1} e ξ_i differisca per meno di $\frac{\theta}{n}$ da $\Phi(\xi_i) - \Phi(\xi_{i-1})$ (ciò che è possibile per il teorema di CANTELLI, e essendo per ipotesi Φ continua nei punti di divisione ξ_h: $\underset{\xi = \xi_h}{limd}\ \Phi(\xi) = \underset{\xi = \xi_h}{lims}\ \Phi(\xi) = \Phi(\xi_h)$). Sia N il massimo degli N_i: la probabilità che, per quante prove si facciano dopo l'N-*esima*, le frequenze dall'N-*esima* in poi siano tutte comprese in uno solo degli n intervalli differisce da $[\Phi(\xi_1) - 0] + [\Phi(\xi_2) - \Phi(\xi_1)] + \ldots + [1 - \Phi(\xi_{n-1})] = 1$ per meno di n volte $\frac{\theta}{n}$, e cioè per meno di θ. Ma se tutte le frequenze sono contenute in uno degli intervalli, la cui lunghezza è $< \varepsilon$, la differenza tra la massima e la minima è $< \varepsilon$, e tale fatto ha quindi a maggior ragione una probabilità che differisce da 1 per meno di θ.

§ 13. — Due casi notevoli.

Consideriamo due casi di particolare importanza, che serviranno anche utilmente come esempi.

Nel caso ben noto in cui la probabilità di un fenomeno è nota a priori e uguale a p,

$$\omega_h^{(n)} = \binom{n}{h} p^h (1-p)^{n-h}, \qquad \omega_n^{(n)} = p^n, \qquad \Omega_n(1+z) = (1+pz)^n,$$

$$\Omega(1+z) = \lim_{n=\infty} \left(1 + \frac{pz}{n}\right)^n = e^{pz}, \qquad \psi(t) = e^{ipt},$$

o anche direttamente (dalla [13])

$$\psi(t) = \sum_0^\infty{}_h\, p^h \frac{i^h t^h}{h!} = e^{ipt}.$$

$\Phi(\xi) = 0, {}^1/_2, 1$ a seconda che $\xi < p, \xi = p, \xi > p$. Al crescere di n la probabilità che la frequenza sia compresa in un intorno $p \pm \varepsilon$ di p tende all'unità; inversamente da tale ipotesi scende che $\psi(t) = \int_0^1 e^{ipt}\, d\,\Phi = e^{ipt}$, e cioè che il fenomeno ha sempre la probabilità costante e uguale a p indipendentemente dall'esito delle altre prove (§ 22). Come casi particolari, per $p = 0$, $p = 1$, si ha $\psi(t) = 1$, $\psi(t) = e^{it}$.

Nel caso in cui tutte le frequenze siano ugualmente probabili si avrà

$$\omega_0^{(n)} = \omega_1^{(n)} = \ldots = \omega_n^{(n)} = \frac{1}{n+1},$$

$$\Omega_n(z) = \frac{1}{n+1}(1 + z + z^2 + \ldots + z^n) = \frac{1}{n+1} \cdot \frac{1 - z^{n+1}}{1 - z},$$

$$\psi_n(t) = \frac{1}{n+1} \cdot \frac{1 - e^{it(n+1)}}{1 - e^{it}}, \quad \Omega(1+z) = \lim_{n=\infty} \frac{1}{n+1} \frac{\left(1 + \frac{z}{n}\right)^{n+1} - 1}{\frac{z}{n}} = \frac{e^z - 1}{z},$$

$$\psi(t) = \frac{e^{it} - 1}{it} = \sum_0^\infty{}_h \frac{t^h}{(h+1)!}, \quad \Phi_n(h) = \frac{h + {}^1/_2}{n+1} \qquad (h = 0, 1, \ldots, n),$$

$\Phi(\xi) = \lim\limits_{n=\infty} \Phi_n(n\xi) = \xi$. Al crescere di n la probabilità che la frequenza sia compresa in un intervallo (ξ_1, ξ_2) tende a $\xi_2 - \xi_1$; inversamente da tale ipotesi scende

$$\psi(t) = \int_0^1 e^{it\xi}\, d\,\xi = \left[\frac{e^{it\xi}}{it}\right]_0^1 = \frac{e^{it} - 1}{it}, \quad \omega_h^{(n)} = \frac{1}{n+1},$$

e il fenomeno ha quindi ugualmente probabili tutte le diverse frequenze su n prove.

CAPITOLO SECONDO.

Operazioni sulle funzioni caratteristiche.

§ 14. — Preliminari.

Passiamo alle operazioni sulle funzioni caratteristiche.

Come osservazione generale possiamo dire che tutte le operazioni che incontreremo sono distributive [1], a meno (ove occorra) di un fattore moltiplicativo che fa assumere il valore 1 alla $\psi(t)$ per $t = 0$, come necessariamente deve aversi. Introducendo l'operatore U:

$$U f(t) = \frac{f(t)}{f(0)}$$

[1] V. PINCHERLE e AMALDI, *Le operazioni distributive e loro applicazioni all'Analisi*, Zanichelli ed.

possiamo dire che le operazioni che ci si presenteranno sono prodotti del tipo UF con F distributiva.

Poichè la funzione di ripartizione $\Phi(\xi)$ è funzione lineare biunivoca della funzione caratteristica $\psi(t)$, ad ogni operazione distributiva sulla ψ corrisponde la trasformata che opera su Φ.

Due operazioni utili per semplificare le notazioni sono P_n (leggere: « polinomio ennesimo ») che applicato a ψ produce Ω_n, e $\begin{bmatrix} h \\ n \end{bmatrix}$ che applicato a ψ produce $\omega_h^{(n)}$. Sarà bene, per evitare l'impressione che si voglia dar luogo a un inutile e vuoto formalismo, spiegare esplicitamente il significato concreto di tali simboli. Col simbolo $\begin{bmatrix} h \\ n \end{bmatrix}\psi$ rappresentiamo la probabilità che un fenomeno di funzione caratteristica ψ si verifichi h volte su n prove. Finora la scrivevamo $\omega_h^{(n)}$, e ciò era lecito finchè non si considerava contemporaneamente più d'una legge di probabilità; non lo sarebbe più in seguito, dove è necessario compaia esplicitamente la funzione ψ da cui essenzialmente dipende. $P_n\psi(t)$ è il polinomio finora indicato $\Omega_n(1+t)$, dove le $\omega_h^{(n)}$ siano quelle relative alla legge ψ (e che abbiamo indicate con $\begin{bmatrix} h \\ n \end{bmatrix}\psi$).

Per chi interessi la definizione formale delle operazioni P_n e $\begin{bmatrix} h \\ n \end{bmatrix}$, si può aggiungere che:

se
$$f(t) = \sum_0^\infty {}_h\, a_h \frac{i^h t^h}{h!}$$

[21]
$$P_n f(z) = \sum_0^n {}_h \binom{n}{h} a_h z^h \qquad \text{(Cfr. la [10])}$$

[22]
$$\begin{bmatrix} h \\ n \end{bmatrix} f = \binom{n}{h} \sum_0^{n-h} {}_i (-1)^i \binom{n-1}{i} a_{h+i} \qquad \text{(Cfr. la [20'])}.$$

Si osservi che in particolare $\begin{bmatrix} n \\ n \end{bmatrix} f = a_n$, come naturalmente vuole il significato.

Per la funzione caratteristica del numero delle prove favorevoli sopra n (la ψ_n del Cap. I) si ha

$$\psi_n(t) = P_n\psi(e^{it}-1).$$

§ 15. — Fenomeno contrario. Operazione K.

Sia $\psi(t)$ la funzione caratteristica di un certo fenomeno aleatorio. La funzione caratteristica del fenomeno contrario è

[23]
$$K\psi(t) = e^{it}\psi(-t).$$

Infatti dire che su n prove quelle favorevoli sono h equivale a dire che sono $n-h$ quelle favorevoli all'evento contrario, ciò che si esprime

[24] $$\begin{bmatrix} h \\ n \end{bmatrix} = \begin{bmatrix} n-h \\ n \end{bmatrix} K,$$

e dà (posto $\omega_h^{(n)} = \begin{bmatrix} h \\ n \end{bmatrix} \psi$):

$$P_n K \psi (z-1) = \omega_n^{(n)} + \omega_{n-1}^{(n)} z + \ldots + \omega_0^{(n)} z^n =$$
$$= z^n \left\{ \omega_0^{(n)} + \omega_1^{(n)} \frac{1}{z} + \ldots + \omega_n^{(n)} \frac{1}{z^n} \right\} = z^n P_n \psi \left(\frac{1}{z} - 1 \right)$$

$$P_n K \psi (e^{it} - 1) = e^{nit} P_n \psi (e^{-it} - 1)$$

e

$$K \psi (t) = \lim_{n=\infty} P_n K \psi \left(e^{i \frac{t}{n}} - 1 \right) = \lim_{n=\infty} \left(e^{i \frac{t}{n}} \right)^n \lim_{n=\infty} P_n \psi \left(e^{-i \frac{t}{n}} - 1 \right) =$$
$$= e^{it} \psi (-t), \text{ c. v. d.}$$

Essendo in particolare $\begin{bmatrix} n \\ n \end{bmatrix} K = \begin{bmatrix} 0 \\ n \end{bmatrix}$:

[25] $$K \psi (t) = 1 + \omega_0^{(1)} i t - \omega_0^{(2)} \frac{t^2}{2!} - \omega_0^{(3)} i \frac{t^3}{3!} + \ldots + \omega_0^{(n)} i^n \frac{t^n}{n!} + \ldots$$

È ovvio che K è operazione distributiva: $K(\psi' + \psi'') = K \psi' + K \psi''$. Per il suo significato è chiaro che K dev'essere un'involuzione, perchè ogni fenomeno coincide col fenomeno contrario del fenomeno contrario. Si verifica del resto immediatamente:

$$K K \psi (t) = K [e^{it} \psi (-t)] = e^{it} [e^{-it} \psi(-[-t])] = \psi (t).$$

E quindi ancora: K è reversibile, e $K^{-1} = K$.

È importante osservare che

[26] $$P_n K \psi \left(e^{i \frac{t}{n}} \right) = K \psi_n \left(\frac{t}{n} \right)$$

ove

$$\psi_n \left(\frac{t}{n} \right) = P_n \psi \left(e^{i \frac{t}{n}} - 1 \right).$$

§ 16. — Corrispondente funzione di ripartizione. Operazione K_Φ.

Se Φ è la funzione di ripartizione corrispondente a ψ, la funzione di ripartizione $K_\Phi \Phi$ corrispondente a $K \psi$ è

[27] $$K_\Phi \Phi (\xi) = 1 - \Phi (1 - \xi).$$

Infatti

$$K_\Phi \Phi(\xi) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{e^{it} - e^{-i\xi t}}{it} K\psi(t)\,dt = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{e^{it} - e^{-i\xi t}}{it} e^{it}\psi(-t)\,dt =$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{2it} - e^{-i(\xi-1)t}}{it}\psi(-t)\,dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-2it} - e^{-i(1-\xi)t}}{-it}\psi(t)\,dt =$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{it} - e^{-2it}}{it}\psi(t)\,dt - \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{it} - e^{-i(1-\xi)t}}{it}\psi(t)\,dt =$$

$$= \Phi(2) - \Phi(1-\xi) = 1 - \Phi(1-\xi).$$

Questa relazione non esprime del resto che un fatto intuitivo: la probabilità che al limite la frequenza d'un evento sia inferiore a ξ è uguale a quella che la frequenza dell'evento contrario sia superiore a $1 - \xi$. La si poteva quindi stabilire a priori, e se ne sarebbe immediatamente dedotto

$$K\psi(t) = e^{it} - it\int_0^1 e^{it\xi}[K_\Phi \Phi(\xi)]\,d\xi = e^{it} - it\int_0^1 e^{it\xi}[1 - \Phi(1-\xi)]\,d\xi =$$

$$= e^{it}\left\{1 - it\int_0^1 e^{-it(1-\xi)}\,d\xi + it\int_0^1 e^{-it(1-\xi)}\Phi(1-\xi)\,d\xi\right\} =$$

$$= e^{it}\left\{1 - it\int_0^1 e^{-itz}\,dz + it\int_0^1 e^{-itz}\Phi(z)\,dz\right\} =$$

$$= e^{it}\left\{1 - it\left[\frac{e^{-it} - 1}{-it}\right] + it\int_0^1 e^{-itz}\Phi(z)\,dz\right\} = e^{it}\left\{e^{-it} + it\int_0^1 e^{-itz}\Phi(z)\,dz\right\} = e^{it}\psi(-t).$$

§ 17. — Trasformazioni in seguito all'esito di una prova. Operazioni R, S.

Quando si conosce l'esito di una prova del fenomeno, la probabilità delle prove successive viene modificata nel senso che abbiamo accennato nel § 3. L'influenza dovuta al fatto di conoscere l'esito di una prova sarà espressa nel modo più completo quando avremo determinato il modo in cui si modifica, per tale effetto, la funzione caratteristica di un fenomeno aleatorio. È quasi superfluo avvertire che, se per semplicità di dizione e per fissare le idee sul caso più spontaneo supporremo di conoscere l'esito delle prime prove, in un certo numero, e parleremo della probabilità della prova immediatamente successiva, tutto ciò non ha alcuna importanza. Dato che l'ordine in cui si susseguono le prove è per ipotesi una nozione del tutto estranea al problema che consideriamo, tanto vale conoscere l'esito di uno stesso numero di prove comunque altrimenti disposte, e la probabilità subordinata a tale fatto si riferisce a qualunque prova di cui non si conosce l'esito, e che potrà essere, rispetto a un eventuale criterio d'ordine, p. es. cronologico, in cui le prove si volessero considerare, sia precedente, sia intermedia, sia successiva e comunque distante da quelle il cui esito è noto. Ciò chiarito, passiamo a studiare come si modifica la funzione caratteristica di un fenomeno aleatorio quando si conosca l'esito di alcune prove.

La funzione caratteristica sia, inizialmente, $\psi(t) = \sum_0^\infty{}_h\, \omega_h^{(h)} \dfrac{i^h t^h}{h!}$; nell'ipotesi che la prima prova sia favorevole essa diviene

$$[28] \qquad R\,\psi = U\,D\,\psi,$$

nell'ipotesi che la prima prova sia sfavorevole

$$[29] \qquad S\,\psi = U\,(i - D)\,\psi$$

e, più in generale, dopo r prove favorevoli ed s sfavorevoli, la funzione caratteristica diviene

$$[30] \qquad R^r\, S^s\, \psi = U\, D^r\, (i - D)^s\, \psi.$$

La probabilità $\omega_n^{(n)}$ che le prime n prove siano tutte favorevoli è infatti uguale al prodotto della probabilità $\omega_1^{(1)}$ che la prima prova sia favorevole per la probabilità che, verificata quest'ipotesi, lo siano le prime $n-1$ prove successive, probabilità che è $\begin{bmatrix} n-1 \\ n-1 \end{bmatrix} R\,\psi$. Il coefficente n-esimo dello sviluppo di $R\,\psi$ è quindi il coefficiente $(n+1)$-esimo, $\omega_{n+1}^{(n+1)}$, dello sviluppo di ψ, diviso per il primo, $\omega_1^{(1)} = -\,i\,D\,\psi(0)$, e quindi

$$R\,\psi(t) = \frac{1}{\omega_1^{(1)}} \left\{ \omega_1^{(1)} + \omega_2^{(2)}\, i\,t - \omega_3^{(3)} \frac{t^2}{2!} + \ldots + \omega_{n+1}^{(n+1)}\, i^n \frac{t^n}{n!} + \ldots \right\} =$$
$$= \frac{-\,i\,D\,\psi(t)}{-\,i\,D\,\psi(0)} = U\,D\,\psi(t).$$

Per dimostrare che $S = U(i - D)$ conviene partire dall'osservazione che S è ovviamente la trasformata di R mediante K: $S = K\,R\,K$. Dalla [23]:

$D\,K\,\psi(t) = e^{it}[i\,\psi(-t) - D\,\psi(-t)]$;
$K\,D\,K\,\psi(t) = e^{it}\left\{e^{-it}[i\,\psi(t) - D\,\psi(t)]\right\} = (i - D)\,\psi(t)$;
$K\,D\,K = i - D$; e per ovvie proprietà di U
$S = K\,R\,K = K\,U\,D\,K = U\,K\,D\,K = U(i - D)$, ossia

$$S\,\psi(t) = \frac{i\,\psi(t) - D\,\psi(t)}{i - D\,\psi(0)} = \frac{1}{1 - \omega_1^{(1)}} \sum_0^\infty{}_h \left(\omega_h^{(h)} - \omega_{h+1}^{(h+1)}\right) \frac{i^h t^h}{h!}.$$

Le operazioni R ed S sono permutabili: $RS = SR$, e quindi $R^r S^s = S^s R^r = S^{s_1} R^{r_1} S^{s_2} R^{r_2} \ldots$ $(s_1 + s_2 + \ldots = s,\ r_1 + r_2 + \ldots = r)$. Ciò prova che dopo r prove favorevoli ed s sfavorevoli, indipendentemente dall'ordine in cui esse si alternano, la funzione caratteristica diviene $R^r S^s \psi$.

§ 18. — Proprietà di R, S (¹).

Si ha

[31] $$P_n D = \frac{i}{n+1} D P_{n+1}$$

Sia infatti

$$\psi(t) = \sum_0^\infty {}_h\, a_h \frac{i^h t^h}{h!}:$$

$$D\psi(t) = \sum_0^\infty {}_h\, i\, a_{h+1} \frac{i^h t^h}{h!} \qquad P_n(D\psi)(z) = i \sum_0^n {}_h \binom{n}{h} a_{h+1} z_h$$

$$D P_{n+1} \psi(z) = D \sum_0^{n+1} {}_h \binom{n+1}{h} a_h z^h = \sum_1^{n+1} {}_h \binom{n+1}{h} h\, a_h z^{h-1} = (n+1) \sum_0^n {}_h \binom{n}{h} a_{h+1} z^h$$

e quindi

$$P_n(D\psi) = \frac{i}{n+1} D P_{n+1} \psi.$$

Per induzione

[32] $$P_n D^h = i^h \frac{n!}{(n+h)!} D^h P_{n+h}.$$

Essendo

$$D^r (i - D)^s = D^r \sum_0^s {}_h (-1)^h i^{s-h} \binom{s}{h} D^h = \sum_0^s {}_h\, i^{s+h} \binom{s}{h} D^{r+h}$$

risulta

$$P_n D^r (i - D)^s = \sum_0^s {}_h\, i^{s+h} \binom{s}{h} P_n D^{r+h} = \sum_0^s {}_h\, i^{s+h} \binom{s}{h} i^{r+h} \frac{n!}{(n+r+h)!} D^{r+h} P_{n+r+h} =$$

$$= n!\, i^{r+s} \sum_0^s {}_h (-1)^h \binom{s}{h} \frac{1}{(n+r+h)!} D^{r+h} P_{n+r+h}.$$

Avremo quindi in generale

[33] $$P_n R^r S^s = U \sum_0^s {}_h (-1)^h \binom{s}{h} \frac{1}{(n+r+h)!} D^{r+h} P_{n+r+h};$$

(¹) Questo § può essere omesso; per dimostrazione della [34] si prenderà allora la giustificazione diretta datane nel § successivo.

come caso particolare ($s = 0$)

$$[33'] \qquad P_n R^r = R^r P_{n+r}$$

(ove, mediante la $R = U D$, si estenda la definizione di R a qualunque funzione).

Si rilevi che $P_n R^r S^s \psi$ è quindi pienamente determinato da $P_{n+r+s}\psi$; la conoscenza di $P_h \psi$ per $h < n + r + s$ è invece insufficente alla sua determinazione. Ciò risulta anche, e in modo suscettibile di interpretazione espressiva, dalla formula seguente, che dà la probabilità che un fenomeno aleatorio (di cui è nota la funzione caratteristica ψ) si verifichi h volte su n ulteriori prove, nell'ipotesi che delle prime $r + s$ quelle favorevoli siano r e s le sfavorevoli:

$$[34] \qquad \begin{bmatrix} h \\ n \end{bmatrix} R^r S^s \psi = \frac{\binom{h+r}{r}\binom{n-h+s}{s}}{\binom{n+r+s}{n}} \cdot \frac{\begin{bmatrix} h+r \\ n+r+s \end{bmatrix}\psi}{\begin{bmatrix} r \\ r+s \end{bmatrix}\psi}.$$

Dimostrazione.

Sviluppando

$$P_n R^r \psi(z) = \sum_0^n {}_h \left(\begin{bmatrix} h \\ n \end{bmatrix} R^r \psi \right) (1+z)^h$$

si ha per la [33']

$$P_n R^r \psi(z-1) = U D^r P_{n+r} \psi(z-1) = U D^r \sum_0^{n+r} {}_h \left(\begin{bmatrix} h \\ n+r \end{bmatrix} \psi \right) z^h =$$
$$= U \sum_0^n {}_h \left(\begin{bmatrix} h+r \\ n+r \end{bmatrix} \psi \right) \frac{(h+r)!}{h!} z^h$$

Il valore di $D^r P_{n+r}(0)$ si calcola più facilmente dalla

$$D^r P_{n+r} \psi(0) = i^{-r} \frac{(n+r)!}{n!} P_n D^r \psi(0) = i^{-r} \frac{(n+r)!}{n!} D^r \psi(0) =$$
$$= i^{-r} \frac{(n+r)!}{n!} i^r \begin{bmatrix} r \\ r \end{bmatrix} \psi = \frac{(n+r)!}{n!} \begin{bmatrix} r \\ r \end{bmatrix} \psi$$

e quindi

$$P_n R^r \psi(z-1) = \frac{n!}{(n+r)! \begin{bmatrix} r \\ r \end{bmatrix} \psi} \sum_0^n {}_h \left(\begin{bmatrix} h+r \\ n+r \end{bmatrix} \psi \right) \frac{(h+r)!}{h!} z^h$$

da cui

$$[34'] \qquad \begin{bmatrix} h \\ n \end{bmatrix} R^r \psi = \frac{n!\,(h+r)!}{(n+r)!\,h!} \frac{\begin{bmatrix} h+r \\ n+r \end{bmatrix} \psi}{\begin{bmatrix} r \\ r \end{bmatrix} \psi}$$

Da questa si passa facilmente alla [34] generalizzando:

$$\begin{bmatrix} h \\ n \end{bmatrix} R^r S^s \psi = \begin{bmatrix} h \\ n \end{bmatrix} K R^s K R^r \psi = \begin{bmatrix} n-h \\ n \end{bmatrix} R^s K R^r \psi =$$

$$= \frac{n!\,(n-h+s)!}{(n+s)!\,(n-h)!} \, \frac{\begin{bmatrix} n-h+s \\ n+s \end{bmatrix} K R^r \psi}{\begin{bmatrix} s \\ s \end{bmatrix} K R^r \psi} = \frac{n!\,(n-h+s)!}{(n+s)!\,(n-h)!} \cdot \frac{\begin{bmatrix} h \\ n+s \end{bmatrix} R^r \psi}{\begin{bmatrix} 0 \\ s \end{bmatrix} R^r \psi} =$$

$$\frac{n!\,(n-h+s)!}{(n+s)!\,(n-h)!} \cdot \frac{\dfrac{(n+s)!\,(h+r)!}{(n+s+r)!\,h!}}{\dfrac{s!\,r!}{(r+s)!}} \; \frac{\dfrac{\begin{bmatrix} h+r \\ n+r+s \end{bmatrix} \psi}{\begin{bmatrix} r \\ r \end{bmatrix} \psi}}{\dfrac{\begin{bmatrix} r \\ r+s \end{bmatrix} \psi}{\begin{bmatrix} r \\ r \end{bmatrix} \psi}} = \frac{\binom{h+r}{r}\binom{n-h+s}{s}}{\binom{n+r+s}{n}} \cdot \frac{\begin{bmatrix} h+r \\ n+r+s \end{bmatrix} \psi}{\begin{bmatrix} r \\ r+s \end{bmatrix} \psi},$$

§ 19. — Significato della formula [34].

L'importante formula

$$[34] \qquad \begin{bmatrix} h \\ n \end{bmatrix} R^r S^s \psi = \frac{\binom{h+r}{r}\binom{n-h+s}{s}}{\binom{n+r+s}{n}} \cdot \frac{\begin{bmatrix} h+r \\ n+r+s \end{bmatrix} \psi}{\begin{bmatrix} r \\ r+s \end{bmatrix} \psi}$$

dimostrata nel precedente § dà la probabilità che un fenomeno aleatorio (di cui è nota la funzione caratteristica ψ) si verifichi h volte su n ulteriori prove, nell'ipotesi che delle prime $r+s$ quelle favorevoli siano r ed s le sfavorevoli.

Il suo significato è molto semplice: quando sulle prime $r+s$ prove se ne ebbero r di favorevoli ed s di sfavorevoli, supporre che sulle prossime n le favorevoli saranno h significa supporre che, complessivamente, sulle prime $r+s+n$ prove quelle favorevoli abbiano ad essere $r+h$. La probabilità che $r+h$ delle prime $r+s+n$ prove siano favorevoli è $\begin{bmatrix} r+h \\ n+r+s \end{bmatrix} \psi$, la probabilità che in tale ipotesi r delle prime $r+s$ prove siano favorevoli è ([1])

$$\binom{h+r}{r}\binom{n-h+s}{s} : \binom{n+r+s}{n}$$

(l'ordine essendo dovuto al caso); dividendo per la probabilità che r delle prime $r+s$ prove siano favorevoli, che è $\begin{bmatrix} r \\ r+s \end{bmatrix} \psi$, si ha la probabilità cercata, $\begin{bmatrix} h \\ n \end{bmatrix} R^r S^s \psi$, dell'evento subordinato, qual'è data dalla [34].

([1]) Cfr. § 5 e nota ([1]) ivi.

Vale la pena di soffermarsi un momento sul caso particolare $h = n = 1$, sia per lo speciale interesse che presenta, sia per chiarire utilmente — fissando le idee su un caso semplice — il ragionamento che precede. Ponendo nella [34] $h = n = 1$ si ha senz'altro

[34*] $$\begin{bmatrix}1\\1\end{bmatrix} R^r S^s \psi = \frac{r+s+1}{r+1} \cdot \frac{\begin{bmatrix}r+1\\r+s+1\end{bmatrix}\psi}{\begin{bmatrix}r\\r+s\end{bmatrix}\psi}$$

che esprime la probabilità che la prossima (cioè la $(r+s+1)$-*esima*) prova sia favorevole dopo r prove favorevoli ed s sfavorevoli.

$\begin{bmatrix}r+1\\r+s+1\end{bmatrix}\psi$ è la probabilità che $r+1$ delle prime $r+s+1$ prove siano favorevoli, e, in tale ipotesi, la probabilità che l'ultima (l'$(r+s+1)$-*esima*) di esse sia favorevole è $(r+s+1)/(r+1)$; il prodotto di queste due probabilità è la probabilità che sulle prime $(r+s+1)$ prove quelle favorevoli siano $r+1$, e tra esse l'ultima. Dividendo per la probabilità $\begin{bmatrix}r\\r+s\end{bmatrix}\psi$ che r delle prime $r+s$ prove siano favorevoli si ottiene la probabilità cercata $\begin{bmatrix}1\\1\end{bmatrix} R^r S^s \psi$ dell'evento subordinato, quale è data dalla [34*].

§ 20. — Le corrispondenti funzioni di ripartizione. Operazioni $R\Phi$, $S\Phi$.

Indichiamo $R\Phi$, $S\Phi$ le operazioni che trasformano la funzione di ripartizione corrispondente a ψ in quella corrispondente a $R\psi$ o a $S\psi$.

Derivando la [18], $\psi(t) = \int_0^1 e^{it\xi}\, d\Phi$, si ha

[35] $$D\psi(t) = i\int_0^1 e^{it\xi}\,\xi\, d\Phi$$

e in generale

[36] $$D^r(i - D)^s\, \psi(t) = i^{r+s}\int_0^1 e^{it\xi}\,\xi^r(1 - \xi)^s\, d\Phi$$

[37] $$D^r(i - D)^s\, \psi(0) = i^{r+s}\int_0^1 \xi^r(1 - \xi)^s\, d\Phi$$

[38] $$R^r S^s\, \psi(t) = \frac{\int_0^1 e^{it\xi}\,\xi^r(1-\xi)^s\, d\Phi}{\int_0^1 \xi^r(1-\xi)^s\, d\Phi} = \int_0^1 e^{it\xi}\, d\left\{\frac{\int_0^{\Phi(\xi)} \xi^r(1-\xi)^s\, d\Phi}{\int_0^1 \xi^r(1-\xi)^s\, d\Phi}\right\}$$

Ciò mostra che la funzione di ripartizione entro $\Big\{\ \Big\}$ ha come funzione caratteristica $R^r S^s \psi$, e quindi abbiamo

$$[39] \qquad R^r_\Phi S^s_\Phi \Phi(\xi) = \frac{\int_0^{\Phi(\xi)} \xi^r (1-\xi)^s \, d\Phi}{\int_0^1 \xi^r (1-\xi)^s \, d\Phi}$$

o integrando per parti

$$[40] \qquad R^r_\Phi S^s_\Phi \Phi(\xi) = \frac{\xi^r (1-\xi)^s \Phi(\xi) - \int_0^\xi [r-(r+s)\lambda]\, \lambda^{r-1} (1-\lambda)^{s-1} \Phi(\lambda)\, d\lambda}{0^s - \int_0^1 [r-(r+s)\lambda]\, \lambda^{r-1} (1-\lambda)^{s-1} \Phi(\lambda)\, d\lambda}$$

dove si conviene che $0^s = 1$ per $s = 0$, mentre per $s > 0$, $0^s = 0$. In particolare

$$[40'] \qquad R_\Phi \Phi(\xi) = \frac{\int_0^{\Phi(\xi)} \xi \, d\Phi}{\int_0^1 \xi \, d\Phi} = \frac{\xi \Phi(\xi) - \int_0^\xi \Phi(\lambda)\, d\lambda}{1 - \int_0^1 \Phi(\lambda)\, d\lambda}$$

$$[40''] \qquad S_\Phi \Phi(\xi) = \frac{\int_0^{\Phi(\xi)} (1-\xi)\, d\Phi}{\int_0^1 (1-\xi)\, d\Phi} = \frac{(1-\xi)\Phi(\xi) + \int_0^\xi \Phi(\lambda)\, d\lambda}{\int_0^1 \Phi(\lambda)\, d\lambda}$$

§ 21. — Una proprietà caratteristica di $R^r_\Phi S^s_\Phi \Phi$.

Applicando il teorema della media si ha dalla [39]

$$[41] \qquad \Big[R^r_\Phi S^s_\Phi \Phi\Big]_{\xi_1}^{\xi_2} = \frac{\bar{\xi}^r (1-\bar{\xi})^s}{\int_0^1 \xi^r (1-\xi)^s \, d\Phi} [\Phi]_{\xi_1}^{\xi_2} \qquad \text{con } \xi_1 \leq \bar{\xi} \leq \xi_2$$

e indicando $[F]_{\xi_1}^{\xi_2} = F(\xi_2) - F(\xi_1)$. Otteniamo così una forma più espressiva di scrivere il precedente risultato.

È facile anzi vedere che tale proprietà caratterizza completamente $R^r_\Phi\, S^s_\Phi\, \Phi$; basta osservare, più in generale, che se f è una qualsiasi funzione continua e Φ una qualunque funzione a variazione limitata, rimane univocamente determinata (a meno d'una costante additiva arbitraria) la funzione F che sodisfa la relazione

$$[F]_{\xi_1}^{\xi_2} = \overline{f(\xi)}\,[\Phi]_{\xi_1}^{\xi_2}$$

ove $\overline{f(\xi)}$ indica un valore compreso fra i limiti inferiore e superiore della funzione f nell'intervallo (ξ_1, ξ_2). Nelle condizioni specificate esiste infatti l'integrale di STIELTJES $\int f(\xi)\, d\,\Phi(\xi)$, per la definizione del quale è

$$F(\xi) = \int_0^{\Phi(\xi)} f(\lambda)\, d\,\Phi(\lambda) + C.$$

Nel nostro caso la costante C è determinata dovendo essere $R^r_\Phi\, S^s_\Phi\, \Phi(\xi) = 0$ per $\xi < 0$, e rimane così dimostrata la proprietà in questione.

La [41] ci sarà utile per dedurne un importante teorema asintotico, che vedremo nel Cap. III (§ 27).

§ 22. — Applicazioni ai casi del § 12.

Come esempio, applichiamo i risultati trovati alle funzioni caratteristiche considerate nel § 12.

Se $\psi(t) = e^{ipt}$, si ha $K\psi(t) = e^{i(1-p)t}$, $R\psi(t) = S\psi(t) = R^r S^s \psi(t) = e^{ipt}$. Non si hanno altre funzioni caratteristiche che rimangano invariate conoscendo l'esito di una prova: se $R\psi = \psi$ o $S\psi = \psi$ scende che $\psi(t) = e^{ipt}$ $(0 \leq p \leq 1)$. Infatti $R\psi = \psi$ significa $D\psi = c\,\psi$, $\psi(t) = a e^{ct}$, $\psi(0) = a = 1$, $\psi(t) = e^{ct}$ e quindi $\psi(t) = e^{ipt}$ $(0 \leq p \leq 1)$. Analogamente per $S\psi$; anche da $R^r\psi = \psi$ o $S^s\psi = \psi$ scende lo stesso risultato. Invece l'equazione $R^r S^s \psi = \psi$ ammette inoltre tutte e sole le soluzioni del tipo

$$\psi(t) = \lambda_1 e^{ip_1 t} + \lambda_2 e^{ip_2 t} \quad (\lambda_1 + \lambda_2 = 1,\ \lambda_1 > 0,\ \lambda_2 > 0,\ 0 \leq p_1 < p_2 \leq 1,$$

$p_1^r(1-p_1)^s = p_2^r(1-p_2)^s$). Ciò riesce intuitivo se si guardi l'espressione di $R^r_\Phi\, S^s_\Phi\, \Phi$.

Se $$\psi(t) = \frac{e^{it}-1}{it}: \qquad K\psi(t) = \frac{e^{it}-1}{it},$$

$$R\psi(t) = \frac{2}{t^2}(-it\,e^{it} + e^{it} - 1) \quad , \quad S\psi(t) = \frac{2}{t^2}(-e^{it} + 1 + it)$$

$$\begin{bmatrix} h \\ n \end{bmatrix} R^r S^s \psi = \frac{\binom{h+r}{r}\binom{n-h+s}{s}}{\binom{n+r+s}{n}} \cdot \frac{r+s+1}{n+r+s+1}.$$

In particolare la probabilità che l' $(r+s+1)^{esima}$ prova sia favorevole dopo r prove favorevoli ed s sfavorevoli, è

$$\begin{bmatrix}1\\1\end{bmatrix} R^r\, S^s\, \psi = \frac{r+1}{r+s+2}.$$

È questa la formula di cui si fece uso e anche abuso nella teoria delle probabilità a posteriori. Essa è rigorosamente esatta quando sia $\psi(t) = \dfrac{e^{it}-1}{it}$, ma solo in tale ipotesi.

§ 23. — Coverificarsi di fenomeni aleatori.

Se si hanno due fenomeni (analogamente per tre o più), indipendenti tra loro, le cui funzioni caratteristiche sono

$$\psi'(t) = \sum_0^\infty {}_h\, a_h \frac{i^h t^h}{h!} \quad , \quad \psi''(t) = \sum_0^\infty {}_h\, b_h \frac{i^h t^h}{h!},$$

il loro coverificarsi è ancora un fenomeno aleatorio e la sua funzione caratteristica è

$$\psi(t) = \sum_0^\infty {}_h\, a_h\, b_h \frac{i^h t^h}{h!} \tag{42}$$

Infatti a_h è la probabilità che il primo fenomeno si verifichi sempre su h prove, b_h l'analoga per il secondo, e di conseguenza $a_h\, b_h$ è la probabilità che su h prove si verifichino sempre entrambi.

In particolare se $\psi''(t) = e^{ipt}$ (fenomeno a probabilità nota p) si ha $\psi(t) = \psi'(pt)$;

se $\psi''(t) = \dfrac{1}{it}(e^{it}-1)$ è $\qquad \psi(t) = \dfrac{1}{t}\displaystyle\int_0^t \psi'(t)\, dt.$

Ad esempio per

$$\psi'(t) = e^{ipt} \ , \ \psi''(t) = \frac{1}{it}(e^{it}-1) \ : \ \psi(t) = \frac{1}{ipt}(e^{ipt}-1).$$

Sia infatti $\quad \psi''(t) = e^{ipt} = \displaystyle\sum_0^\infty {}_h\, p^h \frac{i^h t^h}{h!} \quad , \quad \psi'(t) = \sum_0^\infty {}_h\, a_h \frac{i^h t^h}{h!}:$

è $\qquad \psi(t) = \displaystyle\sum_0^\infty {}_h\, a_h\, p^h \frac{i^h t^h}{h!} = \sum_0^\infty a_h \frac{i^h (pt)^h}{h!} = \psi'(pt).$ Se invece

$$\psi''(t) = \frac{1}{it}(e^{it}-1) = \sum_0^\infty {}_h \frac{1}{h+1}\cdot\frac{i^h t^h}{h!} \quad , \quad \psi'(t) = \sum_0^\infty {}_h\, a_h \frac{i^h t^h}{h!}.$$

$$\psi(t) = \sum_0^\infty {}_h \frac{1}{h+1}\, a_h \frac{1}{h!} = \frac{1}{t}\sum_0^\infty {}_h\, a_h \frac{i^h t^{h+1}}{(h+1)!} = \frac{1}{t}\int_0^t \psi'(t)\, dt.$$

Se il secondo fenomeno non è indipendente dal primo, la formula (42) sussisterà naturalmente purchè $\psi''(t)$ indichi la funzione caratteristica del 2° fenomeno subordinatamente al verificarsi del 1° (o viceversa).

§ 24. — Fenomeni imputabili a più cause.

Se un fenomeno può dipendere da diverse cause (incompatibili) che hanno rispettivamente le probabilità $\lambda_1\ \lambda_2 \ldots \lambda_m$ $(\lambda_1 + \lambda_2 + \ldots + \lambda_m = 1)$, e nelle diverse ipotesi il fenomeno ha rispettivamente le funzioni caratteristiche $\psi^{(1)}(t)$, $\psi^{(2)}(t), \ldots, \psi^{(m)}(t)$, la funzione caratteristica del fenomeno aleatorio è

$$\psi(t) = \lambda_1\, \psi^{(1)}(t) + \lambda_2\, \psi^{(2)}(t) + \ldots + \lambda_m\, \psi^{(m)}(t) \tag{43}$$

Infatti la probabilità che h su n prove siano favorevoli è uguale alla probabilità della prima ipotesi per la probabilità che, se l'ipotesi vera è la prima, h delle n prove siano favorevoli, più i termini analoghi per la 2^a, $3^a, \ldots, m^a$ ipotesi.

Ad esempio l'identità

$$\psi(t) = i^{-1} D\, \psi(t) + i^{-1}(i - D)\, \psi(t)$$

essendo $i^{-1} D\, \psi(0)$ e $i^{-1}(i - D)\, \psi(0)$ rispettivamente la probabilità che la prima prova sia favorevole o sfavorevole, e

$$\frac{D\,\psi(t)}{D\,\psi(t)} = R\,\psi(t) \quad , \quad \frac{(i - D)\,\psi(t)}{(i - D)\,\psi(0)} = S\,\psi(t) \quad ,$$

esprime la funzione caratteristica di un fenomeno aleatorio scomposta relativamente alle due ipotesi che la prima prova sia favorevole, rispettivamente sfavorevole, e facendo quindi intervenire le funzioni caratteristiche $R\,\psi$, $S\,\psi$ dello stesso fenomeno subordinatamente a queste due ipotesi. Analogamente sviluppando $i^{-n}\,[D + (i - D)]^n$ si ottiene la scomposizione di ψ relativamente alle $n + 1$ ipotesi che sulle prime n prove quelle favorevoli siano $0, 1, \ldots, n$. Posto

$$\omega_h^{(n)} = \begin{bmatrix} h \\ n \end{bmatrix} \psi = i^{-n}\, D^h\, (i - D)^{n-h}\, \psi(0) \quad :$$

$$\psi(t) = \omega_0^{(n)}\, S^n\, \psi(t) + \omega_1^{(n)}\, R\, S^{n-1}\, \psi(t) + \omega_2^{(n)}\, R^2\, S^{n-2}\, \psi(t) + \ldots + \\ + \omega_{n-1}^{(n)}\, R^{n-1}\, S\, \psi(t) + \omega_n^{(n)}\, R^n\, \psi(t).$$

L'esempio classico cui si applica tale risultato è quello delle estrazioni da un'urna che è stata scelta a caso in una collezione nota. Se sappiamo che le percentuali di palle nere possono essere $p_1\ p_2 \ldots p_m$ colle probabilità $\lambda_1\ \lambda_2 \ldots \lambda_m$, la funzione caratteristica sarà

$$\psi(t) = \lambda_1\, e^{i p_1 t} + \lambda_2\, e^{i p_2 t} + \ldots + \lambda_m\, e^{i p_m t}\,.$$

Capitolo Terzo.

Probabilità a posteriori e probabilità delle ipotesi.

§ 25. La probabilità e l'esperienza.

Studiare la parte che può avere, nella valutazione di una probabilità, la conoscenza di dati desunti dall'esperienza, è un problema di importanza pratica preminente, perchè nei casi pratici, in cui manca per solito la suddivisione in «eventi ugualmente possibili», ogni previsione si basa essenzialmente sul confronto di casi analoghi gia verificati. La distinzione che si fa ordinariamente fra probabilità matematica e probabilità empirica, probabilità a priori e probabilità a posteriori, è però artificiosa, e corrisponde piuttosto a caratteri esteriori e ad esigenze d'indole pratica che a motivi di significato concettuale. In sostanza, la probabilità, che si suole chiamare *a posteriori*, di un certo evento E, valutata in base alla conoscenza di un certo gruppo di circostanze A, non è altro che il rapporto delle due probabilità, cosiddette *a priori*, del coverificarsi di E e A, e del verificarsi di A.

Abbiamo visto ad esempio come s'interpreti la formula [34*] della probabilità che un certo fenomeno aleatorio si verifichi quando se ne conosca la frequenza su un certo numero di prove già eseguite. Dopo r prove favorevoli ed s sfavorevoli, la probabilità che l'$(r+s+1)^{esima}$ prova risulti favorevole è il rapporto fra la probabilità che delle prime $r+s+1$ prove quelle favorevoli siano $r+1$, e tra esse l'ultima, e la probabilità che sulle prime $r+s$ prove quelle favorevoli siano r.

In quale senso potrà avere allora fondamento la persuasione comune che si possano fare previsioni basandosi sull'analogia coi fatti già osservati, e che, nel caso di un fenomeno aleatorio, conduce ad assumere la frequenza osservata come valore (sia pure approssimato) della probabilità?

È questo il problema generale, non certo nuovo, che nella trattazione di questo Capitolo tenteremo mettere in nuova luce.

§ 26. — Probabilità a posteriori.

Il problema delle probabilità a posteriori consiste nel cercare di determinare la probabilità d'un fenomeno aleatorio in base all'osservazione della frequenza effettivamente constatata in un certo numero di prove.

Nel caso di un fenomeno aleatorio di cui è nota la funzione caratteristica $\psi(t)$, il problema è pienamente determinato dalla formula [34], che dà in particolare la seguente espressione della probabilità del fenomeno dopo r prove favorevoli ed s sfavorevoli

$$[34^*] \qquad \begin{bmatrix} 1 \\ 1 \end{bmatrix} R^r S^s \psi = \frac{r+s+1}{r+1} \cdot \frac{\begin{bmatrix} r+1 \\ r+s+1 \end{bmatrix} \psi}{\begin{bmatrix} r \\ r+s \end{bmatrix} \psi}$$

che si può scrivere mediante Φ

$$[44] \qquad \begin{bmatrix}1\\1\end{bmatrix} R^r S^s \Phi = \int_0^1 \xi \, d \left\{ \frac{\int_0^{\Phi(\xi)} \xi^r (1-\xi)^s \, d\Phi}{\int_0^1 \xi^r (1-\xi)^s \, d\Phi} \right\}.$$

Quello che può interessare qui è invece di vedere se e in quale senso sia lecito affermare che la frequenza, cioè $\frac{r}{r+s}$, è un valore, sia pure approssimato, di tale probabilità, e cioè se e in quale senso un tentativo di inversione del teorema di BERNOULLI potrebbe avere fondamento.

§ 27. — Teorema asintotico sulle probabilità a posteriori.

Il teorema che dimostreremo è, in termini precisi, il seguente:

$$\text{posto } f = \frac{r}{r+s},$$

se per nessun $\varepsilon \neq 0$ è $\Phi(f+\varepsilon) = \Phi(f)$,
è sempre

$$[45] \qquad \lim_{n=\infty} \left(R_\Phi^r S_\Phi^s\right)^n \Phi(\xi) = 0 \text{ o } 1$$

rispettivamente se $\xi < f$ o $\xi > f$; per la corrispondente funzione caratteristica

$$[46] \qquad \lim_{n=\infty} \left(R^r S^s\right)^n \psi(t) = e^{ift},$$

e la tendenza al limite è uniforme in ogni regione finita.

Il teorema si può esprimere dicendo che: qualunque sia la natura di un fenomeno aleatorio, è sempre possibile determinare N (nel nostro caso, $n(r+s) = N$) così grande che, dopo N prove in cui la frequenza abbia un valore prefissato f, esso si possa assimilare con quel grado d'approssimazione che si voglia a un fenomeno che avesse probabilità costante nota a priori e uguale ad f. Ciò ad eccezione del caso che la funzione di ripartizione Φ del fenomeno aleatorio sia costante in un intorno del valore f, a meno cioè che non avesse probabilità nulla, a priori, l'ipotesi che la frequenza su un numero grandissimo di prove cadesse in un intorno di f (1).

In altro modo meno vago: la probabilità che, nell'ipotesi che sulle prime N prove la frequenza osservata risulti uguale ad f, in quelle successive la frequenza tenda a un limite che differisca da f per più di un dato ε, può rendersi piccola quanto si vuole pur di prendere N sufficientemente grande (sempre: per ogni fenomeno aleatorio, e colla restrizione accennata).

La dimostrazione scende immediatamente dalla [41] del § 21

(1) Cfr. il già citato *Probabilismo*, n. 22.

$$\left[R^r_\Phi S^s_\Phi \Phi\right]_{\xi_1}^{\xi_2} = \frac{\bar\xi^r (1-\bar\xi)^s}{\int_0^1 \xi^r (1-\xi)^s \, d\Phi} \, [\Phi]_{\xi_2}^{\xi_1}$$

Considerando due intervalli (ξ_1, ξ_2) e (ξ_3, ξ_4) si ha (purchè $[\Phi]_{\xi_3}^{\xi_4} \neq 0$):

$$\frac{\left[R^r_\Phi S^s_\Phi \Phi\right]_{\xi_1}^{\xi_2}}{\left[R^r_\Phi S^s_\Phi \Phi\right]_{\xi_3}^{\xi_4}} = \frac{\bar\xi^r (1-\bar\xi)^s}{\bar{\bar\xi}^r (1-\bar{\bar\xi})^s} \; \frac{[\Phi]_{\xi_1}^{\xi_2}}{[\Phi]_{\xi_3}^{\xi_4}} \tag{47}$$

(ove $\bar\xi$ è compreso fra ξ_1 e ξ_2, $\bar{\bar\xi}$ fra ξ_3 e ξ_4). Poniamo in particolare, scelto comunque $\varepsilon > 0$, $\xi_1 = 0$, $\xi_2 = f - \varepsilon$, $\xi_3 = f - \frac{1}{2}\varepsilon$, $\xi_4 = f$, ove $f = \frac{r}{r+s}$.
Avremo

$$0 \leq \bar\xi \leq f - \varepsilon \quad , \quad \bar\xi^r (1-\bar\xi)^s \leq (f-\varepsilon)^r (1-f+\varepsilon)^s$$

$$f - \frac{1}{2}\varepsilon \leq \bar{\bar\xi} \leq f \quad , \quad \bar{\bar\xi}^r (1-\bar{\bar\xi})^s \geq (f - \frac{1}{2}\varepsilon)^r (1 - f + \frac{1}{2}\varepsilon)^s$$

$$\frac{(f-\varepsilon)^r (1-f+\varepsilon)^s}{(f-\frac{1}{2}\varepsilon)^r (1-f+\frac{1}{2}\varepsilon)^s} = \theta < 1.$$

Indichiamo brevemente p_n l'espressione

$$\left[\left(R^r_\Phi S^s_\Phi\right)^n \Phi\right]_0^{f-\varepsilon} = \left(R^r_\Phi S^s_\Phi\right)^n \Phi(f-\varepsilon) \;;$$

osserviamo che certo $\left[\left(R^r_\Phi S^s_\Phi\right)^n \Phi\right]_{\xi_3}^{\xi_4} \leq 1$; ricordando che è per ipotesi

$$\left[\Phi\right]_{\xi_3}^{\xi_4} = \Phi(f) - \Phi(f - \tfrac{1}{2}\varepsilon) \neq 0$$

e posto quindi

$$\frac{[\Phi]_{\xi_1}^{\xi_2}}{[\Phi]_{\xi_3}^{\xi_4}} = a \quad , \quad \text{avremo} \quad p_n \leq \theta^n a$$

e quindi $(\theta < 1)$ $\lim\limits_{n=\infty} p_n = 0$, ossia

$$\lim_{n=\infty} \left(R^r_\Phi S^s_\Phi\right)^n \Phi(f-\varepsilon) = 0.$$

Analogamente

$$\lim_{n=\infty} \left(R^r_\Phi S^s_\Phi\right)^n \Phi(f+\varepsilon) = 1.$$

La funzione di ripartizione $\left(R^r_\Phi\ S^s_\Phi\right)^n \Phi$ tende, al crescere di n, salvo forse eventualmente per $\xi = f$, e per Φ qualunque purchè non costante in un intorno di $f\left(f = \dfrac{r}{r+s}\right)$, alla funzione di ripartizione di un fenomeno aleatorio che abbia probabilità costante nota a priori eguale ad f. Per un teorema noto (¹) ne scende che anche la funzione caratteristica $R^r\ S^s\ \psi(t)$ tende uniformemente a e^{ift} , c. v. d.

§ 28. — Sull'inversione del teorema di Bernoulli.

Il teorema del precedente § contiene tutto quello che vi può essere di preciso in un tentativo di inversione del teorema asintotico di BERNOULLI. Dopo un numero indefinitamente crescente di prove, la probabilità di un fenomeno aleatorio tende a divenire uguale alla frequenza (colla restrizione detta a suo luogo). Ma la convergenza non è uniforme per tutte le funzioni caratteristiche, e quindi, per quanto grande sia il numero delle prove già eseguite, non è possibile dedurne che la probabilità sia anche approssimativamente uguale alla frequenza senza conoscere quale fosse a priori la funzione caratteristica del fenomeno. Possiamo dire però che col crescere del numero delle prove divengono sempre meno restrittive le condizioni che si debbono supporre verificate dalla funzione caratteristica del fenomeno aleatorio perchè ne consegua che l'eguaglianza approssimata fra probabilità e frequenza sussista.

Se si vuole, si potrà dire che: quanto maggiore è il numero delle prove, tanto più *è facile* che la probabilità e la frequenza siano approssimativamente uguali. Ma alla frase « *è facile* » non si può dare altro significato che quello precedentemente ben chiarito. In nessun modo si può ritenerlo sinonimo di « *è probabile* », e tentare di tradurre il « *tanto più è facile* » in un'affermazione precisa come « *si può rendere la probabilità comunque prossima ad uno* ».

Ciò parrebbe, prima ancora che errato, privo addirittura di senso.

§ 29. — Probabilità delle cause.

L'altro caso particolare in cui studieremo i rapporti fra probabilità ed esperienza è il problema delle probabilità delle ipotesi o delle cause. Esso consiste nel ricercare la probabilità di un'ipotesi (o « causa ») il cui verificarsi modifica la funzione caratteristica di un fenomeno aleatorio, in base all'osservazione della frequenza con cui il fenomeno s'è verificato in un certo numero di prove.

L'impostazione è ovvia. Se ψ è la funzione caratteristica del fenomeno aleatorio, ψ' la funzione caratteristica dello stesso fenomeno subordinatamente a una certa ipotesi, e la probabilità di quest'ipotesi è p, potremo subito determinare la probabilita x che avrà l'ipotesi dopo che su $r+s$ prove se ne siano presentate r di favorevoli. Infatti la probabilità che r delle prime $r+s$ prove siano favorevoli e che l'ipotesi sia vera è uguale tanto al prodotto della probabilità p dell'ipotesi per la probabilità $\left[\begin{matrix} r \\ r+s \end{matrix}\right]\psi'$ che, subordinatamente, il fenomeno si

(¹) V. LEVY, oppure CASTELNUOVO, *Op. cit.*, Vol. II, Appendice, Art. II, n. 21.

verifichi in r prove su $r+s$, quanto al prodotto della probabilità $\left[\begin{matrix} r \\ r+s \end{matrix}\right]\psi$ che delle prime $r+s$ prove quelle favorevoli siano r per la probabilità x che, in tal caso, l'ipotesi considerata sia vera. E si ha quindi

[48] $$x \doteq p \frac{\left[\begin{matrix} r \\ r+s \end{matrix}\right]\psi'}{\left[\begin{matrix} r \\ r+s \end{matrix}\right]\psi}.$$

Si vede di qui che un problema di probabilità delle ipotesi è pienamente determinato quando e soltanto quando: è noto il fenomeno (è data la sua funzione caratteristica), è nota la precisa influenza dell'ipotesi (è data la funzione caratteristica del fenomeno subordinatamente all'ipotesi), e è nota la probabilità a priori dell'ipotesi stessa.

Altrimenti non ha senso.

§ 30. — Un esempio.

Un esempio che, pur riferendosi, per fissare le idee, all'abusato schema delle estrazioni da un'urna, si avvicina discretamente al tipo dei problemi che si possono presentare nella pratica, è questo. Abbiamo un'urna A contenente n palle tra bianche e nere, che vi sono state immerse da un individuo il quale aveva a sua disposizione $N = c\,n$ palle (c intero maggiore di 1), di cui $H = c\,h$ bianche e $K = c\,k$ nere ($H + K = N$, ossia $h + k = n$). Di tutte le ipotesi possibili riteniamo che soltanto le due seguenti possano essersi verificate:

a) le n palle sono state scelte a caso fra le N disponibili;

b) l'individuo che preparò l'urna ebbe cura di scegliere n palle in modo da conservare le percentuali di palle bianche e nere (e quindi prendendo h palle bianche e k nere).

Conosciamo ancora la probabilità che hanno queste due ipotesi, siano α e β. Il fenomeno consistente nell'estrazione di una palla bianca dall'urna A ha (Cfr. § 24) la funzione caratteristica

$$\psi(t) = \alpha\,\psi^{(\alpha)}(t) + \beta\,\psi^{(\beta)}(t)$$

ove

$$\psi^{(\beta)}(t) = e^{i\frac{h}{n}t}\quad,\quad \psi^{(\alpha)}(t) = \frac{1}{\binom{N}{n}}\sum_0^n{}_l \binom{H}{l}\binom{N-H}{n-l} e^{i\frac{l}{n}t}$$

(essendo $\binom{H}{l}\binom{N-H}{n-l} : \binom{N}{n}$ la probabilità che delle n palle estratte a sorte e immesse nell'urna l siano bianche).

Dopo $r+s$ estrazioni che diedero r palle bianche ed s nere, la funzione caratteristica è

$$R^r S^s \psi(t) = \frac{1}{\left[\begin{matrix} r \\ r+s \end{matrix}\right]\psi}\left\{\alpha.\left[\begin{matrix} r \\ r+s \end{matrix}\right]\psi^{(\alpha)}.\,R^r S^s \psi^{(\alpha)}(t) + \beta.\left[\begin{matrix} r \\ r+s \end{matrix}\right]\psi^{(\beta)}.\,R^r S^s \psi^{(\beta)}(t)\right\}.$$

Derivando si ha, per $t = 0$, la probabilità di ottenere, all' $(r + s + 1)^{\text{esima}}$ estrazione, una palla bianca (determinazione di una probabilità a posteriori); la probabilità che le palle bianche siano state scelte a caso (ipotesi a) dopo r estrazioni di palle bianche e s di palle nere è

$$\alpha . \frac{\begin{bmatrix} r \\ r+s \end{bmatrix} \psi^{(\alpha)}}{\begin{bmatrix} r \\ r+s \end{bmatrix} \psi}$$

(determinazione di una probabilità delle ipotesi).

Per fare un'applicazione numerica, se l'urna contiene 6 palle scelte fra 12, che si avevano a disposizione, di cui 4 bianche e 8 nere, e si ritiene $\alpha = 2/3$, $\beta = 1/3$:

$$\psi^{(\beta)}(t) = e^{\frac{it}{3}}, \quad \psi^{(\alpha)}(t) = \frac{1}{33}\left\{1 + 8\,e^{\frac{it}{6}} + 15\,e^{\frac{it}{3}} + 8\,e^{\frac{it}{2}} + e^{\frac{2it}{3}}\right\},$$

$$\psi(t) = \frac{1}{99}\left\{2 + 16\,e^{\frac{it}{6}} + 63\,e^{\frac{it}{3}} + 16\,e^{\frac{it}{2}} + 2\,e^{\frac{2it}{3}}\right\}.$$

Dopo $r + s$ estrazioni, di cui r diedero palle bianche, la probabilità dell'ipotesi b) (scelta non a caso) è

$$\frac{33 \,.\, 2^r \,.\, 4^s}{16 \,.\, 5^s + 63 \,.\, 2^r \,.\, 4^s + 16 \,.\, 3^{r+s} + 2 \,.\, 4^r \,.\, 2^s}\,.$$

Dopo 6 estrazioni l'ipotesi b) ha le probabilità

0,088.353	**se si ebbero $r = 6$ palle bianche,**
0,176.707	**se $r = 5$,**
0,279.365	**se $r = 4$,**
0,359.918	**se $r = 3$,**
0,389.812	**se $r = 2$,**
0,353.947	**se $r = 1$, e finalmente**
0,260.018	**se $r = 0$, ossia se si estrassero solo palle nere.**

Al crescere del numero delle prove, la probabilità dell'ipotesi b) tende rispettivamente a $^{53}/_{63}$, $^{53}/_{79}$ o a zero a seconda che la frequenza è compresa tra

$$(\log\ ^5/_4)\ /\ (\log\ ^5/_2) = 0{,}243.529 \quad \text{e}$$
$$(\log\ ^4/_3)\ /\ (\log\ 2) = 0{,}415.037,$$

è uguale a uno di questi due limiti, o è esterna.

CAPITOLO QUARTO.

Classi d'eventi equivalenti.

§ 31. Classi d'eventi equivalenti.

Siano $E_1 \dots E_n$ n eventi. Diremo che essi costituiscono una classe d'eventi equivalenti se tutte le loro combinazioni m ad m hanno uguale probabilità ($m = 0, 1, \dots, n$). Precisando: il fatto che degli n eventi se ne verifichino m può avvenire in $\binom{n}{m}$ modi distinti, quante sono le diverse classi di m eventi che si possono formare con $E_1 \dots E_n$. La condizione caratteristica delle classi d'eventi equivalenti è che tutte queste combinazioni abbiano uguale probabilità. Possiamo anche dire che gli eventi $E_1 \dots E_n$ sono equivalenti se la probabilità delle loro combinazioni non varia permutandoli comunque l'uno con l'altro.

È ovvio che ogni sottoclasse di una classe d'eventi equivalenti è classe d'eventi equivalenti.

Una classe infinita di eventi E si dirà classe d'eventi equivalenti se è classe d'eventi equivalenti qualunque sua sottoclasse finita.

Eventi equivalenti sono certo e a fortiori eventi ugualmente probabili; anche subordinatamente al verificarsi di uno o più tra essi i rimanenti sono ancora ugualmente probabili e costituiscono ancora una classe d'eventi equivalenti.

§ 32. Funzione caratteristica d'una classe d'eventi equivalenti.

Se si ha una classe di n eventi (equivalenti o no) il numero di quelli tra essi che si verificheranno è una variabile causale m capace di assumere soltanto i valori $0, 1, \dots, n$. Se si indicano $\omega_0, \omega_1, \dots, \omega_n$ le probabilità di questi $n+1$ casi possibili ($\omega_0 + \omega_1 + \dots + \omega_n = 1$), la funzione caratteristica della variabile casuale considerata sarà

$$\psi(t) = \sum_0^n {}_h\, \omega_h e^{iht}$$

Nel caso di una classe d'eventi equivalenti, la probabilità che gli eventi che si verificheranno siano m eventi assegnati sarà $\dfrac{\omega_m}{\binom{n}{m}}$, poichè per ipotesi tutte le $\binom{n}{m}$ classi di m eventi hanno uguale probabilità. Così varranno pure ancora tutte le altre conseguenze di tale ipotesi, e in particolare la relazione ricorrente [4], che ci darà le funzioni caratteristiche di una qualunque sottoclasse formata con parte degli n eventi equivalenti. Le sottoclassi costituite di m eventi avranno tutte la medesima funzione caratteristica $\psi_m(t)$ che si deduce mediante la [4] dalla $\psi_n(t) = \psi(t)$. Due sottoclassi costituite dello stesso numero (finito) m di eventi appartenenti a una classe d'eventi equivalenti hanno la stessa funzione caratteristica. Lo abbiamo constatato per il caso che appartengano a una classe finita di eventi equivalenti; il caso in cui la classe d'eventi può essere anche infinita è solo apparentemente più generale, perche due sottoclassi di m eventi appartengono sempre a una classe finita di non più di $2\,m$ eventi equivalenti.

Perciò, se si ha una classe infinita di eventi equivalenti, tutte le sue sottoclassi di m eventi hanno la medesima funzione caratteristica $\psi_m(t)$, e $\psi_m\left(\frac{t}{m}\right)$ tenderà a una funzione limite $\psi(t)$ quando $m \to \infty$. Potremo dire la $\psi(t)$, per definizione, funzione caratteristica della classe infinita d'eventi equivalenti.

E allora vale anche per le sottoclassi infinite di una classe d'eventi equivalenti la proprietà che esse avranno tutte la medesima funzione caratteristica.

§ 33. Campo di validità della precedente trattazione.

In tutta la precedente trattazione, svolta pel caso dei fenomeni aleatori, l'unica ipotesi che entra in gioco in modo essenziale è ovviamente soltanto quella d'avere una classe — potenzialmente illimitata — di eventi equivalenti. Il fatto che essi siano altrettante prove distinte di un medesimo fenomeno non ne è che un'interpretazione significativa.

La condizione che la classe non sia limitata non è peraltro necessaria se non per far sì che le funzioni caratteristiche ψ_m abbiano senso per m comunque grande; tutti i calcoli in cui intervengono le sole ψ_m con $m \leq n$ valgono anche nel caso di una classe di n eventi equivalenti.

Quindi: il campo di validità della precedente trattazione è quello delle classi d'eventi equivalenti; trattando di una classe di n eventi non si hanno naturalmente a considerare che le ψ_m con $m \leq n$; i procedimenti in cui m si fa crescere indefinitamente hanno senso soltanto nelle classi infinite.

§ 34. Classi infinite d'eventi equivalenti.

Nel caso di una classe infinita di eventi equivalenti valgono tutte, senza eccezione, le proprietà dei fenomeni aleatori: non avremo quindi che a passare in rassegna le più notevoli per facilitarne la traduzione pel nuovo caso.

La percentuale degli eventi veri per n di essi comunque scelti è una variabile casuale che tende a una legge limite quando n cresce indefinitamente. In una successione infinita di eventi, si ha probabilità 1 che la percentuale degli eventi veri tenda a un limite. e, nel caso che il limite esista, è una variabile casuale che obbedisce alla legge limite.

La classe degli eventi contrari (essa pure classe d'eventi equivalenti) ha la funzione caratteristica $K\psi$; quando si sa che r eventi della classe sono veri ed s falsi, ogni classe infinita formata coi rimanenti ha la funzione caratteristica $R^r S^s \psi$.

§ 35. Caso delle classi finite.

Un caso nuovo ci si può presentare soltanto per le classi finite.

Se si ha una classe di n eventi equivalenti, si è visto che tutti i risultati stabiliti valgono ancora purchè ci si limiti alle ψ_m con $m \leq n$. E si dovrà dire quindi ad esempio: la classe dei fenomeni contrari avrà la funzione caratteristica $\psi'_n\left(\frac{t}{n}\right) = K\psi_n\left(\frac{t}{n}\right)$; quando si sappia che r eventi sono veri, s falsi, la classe residua (di $m = n - r - s$ eventi) avrà la funzione caratteristica $\Omega'_m(e^{it})$ dove, posto (per $k = 0, 1, \ldots, n$) $\Omega_k(z) = \psi_k(-i \log z)$, $\Omega'_m(z)$ si ottiene applicando l'operazione [33]:

$$\Omega'_m(z) = U \sum_{0}^{s}{}_h (-1)^h \binom{s}{h} \frac{1}{(m+r+h)!} D^{r+h}\,\Omega_{m+r+h}(z);$$

e così di seguito.

Ma vi è una differenza essenziale che va studiata a parte.

È ovvio che ogni funzione del tipo $\sum_{0}^{n}{}_h \omega_h e^{iht}$ con $\omega_h \geqq 0$ e $\sum_{0}^{n}{}_h \omega_h = 1$ (tipo che indicheremo H_n^0) può essere la funzione caratteristica di una classe di n eventi equivalenti. È facile infatti costruire una classe di n eventi equivalenti in modo che le probabilità di vederne verificati $0, 1, \ldots, n$ assumano gli $n+1$ valori prefissi $\omega_0, \omega_1, \ldots, \omega_n$. Volendo un esempio, si pensi un'urna U in cui metteremo n palle tra bianche e nere. Il numero delle palle bianche sarà $0, 1, \ldots, n$ a seconda che in un 'esperienza preventiva ove si hanno $n+1$ casi possibili E_i si sia verificato $E_0, E_1, \ldots, E_n$. Se si scelgono gli E_i in modo che le loro probabilità siano $\omega_0 \ldots \omega_n$, le n successive estrazioni da U (senza rimettere le palle estratte) costituiscono una classe di n eventi equivalenti la cui funzione caratteristica è appunto $\sum_{0}^{n}{}_h \omega_h e^{iht}$.

Invece le possibili funzioni caratteristiche di una classe di n eventi che sia sottoclasse di una classe di $n+1$, $n+2, \ldots$ eventi equivalenti sono assoggettate a condizioni sempre più restrittive; le diremo brevemente funzioni caratteristiche del tipo $H_n^1, H_n^2, \ldots$ Per le sottoclassi di una classe infinita è necessario e sufficiente (¹) che siano soddisfatte le condizioni relative a una sottoclasse di classe finita, comunque grande sia il numero d'elementi di questa. Indicando H_n le funzioni caratteristiche di una classe di n eventi sottoclasse di classe infinita, abbiamo cioè che H_n è la parte comune (prodotto logico) di $H_n^0, H_n^1, H_n^2, \ldots$

È ovvio che le H_n^1 sono le funzioni che si deducono dalle H_{n+1}^0 applicando la relazione ricorrente [4'] tra le funzioni caratteristiche, le H_n^2 quelle che si deducono dalle H_{n+1}^1 e quindi dalle H_{n+2}^0, e così via.

Servirà da utile chiarimento un esempio evidente di funzione che può essere funzione caratteristica di una classe di n eventi equivalenti, ma soltanto se non sia sottoclasse di una classe di eventi equivalenti in numero maggiore. Essa è $\psi(t) = e^{iht}$ (dove h può essere $1, 2, \ldots, n-1$), e corrisponde alla certezza che le prove favorevoli siano h; vale quindi ad esempio nel caso che da un'urna contenente n palle di cui h nere, le si estraggano tutte una dopo l'altra. Ma la certezza $(\omega_h^{(n)} = 1)$ che su n prove quelle favorevoli siano h non

(¹) Quest'asserzione mi sembra ovvia, ma una dimostrazione rigorosa sarebbe forse desiderabile.

può aversi evidentemente se gli n eventi appartengono a una classe di eventi equivalenti in numero maggiore, essendo assurdo che tutte le sottoclassi contenenti n eventi possano contenerne esattamente h di verificati (a meno che per $h=0$, $h=n$; casi esclusi). Se da un'urna contenente più di n palle se ne estraggono n, non si può mai esser certi (quali si siano le probabilità delle diverse composizioni) che quelle nere siano h.

§ 36. — Integrazione dell'equazione fondamentale.

Il problema che siamo portati a considerare ci conduce pertanto allo studio dell'integrazione della [4], considerata come equazione differenziale in Ω_n.

L'integrale generale dell'equazione differenziale

$$(n+1)\,\Omega_n(z)=(n+1)\,\Omega_{n+1}(z+(1-z)\,D\,\Omega_{n+1}(z)$$

(lineare, di primo ordine, non omogenea) è

$$[49] \qquad \Omega_{n+1}(z)=(1-z)^{n+1}\left[(n+1)\int_0^z \Omega_n(z)\,.\,(1-z)^{-(n+2)}\,dz+C\right].$$

Perchè $\Omega_n(e^{it})$ sia una H_n^1 è necessario e sufficente che C si possa determinare in modo che i coefficenti di Ω_{n+1} risultino tutti positivi o *al più* nulli. Analogamente integrando più volte si otterrebbe un'espressione del tipo

$$[50] \quad \Omega_{n+h}(z)=F(z)+C_1(1-z)^{n+1}+C_2(1-z)^{n+2}+\ldots+C_h(1-z)^{n+h}$$

dove $F(z)$ è funzione lineare di $\Omega_n(z)$. Perchè $\Omega_n(e^{it})$ sia una H_n^h è necessario e sufficente che $C_1\ldots C_h$ si possano determinare in modo che i coefficienti di Ω_{n+h} risultino tutti positivi.

Per il caso considerato nel precedente §, $\Omega_n(z)=z^h$, l'impossibilità di passare, mediante un'integrazione, a Ω_{n+1} risulta direttamente dalla [2]:

$$\omega_h^{(n)}=\frac{n-h}{n+1}\,\omega_h^{(n+1)}+\frac{h+1}{n+1}\,\omega_{h+1}^{(n+1)}$$

che, essendo $\omega_h^{(n+1)}+\omega_{h+1}^{(n+1)}\leq 1$, dà, per $h=1,2,\ldots,n-1$:

$$\omega_h^{(n)}<1.$$

§ 37. — Sottoclasse finita di classe infinita.

Il caso più interessante e che vale la pena di studiare proseguendo il calcolo sino in fondo è quello delle sottoclassi finite di una classe infinita. Tale ricerca si riconnette poi allo studio dei fenomeni aleatori, perchè la condizione necessaria e sufficente perchè una funzione possa essere la funzione caratteristica di una classe di n prove di un fenomeno aleatorio è appunto quella di appartenere al tipo H_n.

Si può seguire per questo caso un metodo diretto.

Sappiamo che la condizione che caratterizza le funzioni caratteristiche di fenomeni aleatori (o, il che è indifferente, di classi infinite di eventi equivalenti) è che la corrispondente funzione di ripartizione $\Phi(\xi)$ sia nulla per $\xi < 0$ e $= 1$ per $\xi > 1$. Sappiamo che la conoscenza di ψ_n equivale alla conoscenza dei primi $n+1$ momenti $\omega_0^{(0)}, \omega_1^{(1)}, \ldots, \omega_n^{(n)}$ (cfr. la [10]). Si tratta dunque di determinare le condizioni necessarie e sufficenti perchè esista una distribuzione di masse tutta contenuta nell'intervallo (0,1) e che dia luogo agli $n+1$ momenti $\omega_0^{(0)}$, $\omega_1^{(1)}, \ldots, \omega_n^{(n)}$: ci si può basare a tal uopo sulla considerazione dei polinomi che intervengono nella risoluzione del problema dei momenti (CASTELNUOVO, *op. cit.*, V. II, *Appendice*, Art. II, n. 5 e sgg.).

Nel caso di n dispari $(n = 2k - 1)$, condizione necessaria e sufficiente perchè una funzione ψ_n possa essere funzione caratteristica di una sottoclasse di n tra infiniti eventi equivalenti, o anche di una classe di n prove di un fenomeno aleatorio, è che il polinomio

$$f(\xi) = \begin{vmatrix} 1 & \xi & \xi^2 & \ldots & \xi^k \\ \omega_0^{(0)} & \omega_1^{(1)} & \omega_2^{(2)} & \ldots & \omega_k^{(k)} \\ \omega_1^{(1)} & \omega_2^{(2)} & \omega_3^{(3)} & \ldots & \omega_{k+1}^{(k+1)} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \omega_{k-1}^{(k-1)} & \omega_k^{(k)} & \omega_{k+1}^{(k+1)} & \ldots & \omega_{2k-1}^{(2k-1)} \end{vmatrix}$$

abbia tutte le radici nell'intervallo (0,1) (estremi inclusi) (¹)

Il caso di n pari $(n = 2k)$ si può ricondurre al precedente osservando che, perchè esista una distribuzione di masse comprese nell'intervallo (0,1), e i cui primi $2k+1$ momenti siano $1, \omega_1^{(1)}, \ldots, \omega_{2k}^{(2k)}$,

a) è necessario che esista una distribuzione contenuta in (0,1) e i cui primi $2k$ momenti siano $1, \omega_1^{(1)}, \ldots, \omega_{2k-1}^{(2k-1)}$;

b) è necessario e sufficente che, almeno per un valore di $\omega_{2k+1}^{(2k+1)}$, esista una distribuzione contenuta in (0,1) e i cui primi $2k+2$ momenti siano 1, $\omega_1^{(1)}, \ldots, \omega_{2k}^{(2k)}, \omega_{2k+1}^{(2k+1)}$ (²).

(¹) Se la condizione non è sodisfatta, nessuna distribuzione di masse in numero finito avente i dati momenti è interna all'intervallo (0,1) (CASTELNUOVO, l. c., n. 8 : fra due radici di ω_n ve n'è una e una sola di ω_{n-2}). A una distribuzione di infinite masse si può sostituire coll'approssimazione che si vuole una distribuzione di masse in numero finito; eseguendo il passaggio al limite in modo opportuno si può escludere anche l'esistenza di una distribuzione di masse qualunque.

La condizione è quindi necessaria. Che sia sufficente è ovvio, perchè se le radici sono interne all'intervallo, esse stesse, prese con pesi opportuni, danno una soluzione del problema.

(²) Sfruttando la proprietà dimostrata nel § 3 della Sua Memoria *Sul problema dei momenti* (« Boll. Ist. It. Attuari », 1930, n. 2) dal CASTELNUOVO si vede facilmente che la condizione necessaria e sufficiente, nel caso di $n = 2k$, è l'analoga della precedente per $n = 2k - 1$ quando nel determinante si sostituisca ovunque ad $\omega_0^{(0)}, \omega_1^{(1)}, \ldots, \omega_{2k-1}^{(2k-1)}$ ordinatamente $\omega_1^{(1)}, \omega_2^{(2)}, \ldots, \omega_{2k}^{(2k)}$. Quella proprietà assicura infatti che se esistono dei gruppi dell'involuzione I_{k+1} (soluzioni del nostro problema al variare del momento $(2k+1)$-*esimo*) compresi in (0, 1), tale è certamente quello che contiene l'origine; cercando i punti coniugati dell'origine si arriva all'equazione considerata, che è la [19] del CASTELNUOVO, ossia la [9] per $\xi_1 = 0$.

APPENDICE.

Teorema d'esistenza per la legge limite.

1. — È noto che se la funzione caratteristica di una legge variabile tende uniformemente in qualunque intervallo limitato verso la funzione caratteristica di una legge fissa, anche la funzione di ripartizione della legge variabile tende alla funzione di ripartizione della legge fissa (n. 21) ([1]).

Rimaneva però da vedere se fosse effettivamente necessario postulare, come si fa in tale enunciato, l'esistenza di una legge limite corrispondente al limite della funzione caratteristica, o se non era possibile invece dimostrare che il fatto stesso dell'uniforme convergenza a un limite della funzione caratteristica implica necessariamente l'esistenza di una legge limite.

È appunto di tale teorema d'esistenza della legge limite, di notevole importanza di per sè, e per noi necessario nel § 8, che ci occuperemo in questo articolo; lo risolveremo pienamente, e in senso affermativo ([2]).

2. — Dall'Oss. al n. 18 scende ([3]) che, fissato comunque piccolo ε, ponendo $c = \frac{4}{\varepsilon}$, c risulta sufficentemente grande perchè ([4])

$$[1] \qquad \left| 2\int_{-\infty}^{\infty} \frac{sen\, x}{x}\left[\Phi\left(y+\frac{x}{a}\right)-\Phi\left(\frac{x}{a}\right)\right]dx - 2\int_{-c}^{+c}\frac{sen\, x}{x}\left[\Phi\left(y+\frac{x}{a}\right)-\Phi\left(\frac{x}{a}\right)\right]dx\right| < \varepsilon$$

quali si siano la funzione di ripartizione Φ e il numero a. Scelto in tal modo c, risulta pure che

$$[2] \qquad \left| 2\pi\left[\Phi(y)-\Phi(0)\right] - 2\int_{-c}^{+c}\frac{sen\, x}{x}\left[\Phi(y)-\Phi(0)\right]dx\right| < \varepsilon$$

e quindi, essendo

$$\Phi\left(y+\frac{x}{a}\right)-\Phi\left(\frac{x}{a}\right)=\left[\Phi\left(y+\frac{x}{a}\right)-\Phi(y)\right]-\left[\Phi\left(\frac{x}{a}\right)-\Phi(0)\right]+\left[\Phi(y)-\Phi(0)\right]$$

([1]) Il richiamo « n. 21 » e gli altri nel seguito si riferiscono ai numeri dell'Art. II in Appendice al *Calcolo delle Probabilità* del CASTELNUOVO. I richiami con «§» rimandano ai paragrafi di questa memoria.

([2]) Il LEVY sembra affermi a p. 202 tale risultato, e sembra ne ritenga superflua la dimostrazione, ma il suo ragionamento non mi pare sufficiente.

([3]) A rigore, scende solo per i c multipli di 2π, ma tale restrizione si toglie agevolmente.

([4]) Il simbolo $\int_{-\infty}^{+\infty}$ significherà sempre $\lim_{a=\infty}\int_{-a}^{+a}$ (valore principale secondo CAUCHY).

sarà ancora per qualunque Φ, come si ricava sommando

$$\left| 2\int_{-\infty}^{\infty} \frac{sen\, x}{x}\left[\Phi\left(y+\frac{x}{a}\right)-\Phi\left(\frac{x}{a}\right)\right]dx - 2\int_{-c}^{+c}\frac{sen\, x}{x}\left[\Phi\left(y+\frac{x}{a}\right)-\Phi(y)\right]dx + \right.$$

$$\left. +2\int_{-c}^{+c}\frac{sen\, x}{x}\left[\Phi\left(\frac{x}{a}\right)-\Phi(0)\right]-2\pi\left[\Phi(y)-\Phi(0)\right]\right| < 2\varepsilon$$

e finalmente

$$[3] \qquad \left| 2\pi\left[\Phi(y)-\Phi(0)\right] - 2\int_{-\infty}^{\infty}\frac{sen\, x}{x}\left[\Phi\left(y+\frac{x}{a}\right)-\Phi\left(\frac{x}{a}\right)\right]dx\right| <$$

$$< 2\varepsilon + \left|\int_{-c}^{+c}\frac{sen\, x}{x}\left[\Phi\left(y+\frac{x}{a}\right)-\Phi(y)\right]dx - \int_{-c}^{+c}\frac{sen\, x}{x}\left[\Phi\left(\frac{x}{a}\right)-\Phi(0)\right]dx\right|.$$

3. — Supponiamo ora che la funzione di ripartizione Φ ammetta in ogni punto densità limitata non superiore a un numero finito μ, così che si abbia (per ogni y e ogni h):

$$[4] \qquad |\,\Phi(y+h)-\Phi(y)\,| \leq |\,h\,|\,\mu.$$

Allora

$$\left|\int_{-c}^{+c}\frac{sen\, x}{x}\left[\Phi\left(y+\frac{x}{a}\right)-\Phi(y)\right]dx\right| < \frac{\mu}{a}\int_{-c}^{+c}|\,sen\, x\,|\,dx < \frac{2\mu c}{a} < \varepsilon$$

pur di prendere

$$a > \frac{2\mu c}{\varepsilon} = \frac{8\mu}{\varepsilon^2}.$$

Con ciò risulta che per ogni funzione di ripartizione Φ a densità limitata non superiore a μ

$$[5] \qquad \left| 2\pi\left[\Phi(y)-\Phi(0)\right] - 2\int_{-\infty}^{\infty}\frac{sen\, x}{x}\left[\Phi\left(y+\frac{x}{a}\right)-\Phi\left(\frac{x}{a}\right)\right]dx\right| < 4\varepsilon$$

purchè $a \geq \frac{8\mu}{\varepsilon^2}$. Ricordando che, se $\psi(t)$ è la funzione caratteristica della Φ si ha (n. 18).

$$2\int_{-\infty}^{\infty} \frac{sen\, x}{x} \left[\Phi\left(y + \frac{x}{a}\right) - \Phi\left(\frac{x}{a}\right)\right] d\,x = \int_{-a}^{+a} \frac{1 - e^{-iyt}}{i\,t}\, \psi(t)\, d\,t$$

risulta finalmente che nelle stesse ipotesi:

[6] $$\left| 2\pi [\Phi(y) - \Phi(0)] - \int_{-a}^{+a} \frac{1 - e^{-iyt}}{i\,t}\, \psi(t)\, d\,t \right| < 4\,\varepsilon.$$

4. — Se Φ_1 e Φ_2 sono due funzioni di ripartizione, entrambe a densità limitata non superiore a μ, abbiamo per differenza

[7] $$\begin{aligned} &| 2\pi [\Phi_1(y) - \Phi_2(y)] - 2\pi [\Phi_1(0) - \Phi_2(0)] | = \\ &= \left| 2\pi [\Phi_1(y) - \Phi_1(0)] - 2\pi [\Phi_2(y) - \Phi_2(0)] \right| < \\ &< 8\,\varepsilon + \left| \int_{-a}^{+a} \frac{1 - e^{-iyt}}{i\,t} [\psi_1(t) - \psi_2(t)]\, d\,t \right|. \end{aligned}$$

Sia ora $\Phi_1\, \Phi_2 \ldots \Phi_h \ldots$ una successione di funzioni di ripartizione le cui funzioni caratteristiche $\psi_1\, \psi_2 \ldots \psi_h \ldots$ tendano uniformemente in ogni regione finita dell'asse reale a una funzione limite ψ: si possa cioè, assegnati comunque θ e a, determinare N tale che per $n > N$ sia sempre

[8] $$| \psi(t) - \psi_n(t) | < \theta$$
per t reale e $| t | < a$.

Consideriamo dapprima il caso in cui tutte le Φ_h abbiano densità limitata non superiore a un numero finito μ. Fissati ε e θ, scelto $a > \frac{8\,\mu}{\varepsilon^2}$, e successivamente N in modo che, per tali valori di a e di θ, sussista la [8], si ha che, per $n > N$, $m > N$, è pure

[9] $$| \psi_n(t) - \psi_m(t) | < 2\,\theta$$
in tutto il segmento reale $-a < t < a$,

e la [7] è soddisfatta dalle due funzioni Φ_n e Φ_m:

[10] $$\left| 2\pi [\Phi_n(y) - \Phi_n(0)] \quad 2\pi [\Phi_m(y) - \Phi_m(0)] \right| < 8\,\varepsilon + 2\theta \int_{-a}^{a} \left| \frac{1 - e^{-iyt}}{i\,t} \right| dt.$$

Il secondo membro si può rendere piccolo ad arbitrio fissando opportunamente prima ε e poi θ; ciò basta a provare che, per un dato y, la successione $\Phi_n(y) - \Phi_n(0)$ tende, al crescere di n, a un limite che sarà certo funzione non decrescente di y.

5. — Esso tenderà a sua volta, per $y \to \pm\infty$, a due limiti la cui differenza non potrà superare 1, e precisamente è uguale o minore di 1 a seconda che $\Phi_n(y) - \Phi_n(0)$ tende al limite in modo uniforme per ogni y oppure no. Dimostriamolo. Se la tendenza al limite è uniforme, fissato comunque ε, scelto n in modo che, qualunque sia y, $\Phi_n(y) - \Phi_n(0)$ differisca dal proprio limite per meno di ε, e presi successivamente y_1 e y_2 in modo che $\Phi_n(y_2) - \Phi_n(y_1)$ differisca da 1 per meno di ε, avremo che la differenza dei limiti di $\Phi_n(y_2) - \Phi_n(0)$ e $\Phi_n(y_1) - \Phi_n(0)$ differirà da 1 per meno di $3\,\varepsilon$. Reciprocamente, se la differenza dei limiti è 1, $\Phi_n(y)$ tende uniformemente a

$$\Phi_\infty(t) = \lim_{n=\infty}[\Phi_n(y) - \Phi_n(0)] - \lim_{z=-\infty}\lim_{n=\infty}[\Phi_n(z) - \Phi_n(0)]\,;$$

infatti $\Phi_n(y) - \Phi_n(0)$ tende a $\Phi_\infty(y) - \Phi_\infty(0)$, scelti quindi $y_1\, y_2$ in modo che $\Phi_\infty(y_1) - \Phi_\infty(y_2)$ differisca da 1 per mono di ε, e fissato n così che, nei punti $y = y_1$ e $y = y_2$, $\Phi_n(y) - \Phi_n(0)$ differisca da $\Phi_\infty(y) - \Phi_\infty(0)$ per meno di ε, $\Phi_n(y_2) - \Phi_n(y_1)$ differisce da 1 per meno di $3\,\varepsilon$, ed essendo $0 \leq \Phi_n(y) \leq 1$, $\Phi_n(y_1)$ differirà per meno di $3\,\varepsilon$ da 0 e $\Phi_n(0)$ differirà da $\Phi_\infty(0)$ per meno di $4\,\varepsilon$; $\Phi_n(0)$ tende dunque a $\Phi_\infty(0)$ e ciò implica che per qualunque y

$$\Phi_n(y) = \Phi_n(0) + [\Phi_n(y) - \Phi_n(0)]$$

tenda a $\Phi_\infty(y)$. Le Φ_h sodisfanno la restrizione [4], e altrettanto si può quindi affermare per la Φ_∞, loro limite. Ciò prova che la funzione di ripartizione Φ_∞ è continua: allora vale il teorema di Polya (n. 3) che assicura dell'uniforme convergenza.

La [10] non permette di asserire che nelle nostre ipotesi la convergenza sia uniforme, perchè il secondo membro dipende da y e cresce indefinitamente con y. Tralasciando di esaminare a parte se e in quali ipotesi la convergenza non uniforme sia effettivamente possibile, è certo che in tali casi non si può parlare di legge limite di probabilità: ammettendo anche che $\Phi_n(y)$ tendesse (non uniformemente) a $\Phi_\infty(y)$, il fatto che

$$\lim_{y=\infty}\Phi_\infty(y) - \lim_{y=-\infty}\Phi_\infty(y) < 1$$

escluderebbe già che Φ_∞ possa essere una funzione di ripartizione (almeno nel senso ordinario).

Osserviamo soltanto per ora che condizione necessaria e sufficiente perchè la convergenza sia uniforme è che le $\Phi_n(y)$ tendano uniformemente a zero quando y tende a $-\infty$ e tendano uniformemente ad 1 quando y tende a $+\infty$.

6. — In questo caso possiamo asserire (n. 21) che la funzione caratteristica di Φ_∞ è la $\psi(t) = \lim_{n=\infty}\psi_n(t)$, e otteniamo così la espressione di $\Phi_\infty(y)$:

$$\Phi_\infty(y) - \Phi_\infty(0) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{1 - e^{-iyt}}{it}\,\psi(t)\,dt\,.$$

Ricapitolando: se al crescere di n la funzione caratteristica ψ_n di una legge variabile Φ_n tende uniformemente in ogni regione finita dell'asse reale a una funzione limite ψ, e se le $\Phi_n(y)$ hanno tutte densità limitata non superiore a un certo numero μ, e tendono uniformemente a 0 per $y \to -\infty$ ed a 1 per $y \to +\infty$, l'integrale

$$\int_{-\infty}^{+\infty} \frac{1-e^{-iyt}}{it}\,\psi(t)\,dt$$

esiste sempre, e la legge variabile tende alla legge fissa $\Phi_\infty(y)$ di cui $\psi(t)$ è la funzione caratteristica.

7. — Cerchiamo di togliere la restrizione finora posta, che le Φ_n abbiano densità uniformemente limitate. Componendo la variabile casuale X_n, che ha funzione di ripartizione $\Phi_n(x)$, colla variabile ausiliaria Y_μ che ha densità costante μ nell'intervallo $(-l,0)$, dove $l = \dfrac{1}{\mu}$, e densità nulla altrove, si ottiene una variabile casuale $Z_{n,\mu} = X_n + Y_\mu$ la cui funzione di ripartizione $\Phi_{n,\mu}(z)$ è continua e ha densità mai superiore a μ (n. 20). La funzione caratteristica è $\psi_{n,\mu}(t) = \psi_n(t) \cdot \chi_\mu(t)$, dove χ_μ è la funzione caratteristica di Y_μ, ed è una funzione intera.

Poichè per ipotesi ψ_n al crescere di n tende uniformemente in ogni intervallo reale verso una funzione limite ψ, così $\psi_{n,\mu}$ tenderà uniformemente verso la funzione $\psi\,\chi_\mu$. Ora, fissato μ, la successione di variabili casuali $Z_{n,\mu}$ sodisfa le ipotesi precedenti purchè sia uniforme la tendenza rispettivamente a 0 e ad 1 di $\Phi_n(y)$ per $y \to \pm\infty$, e quindi $\Phi_{n,\mu}(z)$ tende uniformemente, al crescere di n, verso $\Phi_{\infty,\mu}$, dove

$$[11] \qquad \Phi_{\infty,\mu}(z) - \Phi_{\infty,\mu}(0) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{1-e^{-izt}}{it}\,\psi(t)\,\chi_\mu(t)\,dt\,.$$

8. — Teniamo ora fermo z, e facciamo tendere μ ad ∞. Di $\Phi_{\infty,\mu}(z)$ consideriamo il massimo e il minimo limite per $\mu \to \infty$: otteniamo due funzioni non decrescenti di z che hanno uguali limiti destro e sinistro. Per dimostrarlo facciamo vedere anzitutto che se $y > x$

$$\max_{\mu=\infty} \lim \Phi_{\infty,\mu}(x) \leq \min_{\mu=\infty} \lim \Phi_{\infty,\mu}(y)\,.$$

Scelto ε ad arbitrio, si prenda $\bar\mu$ maggiore di $\dfrac{1}{y-x}$ e tale che $\Phi_{\infty,\bar\mu}(x)$ superi $\max\lim_{\mu=\infty} \Phi_{\infty,\mu}(x) - \varepsilon$, e successivamente si fissi N così che per $n > N$ $\Phi_{n,\bar\mu}(x)$ differisce per meno di ε da $\Phi_{\infty,\bar\mu}(x)$. Per proprietà note (n. 20) esisterà un numero $\xi \leq 1/\bar{\mu} < y - x$ tale che $\Phi_n(y) = \Phi_{n,\bar{\mu}}(y - \xi)$, e, risultando $y - \xi > x$, $\Phi_{n,\bar{\mu}}(x) \leq \Phi_{n,\bar{\mu}}(y - \xi)$,

[12] $$\Phi_n(y) \geq \Phi_{n,\bar{\mu}}(x) \geq \underset{\mu=\infty}{max\ lim}\ \Phi_{\infty,\mu}(x) - 2\varepsilon .$$

Per le stesse proprietà (n. 20) si sa che, qualunque sia μ, $\Phi_{n,\mu}(z) \geq \Phi_n(z)$ in ogni punto z. Quindi per $n > N$ e per μ arbitrario:

$$\Phi_{n,\mu}(y) \geq \underset{\mu=\infty}{max\ lim}\ \Phi_{\infty,\mu}(x) - 2\varepsilon ;$$

tenendo fisso μ e facendo tendere n all'infinito si ha

$$\Phi_{\infty,\mu}(y) \geq \underset{\mu=\infty}{max\ lim}\ \Phi_{\infty,\mu}(x) - 2\varepsilon ;$$

valendo tale disuguaglianza per μ qualunque, e potendosi supporre ε piccolo a piacere, ne scende senz'altro

[13] $$\underset{\mu=\infty}{min\ lim}\ \Phi_{\infty,\mu}(y) \geq \underset{\mu=\infty}{max\ lim}\ \Phi_{\infty,\mu}(x) \qquad \text{c. d. d.}$$

9. — Facendo tendere y ad x:

$$\underset{y=x}{lims} \left\{ \underset{\mu=\infty}{min\ lim}\ \Phi_{\infty,\mu}(y) \right\} \geq \underset{\mu=\infty}{max\ lim}\ \Phi_{\infty,\mu}(x) ,$$

e poichè il limite sinistro del limite sinistro è ancora il limite sinistro

$$\underset{y=x}{lims} \left\{ \underset{\mu=\infty}{min\ lim}\ \Phi_{\infty,\mu}(y) \right\} \geq \underset{y=x}{lims} \left\{ \underset{\mu=\infty}{max\ lim}\ \Phi_{\infty,\mu}(y) \right\}$$

da cui, non potendo evidentemente sussistere la relazione $>$, scende necessariamente

[14] $$\underset{y=x}{lims} \left\{ \underset{\mu=\infty}{min\ lim}\ \Phi_{\infty,\mu}(y) \right\} = \underset{y=x}{lims} \left\{ \underset{\mu=\infty}{max\ lim}\ \Phi_{\infty,\mu}(y) \right\} .$$

Facendo tendere nella [13] x ad y e procedendo allo stesso modo si ha ancora

[15] $$\underset{y=x}{limd} \left\{ \underset{\mu=\infty}{min\ lim}\ \Phi_{\infty,\mu}(y) \right\} = \underset{y=x}{limd} \left\{ \underset{y=\infty}{max\ lim}\ \Phi_{\infty,\mu}(y) \right\}$$

Ponendo

$$\Phi_{\infty,\infty}(x) = \frac{1}{2} \left[\underset{y=x}{lims} \left\{ \underset{\mu=\infty}{max\ lim}\ \Phi_{\infty,\mu}(y) \right\} + \underset{y=x}{limd} \left\{ \underset{\mu=\infty}{min\ lim}\ \Phi_{\infty,\mu}(y) \right\} \right]$$

si ha una funzione che ha ancora lo stesso limite destro e sinistro delle precedenti, non è mai decrescente, e tende a 0 e 1 quando x tende a $\pm\infty$ (come segue facilmente dall'ipotesi stabilita che la tendenza rispettivamente a 0 ed 1 di $\Phi_n(y)$ per $y \to \pm\infty$ sia uniforme). $\Phi_{\infty,\infty}$ sarà al più discontinuo in un insieme finito o numerabile di punti, in cui ha per valore la semisomma dei limiti destro e sinistro, ed è quindi una funzione di ripartizione.

Dimostriamo ora che nei punti in cui $\Phi_{\infty,\infty}(x)$ è continua si ha

$$\mathop{lims}_{n=\infty} \Phi_n(x) = \Phi_{\infty,\infty}(x),$$

e in generale si ha

$$\mathop{lims}_{y=x} \Phi_{\infty,\infty}(y) \leq \mathop{min\ lim}_{n=\infty} \Phi_n(x) \leq \mathop{max\ lim}_{n=\infty} \Phi_n(x) \leq \mathop{limd}_{y=x} \Phi_{\infty,\infty}(y).$$

La [12] mostra infatti che $\mathop{min\ lim}_{n=\infty} \Phi_n(y) \leq \mathop{max\ lim}_{\mu=\infty} \Phi_{\infty,\mu}(x) - 2\varepsilon$ essendo $x < y$ e $\varepsilon > 0$ arbitrario, da cui

$$[16] \qquad \mathop{min\ lim}_{n=\infty} \Phi_n(y) \geq \mathop{lims}_{x=y} \left\{ \mathop{max\ lim}_{\mu=\infty} \Phi_{\infty,\mu}(x) \right\} = \mathop{lims}_{x=y} \Phi_{\infty,\infty}(x).$$

Dalla $\Phi_{n,\mu}(y) \geq \Phi_n(y)$ scende poi $\Phi_{\infty,\mu}(y) \geq \mathop{max\ lim}_{n=\infty} \Phi_n(y)$ per qualunque μ, e quindi

$$[17] \qquad \mathop{max\ lim}_{n=\infty} \Phi_n(y) \leq \mathop{max\ lim}_{\mu=\infty} \Phi_{\infty,\mu}(y) \leq \mathop{limd}_{x=y} \Phi_{\infty,\infty}(x).$$

10. — Abbiamo con ciò dimostrato che la legge di probabilità Φ_n tende alla legge limite $\Phi = \Phi_{\infty,\infty}$ (nel senso in cui la tendenza al limite va intesa per le funzioni di ripartizione, come è spiegato dal Lévy, *op. cit.*, a pag. 194); ciò implica di per sè (Lévy, p. 195) che la funzione caratteristica corrispondente a Φ è la $\psi(t)$ cui tende per ipotesi $\psi_n(t)$, e possiamo quindi rispondere affermativamente alla nostra questione mediante il seguente

Teorema.

Se si ha una successione di leggi di probabilità le cui funzioni caratteristiche tendono uniformemente in ogni intervallo limitato verso una funzione limite ψ, e per le quali la probabilità di valori superiori in modulo a* x *tende uniformemente a zero al crescere di* x*, la successione tende a una legge limite, di cui ψ, di conseguenza, è la funzione caratteristica.

Analiticamente, ciò significa che esiste, è reale, mai decrescente, e ha variazione totale (da $-\infty$ a $+\infty$) $= 1$, l'integrale

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1 - e^{-iyt}}{it} \psi(t)\, dt$$

considerato come funzione di y; postolo uguale a $\Phi(y) - \Phi(0)$, ove $\Phi(0)$ è tale che $\lim_{y=-\infty} \Phi(y) = 0$, si ha che $\Phi_n(y) \rightarrow \Phi(y)$ per ogni y all'infuori, tutt'al più, dei punti di discontinuità, che costituiscono, nella peggiore delle ipotesi, un insieme numerabile. L'importanza di tale possibile eccezione appare del resto concettualmente nulla se si pensa che è soltanto per effetto di una convenzione che nei punti di discontinuità si prende come valore della Φ la semisomma dei limiti destro e sinistro.

11. Nel caso che interessa la nostra precedente ricerca, e cioè per le funzioni caratteristiche di fenomeno aleatorio, può avere però qualche interesse osservare che l'eccezione accennata non si verifica mai, e cioè che la Φ_n tende a Φ anche negli eventuali punti di discontinuità.

Rammentiamo che se si ha un fenomeno di probabilità costante p la legge delle frequenze tende alla distribuzione gaussiana simmetrica rispetto a p, e quindi in tal caso $\lim_{n=\infty} \Phi_n(p) = \frac{1}{2}$. Il più generale fenomeno aleatorio per cui la $\Phi(y)$ ha una discontinuità per $y = p$ si ottiene combinando (nel senso del § 24) il fenomeno precedente e uno che non presenta discontinuità per $y = p$. E si vede senz'altro che

$$\lim_{n=\infty} \Phi_n(p) = \frac{1}{2}\left[\operatorname*{limd}_{y=p} \Phi(y) + \operatorname*{lims}_{y=p} \Phi(y)\right] = \Phi(p).$$